Agreement in Classifying Bloodstream Infections Among Multiple Reviewers Conducting Surveillance

Jeanmarie Mayer,[1] Tom Greene,[1] Janelle Howell,[2] Jian Ying,[1] Michael A. Rubin,[1,2] William E. Trick,[3] and Matthew H. Samore,[1,2] for the CDC Prevention Epicenters Program[4]

[1]University of Utah School of Medicine, and [2]Salt Lake City Veterans Affairs Medical Center IDEAS Center, Salt Lake City, Utah; [3]Stroger Hospital of Cook County, Chicago, Illinois; and [4]Centers for Disease Control and Prevention, Atlanta, Georgia

Background. Mandatory reporting of healthcare-associated infections (HAIs) is increasing. Evidence for agreement among different reviewers applying HAI surveillance criteria is limited. We aim to characterize agreement among infection preventionists (IPs) conducting surveillance for central line–associated bloodstream infection (CLABSI) with each other and as compared with simplified laboratory-based definitions.

Methods. Abstracted electronic health records were assembled from inpatients with positive blood cultures at a tertiary-care Veterans Affairs (VA) hospital over a 5-year period. Identical patient records were made available to VA IPs from different facilities to report on CLABSI using their usual surveillance methods. Positive blood cultures were also evaluated using laboratory-based definitions. Standard indices of interrater agreement, expressed as a κ statistic, were computed between IPs, and between IPs and simplified laboratory-based methods.

Results. Overall, 114 patient records were reviewed by 18 IPs, the majority of whom specified they followed National Healthcare Safety Network criteria. The overall agreement among IPs by κ statistic was 0.42 (standard error [SE], 0.06). IPs had better agreement with a simple laboratory-based definition with an average κ of 0.55 (SE, 0.05). The proportion of patient records that 18 IPs reported with CLABSI ranged from 14% to 39% (overall mean, 28% with a coefficient of variation of 25%). When simple laboratory-based methods were applied to different sets of patient records, classification was more consistent with CLABSI assigned in a proportion ranging from 36% to 42% (overall mean, 39%).

Conclusions. Reliability of IP-conducted surveillance to identify HAI may not be ideal for public reporting goals of interhospital comparisons.

Although public reporting of healthcare-associated infections (HAIs) is increasing [1, 2], the determination of HAI is not straightforward and comparisons across institutions can be misleading [3]. Surveillance conducted by infection preventionists (IPs) is time consuming, and the traditional reference standards they use are Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN)

standardized case definitions [4, 5]. Application of these partially subjective criteria requires considerable clinical judgment, extensive education, and training and can be complicated by variation in skill, knowledge, and experience [6–9]. Several studies have compared prospective IP evaluations with retrospective review to classify bloodstream infections (BSIs), with interrater correlation coefficients, or κ statistics, ranging from 0.31 to 0.66 [10–13]. For tracking trends within an institution, variation is unlikely a problem in facilities with low IP turnover, established training programs, and stable supervision. However, using a subjective system across institutions raises comparability concerns, and simpler laboratory-based definitions that can be automated have been proposed [3, 11, 14–18].

Studies assessing interrater reliability of IP-applied surveillance in practice would assist policy makers in

determining optimal methods for public reporting. To our knowledge, our study is the first to characterize agreement among multiple IPs reviewing identical patient records presented in a standardized "actual-practice" format. IPs were asked to use their traditional surveillance methods, with agreement among IPs in assigning the determination of central line–associated BSI (CLABSI) at the individual patient record and as a proportion of a set of patient records. We hypothesized that IP disagreement would occur and potentially was more likely if IPs reviewed ambiguous or complicated patient records or had different thresholds in reporting a CLABSI.

METHODS

Patient Sample and IP Participants

Patients admitted to a 110-bed tertiary-care Veterans Affairs (VA) Health Care System hospital between January 2001 and December 2005 with blood cultured >2 days from admission growing a microorganism(s) met criteria for inclusion. Patients defined as sensitive (*International Classification of Diseases, Ninth Revision* codes for human immunodeficiency virus, substance abuse, or mental health treatment) were excluded [19]. From a pool of 441 eligible patients, 114 admissions were randomly selected.

Study participants were recruited via email to IPs at 138 acute inpatient VA facilities. The first respondent from each facility with at least 6 months of experience was eligible to participate during enrollment (September–December 2009). Eighteen participants enrolled and were free to withdraw at any time, their responses were kept confidential, and compensation was based on the number of patient records reviewed.

The study was approved by the institutional review boards of the University of Utah and the Salt Lake City VA Health Care System.

Web Application to Display and Report on Patient Records

Data extracted from the data warehouse included microbiology, unit location and room transfers, antimicrobial therapy, white blood cell counts, vital signs, and selected clinical narratives over a hospitalization. All documents of a particular category (admission and discharge summaries, infectious disease consultations) and selected documents around the time of positive blood cultures (physician/nursing notes and radiograph reports) were used. Patient records lacking physician notes were excluded. A deidentification software tool removed all identifiers except dates [20, 21]. Information on central venous catheters was not standardized and reviewers needed to infer catheter use from clinical narratives and radiographic reports. Patient records were displayed and accessible to IPs in a familiar electronic health record "search and select" format through a password-protected Web-based interface on a secure VA site. Webinar training and an online manual informed participants how to access data and report surveillance decisions.

Surveillance by IP Review and Laboratory-Based Definitions

An efficient allocation strategy was used to simulate a "real-practice" workload for IPs while maximizing the spectrum of reviewers and patient records. The study dataset of 114 patient records was randomly divided into 4 smaller workable sets. Eighteen IPs were randomly divided into 4 groups, with each IP group randomly assigned to a set of patient records. IPs were instructed to perform surveillance for CLABSI as they normally would and were given a deadline. To explore the level of certainty related to dichotomous yes/no decision making typical for surveillance [4], IPs were asked to provide their level of confidence for CLABSI on a 7-point Likert scale ("definite no" to "definite yes"), describe their decisions, and comment on adequacy of the data and, upon completion of the full review, surveillance criteria used. IPs were considered in agreement if all identified no CLABSI or if all IPs identified ≥1 CLABSI per patient record.

An algorithm to identify CLABSI comprised solely of simplified laboratory-based definitions approximating NHSN criteria as previously published was developed for reference comparison [3, 11]. Because not all data were in an easily retrievable electronic format, a document to standardize manual application of the algorithm was developed. A series of rules applied to blood cultures classified distinct BSI episodes. Blood positive for organisms spanning a 5-day period were defined as a discrete episode, or 30-day period when the same organism was recovered. An episode was classified as a true BSI or contaminant based on organism type and number of positive cultures. A true BSI was identified as secondary if the same organism was cultured from other specified sources 5 days before through 10 days after BSI onset. Primary BSI was classified as CLABSI if a central venous catheter was in place within 48 hours prior to onset. Two investigators independently applied the algorithm to each patient record and adjudicated oppositional cases to consensus.

Statistical Analysis

Power calculations using statistical simulation supported a strategy where patient records were randomly divided into a number of workable sets, with each set randomly assigned to one of multiple IP groups. The motivation for this strategy was the limited time available to IP participants; it also provided a more precise estimate of the interrater κ statistic than the alternative strategy of having all IPs review the same patient records from a smaller dataset. The projected standard error (SE) for κ with 114 records divided into 4 sets of patient

records, where each set was given to 1 of 4 IP groups, was 0.058, compared to a projected SE of 0.079 if a single set of 30 patient records was reviewed by all 18 IPs.

IP agreement was assessed using Fleiss's generalization of Cohen κ statistic to multiple raters [22, 23]. Systematic variation between raters in the proportion of positive classifications was evaluated by the coefficient of variation of proportion of positive ratings across IPs, with correction for random sampling error. The coefficient of variation scales the standard deviation of the respective rater's proportions of positive ratings to the average proportion, and indicates the extent to which differences among IPs could lead to systematic biases in comparisons of their infection rates. Because the study design subdivided patient records into multiple sets, each reviewed by a separate group of IPs, separate estimates of κ and the coefficient of variation were computed for each group of IPs and then pooled to produce overall estimates.

RESULTS

Characteristics of IP Study Participants and the Patient Sample

IPs reported a median of 4.3 years of surveillance experience (range, 1.75–25 years). Seventy-two percent were certified in infection prevention and control. Although only 22% of IPs formally reported to NHSN, 83% specified that they adhered to NHSN surveillance criteria. IP hospitals were representative of VA acute inpatient facilities nationwide in size and geographic location. The median bed size for the 18 IP hospitals was 125 (range, 21–347) while that of VA nationwide was 114 (range, 9–397). IP hospitals were from 14 of 21 geographical regions. The number of IPs and their makeup by years of experience and facility size were similar across IP groups.

Patient characteristics and microbiology were similar across sets of patient records (Table 1). The overall distribution of organisms for review comprised common skin commensals in 38%, gram-positive recognized pathogens in 31%, gram-negatives in 19%, yeast in 8%, or polymicrobial with recognized pathogens in 4%. More than one set of blood cultures was positive in 50% of episodes for review. Thirty-six percent of patients with positive episodes were located on a critical care unit.

CLABSI Assignment by Algorithm and IP Review

Classifications assigned by researchers applying the laboratory-based algorithm are provided in Table 2. There were 2 disagreements in 114 patient records, which were resolved upon further review (overlooked nonblood culture; difficulty determining central line presence). In addition, when researchers applied the algorithm, 87% of episodes (46 of 53) from common skin commensals were assigned as

Table 1. Characteristics of the Patient Sample (N = 114 Patient Records)

Description of Individual Records[a]	Median	Range
Patient age, y	66	32–89
Length of stay, d	24	3–107
Clinical notes, no.	43	10–124
Microbiology cultures, no.	11.5	1–61
Blood culture sets, no.	5	1–28
Overall microbiologic data	**n**	
All blood culture sets (positive sets)	750 (210)	
Episodes of positive blood cultures for review[b]	138[c]	
Episodes by organism		
Common skin commensals[d]	53	
Staphylococcus aureus	26	
Enterococcus spp	13	
Klebsiella spp	10	
Candida spp	11	
Enterobacter spp	8	
β-hemolytic streptococcus	3	
Escherichia coli/Serratia/Pseudomonas	2 each	
Acinetobacter/Proteus/Mycobacterium fortuitum	1 each	
Mixed	4	

[a] Ninety-nine percent of patients were male.

[b] Positive cultures were grouped into episodes for review over a 5-day period, or, if the same organism was recovered from an initial infection, a 30-day period.

[c] 20 records had >1 episode for review.

[d] Common skin commensals as defined by the National Healthcare Safety Network.

"contaminants" whereas 43% of episodes (37 of 85) associated with recognized pathogens were assigned as CLABSI.

The proportion of patient records classified as CLABSI by algorithm was similar across sets of patient records and

Table 2. Classification of Culture-Positive Episodes (n = 138)[a] Applying a Simple Laboratory-Based Definition

Classification Category	Episodes, n
Contaminant	46
Secondary BSI	30
Primary BSI without central venous catheter	14
CLABSI	48 (from 44 records)
Due to common skin commensal vs recognized pathogen[b]	11 vs 37
Occurred in ICU vs non-ICU location	20 vs 28

Abbreviations: BSI, bloodstream infection; CLABSI, central line–associated bloodstream infection; ICU, intensive care unit.

[a] Consensus determinations of 2 research reviewers applying a laboratory-based definition.

[b] As defined by the National Healthcare Safety Network.

Table 3. The Proportion of Records With Agreement Among Infection Preventionists and a Laboratory-Based Definition in Assigning Central Line–Associated Bloodstream Infection

			Proportion of Records Assigned With CLABSI			
				IP		
Group	IPs, n	Patient Records, n	Laboratory-Based Definitions	Mean	Range	Coefficient of Variation (%)
1	4	29	0.42	0.30	0.21–0.38	24
2	4	29	0.39	0.26	0.17–0.350	32
3	5	28	0.36	0.21	0.14–0.290	26
4	5	28	0.39	0.33	0.25–0.390	9
Overall[a]	18	114	0.39	0.28	0.14–0.390	25

Abbreviations: CLABSI, central line–associated bloodstream infection; IPs, infection preventionists.

[a] Separate estimates of coefficient of variation computed for each group of IPs and pooled across groups to produce overall estimates.

ranged from 36% to 42% (mean, 39%) (Table 3). In contrast, the proportion of patient records reported by IPs as having CLABSI varied significantly across IPs, even within the same patient record sets. Overall, IPs reported a minimum of 14% to a maximum of 39% of patient records with CLABSI, for an overall mean of 28% and a coefficient of variation of 0.25.

Agreement on surveillance decisions for individual records within IP groups were calculated as separate κ statistics and pooled across groups to produce the overall κ of 0.42 (SE, 0.06) (Table 4). There was a trend for IPs to have better agreement with classifications made by algorithm than with each other (Table 4), with an overall average κ across 18 IPs of 0.55 (SE, 0.05). In addition to data provided in tables, IPs uniformly agreed with each other when reviewing identical patient records in 55% of records. Discordant classifications were noted in 5% of patient records where an IP identified a CLABSI but the algorithm did not, and in 16% of patient records where the algorithm identified a CLABSI and the IP did not.

Table 4. Agreement Among Infection Preventionists and With a Laboratory-Based Definition Assigning Central Line–Associated Bloodstream Infection for Individual Patient Records

			Interrater Agreement for Individual Patient Records	
Group	IPs, n	Patient Records, No.	IP vs IP, κ (SE)	IP vs Laboratory-Based, κ (SE)
1	4	29	0.50 (0.11)	0.66 (0.09)
2	4	29	0.46 (0.13)	0.62 (0.10)
3	5	28	0.30 (0.13)	0.29 (0.13)
4	5	28	0.43 (0.09)	0.65 (0.09)
Overall[a]	18	114	0.42 (0.06)	0.55 (0.05)

Abbreviations: IPs, infection preventionists; SE, standard error.

[a] Separate estimates of κ computed for each group of IPs and pooled across groups to produce overall estimates.

There were 4 patient records where the algorithm identified CLABSI but all IPs did not. IP discrepancies for these cases included: (1) enterococcus considered secondary (though blood and secondary isolates had different susceptibilities); (2) candida considered secondary to respiratory source (laboratory-based definition classified yeast in sputum a contaminant); (3) candida from blood considered a contaminant; and (4) a second BSI episode not addressed.

Patient records for IP review contained actual data, and some had inconsistencies or difficult-to-locate or missing information as found in real practice. The majority of IPs reported data quality problems in 14 of 114 patient records, with 12 due to difficulties in determining the duration, though not presence, of central venous catheterization.

IP Uncertainty

We found differences among IPs on their self-reported certainty that a CLABSI was or was not present. The "least sure" IP was definitely certain in only 20% of patient records reviewed, as compared to the "most sure" who was definitely certain in 97% of records (median, 58%). However, even when certainty was not definite, responses still leaned toward a yes or no. "Completely unsure" was selected only 23 times in 512 reviews (4.5%) by 8 IPs (Figure 1). Definite yes or definite no certainty was most common in patient records where all IPs reviewing uniformly agreed (Figure 1). For 233 reviews associated with disagreement, the distribution of certainty was bimodal (Figure 1) with definite/probable yes or definite/probable no most commonly selected. Thus, although there was a nearly linear relationship for the average certainty score from definite no through definite yes, the full distribution of IP certainty levels revealed IP disagreement rather than uncertainty. In other words, there were 2 distinct camps in which some IPs were certain CLABSI was present, while others were certain it was not.
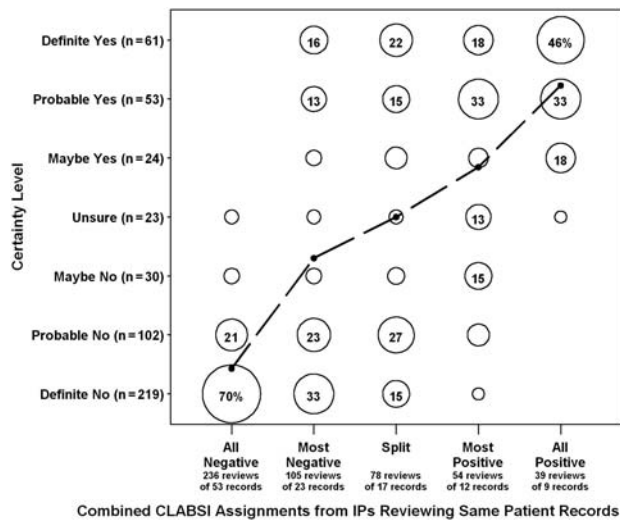
Figure 1. The distribution of infection preventionist (IP) certainty for central line–associated bloodstream infection (CLABSI) within categories of combined CLABSI assignments when 18 IPs evaluated the same 114 patient records in 512 reviews. The categories where IPs all agreed CLABSI was not present is "All Negative" and where all agreed CLABSI was present is "All Positive;" other categories had some IP disagreement. Numbers within bubbles represent percentages of IPs with that level of certainty for CLABSI within that combined CLABSI assignment category (columns sum to 100%). The mean IP certainty level for each combined CLABSI assignment category is the average of individual IP reported certainty levels ("Definite No" as 1 through "Definite Yes" as 7) represented by the dashed line.

DISCUSSION

Surveillance is defined as the ongoing systematic collection, interpretation, and dissemination of data reflecting the status of a population [24, 25]. Various stakeholders use HAI surveillance for different reasons, and different approaches may be needed to address divergent goals [26]. When evaluating the impact of improvement processes within facilities, it is most important to accurately identify each infection. On the other hand, the objectives of public reporting to benchmark and accurately rank facilities call for criteria providing the most consistent outcomes, even at the expense of clinical accuracy [18]. Conventional IP surveillance applying NHSN definitions to identify HAIs has been the practice standard [4]. NHSN criteria, though standardized, require subjective judgment and are not explicit, such as in determining a primary BSI when "a recognized pathogen cultured from one or more blood cultures and the organism cultured from blood is not related to an infection at another site" [4]. Although a uniform and reliable system is essential for comparison across facilities, several recent reports documented variation in IP surveillance determinations. One state health department found that >50%

misclassified NHSN-reported CLABSIs, the bulk of discordant assignments representing underreporting [7]. Another study found that more than half of CLABSIs were missed in Australian hospitals [27].

We present the results of a large, systematic study assessing reliability of IP surveillance using multiple reviewers examining identical patient records presented in a simplified real-life electronic health record format and then compared to a solely laboratory-based method. Neither explicit surveillance criteria nor standardized abstraction forms were provided in order to assess reproducibility of IP surveillance as normally practiced. We found substantial variation in IP surveillance for CLABSI, considered the most straightforward HAI, at both an ecologic (proportion of patient records reported with CLABSI) and an individual patient record level (interrater agreement). A novel approach, soliciting IP-perceived certainty for CLABSI, was used to explain differences in IP decisions: in the majority of patient records where IPs disagreed, IPs were certain. When we compared surveillance classifications using a laboratory-based definition to that assigned by IPs, although the simplified algorithm called more infections, individual IPs agreed more often with the laboratory-based definition than with each other. There are several potential explanations for differences in IP surveillance decisions. Discrepancies may have been due to oversights or differences in locating data, but an important contributor is the CLABSI surveillance criteria. These criteria require some subjective judgment and differences may arise with different mental models in approaching surveillance. Comments from IPs revealed variation in how criteria were applied and data were interpreted, and in the level of evidence needed to assign CLABSI to a patient record.

Clinicians implementing preventive measures expect reductions in infections and covet specific definitions that identify confirmed clinical infections [28, 29]. Despite an expectation that trained IPs conducting detailed reviews identify only unambiguous infections, there remains risk for variation. Lin et al [3] used a retrospective cohort of 20 intensive care units across multiple medical centers to examine the correlation between prospective IP-derived CLABSI rates and a retrospectively applied computer algorithm reference standard. Median IP rates (3.3 per 1000 line days; interquartile range, 2.0–4.5) and algorithm rates (9.0 per 1000 line days; interquartile range, 6.3–11.3) had weak correlation ($P = .34$) [3]. Regression modeling revealed significant heterogeneity between IP and algorithm rates among institutions, suggesting local variation in surveillance practices. Interestingly, the facility with the lowest IP rate had the highest rate by algorithm, highlighting how variation in surveillance can complicate interfacility comparisons [3]. Likewise, higher interrater agreement with original surveillance decisions by blinded IPs conducting evaluations for ventilator-associated pneumonia

was found with a streamlined version using objective quantitative measures ($\kappa = 0.79$; confidence interval, .62–.97) than with conventional NHSN criteria ($\kappa = 0.45$; confidence interval, .26–.64) [30]. The pressure to "get to zero" raises concerns that partially subjective surveillance definitions applied inconsistently could be exploited or prone to subconscious cognitive bias to lower infection rates [31]. Variability may arise from the same cognitive biases described in clinical decision making, including the tendency to opt for decisions that lead to good outcomes (outcome bias), the benign condition during situations of ambiguity (frequency gambling), and affective sources of error (visceral bias) [32]. These biases can be removed through application of simplified laboratory-based definitions, albeit automated or manually.

Our study and those of others have shown that surveillance done across hospitals using laboratory-based definitions perform more uniformly and is a preferred signal to traditional surveillance when the goal is interinstitutional comparison [3, 18, 30]. Although simulation models suggest that laboratory-based definitions likely overcall CLABSI compared to clinical criteria, laboratory-based definitions provide more accurate rankings of institutions [18]. Evaluation for the future should address implications of reliability in surveillance methods when the outcome of interest is a rare occurrence, as is becoming the case for CLABSI [33, 34]. Consideration of a "blended" surveillance method, where IPs could report disagreements with algorithmic decisions, would provide the standardization of laboratory-based criteria yet improve the positive predictive value through expert confirmation. Such a system was used and supported by IPs across healthcare systems from 2008 to 2010 in the state of Utah (local experience).

Our study had several limitations. IP review was retrospective and we could not replicate some practices, including consultation with clinicians or unlimited access to patient information. Although reviewers were told there was no "right answer" and to work normally, they still may have altered their decision making. The population under study was comprised of mainly men within the VA Health Care System, and thus patient comorbidities and microbiologic and surveillance practices may not be generalizable to other healthcare organizations. The patient records used were from a time when national CLABSI rates were higher, and while this allowed us to maximize efficiency with essentially an enriched sample, the positive predictive value of the laboratory-based definitions may have been inflated. However, given a setting of lower CLABSI prevalence, the objective criteria we used could be modified to minimize the potential for false positives.

In summary, our study adds to previous literature evaluating HAI surveillance systems by allowing multiple IPs to review and report on the same records. We showed, as did others using single reviewers, that laboratory-based definitions had less heterogeneity for HAI detection between hospitals than did routine IP surveillance [3, 30]. Furthermore, we found that IP discordant surveillance classifications were due to disagreement, rather than from uncertainty or ambiguous cases. Our results support the use of laboratory-based definitions as better suited for public reporting of infection rates when the primary goal is to rank facilities [3, 18, 26]. There is a need for infection preventionists, healthcare epidemiologists, and policy makers to come to consensus on the value and limitations of more streamlined objective quantitative measures of identifying HAIs as compared with traditional manual methods for the purposes of reproducible public reporting and performance measurements.

Notes

References

1. Rosenthal MB. Nonpayment for performances? Medicare's new reimbursement rule. N Engl J Med 2007; 357:1573–5.
2. Association for Professionals in Infection Control and Epidemiology. Legislation in progress. 2008. Available at: http://www.apic.org/script content/custom/dyncontent/legislation/index.cfm?section=government_advocacy. Accessed 18 October 2010.
3. Lin MY, Hota B, Khan YM, et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. JAMA 2010; 304:2035–41.
4. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control 2008; 36:309–32.
5. Stricof RL, Schabses KA, Tserenpuntsag B. Infection control resources in New York state hospitals. Am J Infect Control 2008; 36:702–5.
6. Burke JP. Patient safety: infection control—a problem for patient safety. N Engl J Med 2003; 348:651–6.
7. Backman LA, Melchreit R, Rodriguez R. Validation of the surveillance and reporting of central line-associated bloodstream infection data to a state health department. Am J Infect Control 2010; 38:832–8.
8. Nieder MF. 2008 National Association of Children's Hospitals and Related Institutions Pediatric Intensive Care Unit Patient Care FOCUS group. Am J Infect Control 2010; 38:585–95.
9. Worth LJ, Brett J, Bull AL, et al. Impact of revising the National Nosocomial Infection Surveillance System definition for catheter-related bloodstream infection in ICU: reproducibility of the National Healthcare Safety Network case definition in an Australian cohort of infection control professionals. Am J Infect Control 2009; 37:643–8.

10. Yokoe DS, Anderson J, Chambers R, et al. Simplified surveillance for nosocomial bloodstream infections. Infect Control Hosp Epidemiol **1998**; 19:657–60.

11. Trick WE, Zagorski BM, Tokars JI, et al. Computer algorithms to detect bloodstream infections. Emerg Infect Dis **2004**; 10:1612–20.

12. Emori TG, Edwards JR, Culver DH, et al. Accuracy of reporting nosocomial infections in intensive care unit patients to the National Nosocomial Infections Surveillance System: a pilot study. Infect Control Hosp Epidemiol **1998**; 19:308–16.

13. Haley RW, Schaberg DR, McGlish DK, et al. The accuracy of retrospective chart review in measuring nosocomial infection rates. Amer J Epidemiol **1980**; 111:516–33.

14. Tokars JI, Richards C, Andrus M, et al. The changing face of surveillance for health care-associated infections. Clin Infect Dis **2004**; 39:1347–52.

15. Klompas M, Yokoe DS. Automated surveillance of health care-associated infections. Clin Infect Dis **2009**; 48:1268–75.

16. Woeltje KF, Butler AM, Goris AJ, et al. Automated surveillance for central-line associated bloodstream infection in intensive care units. Infect Control Hosp Epidemiol **2008**; 29:842–6.

17. Leal J, Gregson DB, Ross T, et al. Development of a novel electronic surveillance system for monitoring of bloodstream infections. Infect Control Hosp Epidemiol **2010**; 31:740–7.

18. Rubin MA, Mayer J, Greene T, et al. An agent-based model for evaluating surveillance methods for catheter-related bloodstream infection. AMIA Annu Symp Proc **2008**; 6:631–5.

19. Veterans Health Administration (VHA) directive 1605.1, privacy program, U.S.C. 7332, implemented by 38 CFR §1.460–1.496.

20. Friedlin FJ, McDonald CJ. A software tool for removing patient identifying information from clinical documents. J Am Med Inform Assoc **2008**; 15:601–10.

21. Health Insurance Portability and Accountability Act. Standards for privacy of individually identifiable health information. Federal Register 53181 (2002) (codified at 45 CFR 160 and 164).

22. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement **1960**; 20:37–46.

23. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin **1971**; 76:378–82.

24. Thacker SB, Berkelman RS. Public health surveillance in the United States. Epidemiol Rev **1988**; 10:164–90.

25. Centers for Disease Control and Prevention. Updated guidelines for evaluating public health surveillance systems: recommendations from the guidelines working group. MMWR Recomm Rep **2001**; 50 (RR13):1–35.

26. Magill SS, Fridkin SK. Improving surveillance definitions for ventilator-associated pneumonia in an era of public reporting and performance measurement. Clin Infect Dis **2012**; 54:378–80.

27. McBryde ES, Brett J, Russo PL, et al. Validation of statewide surveillance system data on central line-associated bloodstream infection in intensive care units in Australia. Infect Control Hosp Epidemiol **2009**; 30:1045–9.

28. Sexton DJ, Chen L, Anderson DJ. Current definitions of central line-associated bloodstream infection: is the emperor wearing clothes? Infect Control Hosp Epidemiol **2010**; 31:1286–9.

29. Shuman EK, Washer LL, Arndt JL, et al. Analysis of central line-associated bloodstream infections in the intensive care unit after implementation of central line bundles. Infect Control Hosp Epidemiol **2010**; 31:551–3.

30. Klompas M, Kleinman K, Khan Y, et al. Rapid and reproducible surveillance for ventilator-associated pneumonia. Clin Infect Dis **2012**; 54:370–7.

31. Klompas M. Ventilator associated pneumonia: is zero possible? Clin Infect Dis **2010**; 51:1123–6.

32. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med **2003**; 78:775–80.

33. Dudeck MA, Horan TC, Peterson KD, et al. National Healthcare Safety Network (NHSN) report, data summary for 2009, device-associated module. Available at: www.cdc.gov/nhsn/PDFs/dataStat/2010NHSNReport.pdf. Accessed 30 January 2012.

34. Schulman J, Stricof R, Stevens TP, et al. Statewide NICU central line associated bloodstream infection rates decline after bundles and checklists. Pediatrics **2011**; 127:436–44.