The Dilemma of Assessment Bias in Infection Control Research

Michael Y. Lin1 and Marc J. M. Bonten2,3

1Department of Internal Medicine, Section of Infectious Diseases, Rush University Medical Center, Chicago, Illinois; 2Department of Medical Microbiology, and 3Julius Center for Public Health and Primary Care, University Medical Center Utrecht, The Netherlands

Infection control studies often rely on infection endpoints to determine whether interventions are effective. However, many infection outcomes, including those defined by standardized surveillance criteria, involve some subjective judgment for determination. Studies that use unblinded ascertainment of subjective infection endpoints are at risk of assessment bias. Unfortunately, infection control studies have not routinely accounted for assessment bias. To ensure validity, infection control studies should incorporate study design elements to control assessment bias, such as blinded assessment or use of objective outcome measures.

ASSESSMENT BIAS

Bias refers to any systematic flaw in study methodology that tends to shift study results away from the true result [2]. Bias compromises the validity of a study, and its effect cannot be removed through statistical adjustment. Assessment bias, which is also called ascertainment bias, diagnostic bias, or observer bias, occurs when assessment of a study subject’s outcome is influenced by the knowledge of the subject’s exposure status. The major cause of assessment bias is lack of blinding, and risk of bias is greatest when assessment of the study outcome requires subjective judgment [3].

Assessment bias is powerful, and an investigator’s prior expectations can lead to falsely positive results when in fact, no difference exists (type 1 error). To illustrate, when human experimenters were asked to assess groups of rats for number of correct responses and speed in a maze test, they found that rats they were led to believe were “maze-bright” had superior performance...
compared with rats they believed were “maze-dull”; in actuality, the groups were randomly distributed standard laboratory rats. Remarkably, the experimenters’ expectations insidiously biased their measurements and perceptions to create a falsely significant difference in the direction supporting their prior beliefs [4].

Lack of blinding generally biases a study’s outcome toward efficacy; the magnitude of the bias has been estimated at about 17% odds ratio exaggeration in favor of efficacy, even when including studies with completely objective outcomes such as mortality [5]. Furthermore, studies with composite endpoints that mix subjective (clinician-dependent) and objective (clinician-independent) elements are twice as likely to report a significant effect compared with studies with objective elements alone [6].

**INFECTION OUTCOMES AND THE RISK OF BIAS**

Most investigators recognize that clinical diagnosis of infection requires substantial judgment due to the limits of available clinical information and lack of a gold-standard test. For example, the recovery of pathogenic bacteria from endotracheal secretions of a febrile mechanically ventilated intensive care unit (ICU) patient may or may not represent hospital-acquired pneumonia, and the physician must make a judgment using clinical, laboratory, and radiographic data. Thus, infection-control investigators seeking a more robust study outcome often utilize public health surveillance definitions of infection, such as those developed by the Centers for Disease Control and Prevention’s National Healthcare Safety Network (NHSN) [7] as well as those utilized by other public health surveillance networks [8–10]. Because surveillance definitions were developed to minimize subjectivity and improve reliability, the implicit belief is that they can be standardized across healthcare facilities and immune to external influences. However, recent studies have shown that even standardized surveillance definitions, like clinical determinations, contain ample opportunities for subjectivity that need to be recognized in the context of research.

**Central Line–Associated Bloodstream Infection**

Central line–associated bloodstream infections (CLABSIs) are an infection outcome for many infection control studies. Despite the utilization of standard definitions and training, there may be variability in the performance of CLABSI surveillance [11]. In a study of identical patient records reviewed by multiple infection preventionsists, up to 40% of the cases were described as “uncertain,” and the overall agreement among multiple raters was only moderate (κ, 0.45) [12]. Poor-to-moderate interrater reliability has been demonstrated in other studies [13, 14]. Such variability likely reflects the underlying clinical uncertainty in diagnosing many catheter-related bloodstream infections. Although common CLABSI surveillance definitions contain objective elements (“patient has a recognized pathogen cultured from 1 or more blood cultures”), there are subjective elements that rely on the assessor’s judgment (“organism cultured from blood is not related to an infection at another site”) [15]. In practice, incomplete clinical data or subjective assessments of infections at non-blood body sites can complicate judgment of whether a bloodstream pathogen originated from the bloodstream; for example, *Escherichia coli* in the blood, recovered from a patient with a central venous catheter who had recent abdominal surgery, could reasonably represent either a CLABSI or a surgical site infection. Anecdotal disagreement in CLABSI determinations among expert reviewers underscores the surveillance definition’s uncertainty in some clinical contexts [16].

**Ventilator-Associated Pneumonia**

Ventilator-associated pneumonias (VAPs) are even more clinically challenging than catheter-related bloodstream infections to diagnose because many common conditions can mimic VAP, such as acute respiratory distress syndrome, thromboembolic disease, pulmonary hemorrhage, congestive heart failure, and atelectasis [17]. The clinical challenges in diagnosing VAP are reflected in many standardized definitions of VAP. Commonly used definitions contain multiple subjective elements, requiring the assessor to judge “new or progressive and persistent infiltrate” or “new onset of purulent sputum, or change in character of sputum, or increased respiratory secretions, or increased suctioning requirements” [18]. In a study of 50 ICU patients on mechanical ventilation, 2 experienced assessors used NHSN criteria to assess for VAP; 1 assessor identified nearly twice the number of VAPs (20) as the other (11), for only moderate agreement (κ, 0.50) [19]. Including microbiologic criteria such as respiratory cultures (not uniformly available and thus optional in most surveillance definitions) increases specificity, but cultures lack sensitivity [17] and, more importantly, do not eliminate the subjective criteria.

A standardized pneumonia measure, the clinical pulmonary infection score, combines clinical, radiologic, and microbiologic criteria for diagnosis of VAP. Yet when the clinical pulmonary infection score was compared with quantitative bronchoalveolar lavage cultures as a reference standard, specificity was low and interrater agreement between 2 intensivist assessors was imperfect [20].

**Other Infection Outcomes**

Other infection outcomes that are defined by surveillance definitions contain elements requiring subjective interpretation (Table 1) [7]. Common definitions for catheter-associated urinary tract infections require at least 1 of the following signs
or symptoms—fever, suprapubic tenderness, or costovertebral angle tenderness—with “no other recognized cause” [21]. In actuality, symptoms are often subtle or nonexistent among bladder-catheterized patients [22], and determining the source of fever requires substantial judgment. Similarly, *Clostridium difficile* infection surveillance requires the assessor to judge whether patient has “liquid stool” [23, 24].

Thus, infection outcomes used for research, even those that are driven by protocolized surveillance definitions, can be highly subjective. This is not a criticism of the surveillance definitions per se, because the subjectivity was originally designed to allow trained assessors in routine surveillance settings to use clinical judgment to potentially improve the specificity of their determination. Nevertheless, in research settings in which an infection control intervention is being evaluated, unblinded assessment of infection outcomes raises the danger of assessment bias.

**EXAMPLES OF STUDIES WITH POTENTIAL ASSESSMENT BIAS**

Given the subjectivity of infection surveillance definitions, it would be reassuring if publications of infection control research routinely utilized some kind of protection against assessment bias. Unfortunately, this is not the case. We highlight 2 papers as examples of studies that, although using the prevailing methods for outcome assessment, likely contain assessment bias.

Investigators examined the ability of a bundled group of prevention interventions to decrease the rate of VAP among 112 ICUs in Michigan [25]. The primary outcome was the NHSN-defined VAP, which was assessed by infection preventionists as part of their usual surveillance. Although the infection preventionists were described as “independent” of the project, it is unreasonable to expect that they were blinded to the intervention itself, which was high profile and included other hospital-wide safety and communication interventions. In fact, the study design specifically partnered infection preventionists with local ICUs, and the infection preventionists regularly fed back VAP numbers and rates as a critical part of the intervention. Thus, the outcome assessor (infection preventionist) was unblinded and an active participant in the intervention! The assessed treatment effect in this intervention was biased toward efficacy, and it is impossible given the data presented to separate bias from intervention effect.

Another study examined the effectiveness of an intervention bundle to decrease methicillin-resistant *Staphylococcus aureus* (MRSA) healthcare-associated infections as a quality improvement project [26]. The “MRSA bundle” included universal nasal surveillance for MRSA, contact precautions of MRSA-colonized or -infected patients, hand hygiene promotion, and a change in institutional culture. The outcome of the intervention was the “prevalence of MRSA colonization or infection,” which was composed of 4 NHSN-defined MRSA infection outcomes: pneumonia, bloodstream infection, urinary tract infection, and skin and soft tissue infection. A subset of participating hospitals also performed surveillance on healthcare-associated vancomycin-resistant *Enterococcus* and *C. difficile* infections. Assessment of the infections was performed by “a physician or other professional in infection prevention and control” who reviewed the patient’s record to

---

**Table 1. Subjective Elements Associated With Surveillance Definitions of Healthcare-Associated Infections That May Be Susceptible to Assessment Bias**

<table>
<thead>
<tr>
<th>Infection</th>
<th>Examples of Subjective Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catheter-associated urinary tract infection</td>
<td>“At least 1 of the following signs or symptoms with no other recognized cause: fever, suprapubic tenderness, or costovertebral angle pain or tenderness”</td>
</tr>
<tr>
<td>Central line–associated bloodstream infection</td>
<td>“Recognized pathogen cultured from 1 or more blood cultures and organism cultured from blood is not related to an infection at another site”</td>
</tr>
<tr>
<td><em>Clostridium difficile</em> infection (gastroenteritis)</td>
<td>“Acute onset of diarrhea (liquid stools for more than 12 h)” “No likely noninfectious cause”</td>
</tr>
<tr>
<td>Skin and soft tissue infection</td>
<td>“Pain or tenderness, localized swelling, redness, or heat” “...purulent drainage...” “...fever or localized pain...”</td>
</tr>
<tr>
<td>Surgical site infection</td>
<td>“Diagnosis … of an SSI by a surgeon or attending physician” “New or progressive and persistent infiltrate”</td>
</tr>
<tr>
<td>Ventilator-associated pneumonia</td>
<td>“New onset of purulent sputum, or change in character of sputum, or increased respiratory secretions, or increased suctioning requirements” “New-onset or worsening cough, or dyspnea, or tachypnea”</td>
</tr>
</tbody>
</table>

Definitions are from the Centers for Disease Control and Prevention’s National Healthcare Safety Network [7]. Similar definitions are found among other surveillance networks [9].
Abbreviation: SSI, surgical site infection.
determine “whether the criteria for infection had been met.” The assessors were not blinded to the MRSA bundle; rather, the assessors (infection preventionists and hospital epidemiologists) were systematically involved with promoting the bundle and were part of the widely publicized institutional culture change. Thus, the study outcome was biased toward lower MRSA infection rates, and it is conceivable that assessment bias also contributed to greater-than-expected declines in vancomycin-resistant Enterococcus (VRE) and C. difficile infection rates, which were voluntarily reported.

In the examples above, it is certainly possible that the interventions being tested are truly effective. However, given the lack of protection against assessment bias, it is likely that the intervention’s degree of effectiveness is overestimated. At worst, the assessment bias was strong enough to make an ineffective intervention appear falsely effective.

**MANAGEMENT OF ASSESSMENT BIAS**

Several strategies can be used to manage assessment bias in infection control studies that rely on infection as an outcome (Table 2). The major strategies are blinding the assessor and using objective outcome measures.

**Blinding**

If a subjective outcome is used for a study, such as NHSN surveillance-defined infections, then the ideal defense against assessment bias is to blind the assessor to the allocation of the intervention. This method can be feasibly performed in retrospective studies of historical infection rates, if the assessor (infection preventionist) is unaware of a study or intervention occurring. For example, in a study comparing prospectively determined infection preventionist CLABSI rates with computer algorithm CLABSI rates determined retrospectively, the infection preventionists could be blinded to the study protocol [11]. Blinding becomes more difficult for prospective infection control interventions that are designed to alter infection control practice, because infection preventionists who perform outcome assessment are usually closely involved with instituting infection control interventions as part of their duties. Blinded assessment by infection preventionists is feasible in situations where there is randomization of the intervention and masking of allocation. For example, in a study of the efficacy of stop orders to reduce catheter-associated urinary tract infection as one of several endpoints, patients were randomized to stop orders or usual care. The assessors in the study were explicitly blinded to the intervention assignment and found that there was no difference in infection outcome between the 2 groups [27].

**Objective Infection Measures**

Another general approach to limit assessment bias is to use objective outcomes that are less susceptible to bias. Commonly available objective outcomes include mortality, length of stay, antimicrobial use, or incidence/prevalence of pathogen-specific colonization. For example, investigators of decontamination of the digestive tract and oropharynx in ICU patients chose 28-day mortality as their primary endpoint instead of VAP [28], because they recognized the subjectivity of the pneumonia outcome. Further, the investigators recognized that even in-hospital mortality could be biased by physicians who, knowing treatment allocation, could influence discharge decisions among patients in one intervention group versus another; thus, a 28-day mortality outcome that included out-of-hospital events was selected.

For investigators who are interested in objective measures of infection, several options are available. For example, NHSN has developed objective surveillance definitions that only rely on clinical culture results obtained from the laboratory (“laboratory-identified events”) [29]. Laboratory-identified events can be reported to NHSN for organisms such as methicillin-resistant and methicillin-susceptible Staphylococcus aureus, VRE, carbapenem-resistant Klebsiella spp., and C. difficile, and the surveillance rules can be adapted to any bacteria of interest. The laboratory-identified event is designed to be determined either by computers or by humans.

A simpler and practical form of a laboratory-identified event would be to use the presence of nosocomial clinical cultures identified by the microbiology laboratory for pathogens and culture sites of interest. Such laboratory-identified culture-based methods are potentially scalable; a single hospital, or a large number of hospitals, can obtain such an outcome as long as microbiology data are accessible. For example, a cluster randomized trial of 3 different interventions to reduce MRSA disease in hospital ICUs among 45 hospitals utilized a primary outcome of “number of ICU patients who have MRSA-positive clinical cultures occurring at least 2 days after ICU admission through 2 days after ICU discharge” [30].

**Table 2. Study Design Strategies to Limit Assessment Bias in Infection Control Studies**

<table>
<thead>
<tr>
<th>Strategy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Blind the assessor to the allocation of the study intervention</td>
<td></td>
</tr>
<tr>
<td>Use objective outcomes</td>
<td></td>
</tr>
<tr>
<td>Mortality, length of stay, antibiotic use, incidence/prevalence of pathogen colonization</td>
<td></td>
</tr>
<tr>
<td>Laboratory-identified infection event</td>
<td></td>
</tr>
<tr>
<td>Positive clinical cultures</td>
<td></td>
</tr>
<tr>
<td>Computer algorithm-defined infections</td>
<td></td>
</tr>
<tr>
<td>Miscellaneous strategies</td>
<td></td>
</tr>
<tr>
<td>Multiple assessors with consensus</td>
<td></td>
</tr>
<tr>
<td>Assessors external to the study</td>
<td></td>
</tr>
</tbody>
</table>

**HEALTHCARE EPIDEMIOLOGY • CID 2012:54 (1 May) • 1345**
Other objective measures of infection include the use of automated computer systems to perform infection surveillance [31]. Automated surrogate measures of infection have been developed for CLABSI [14, 32], catheter-associated urinary tract infection [33, 34], surgical site infection [35, 36], and VAP [37]. Commercially available automated measures of infection have also been developed [34, 38]. These infection surveillance measures can be used either as primary or secondary outcomes; in particular, the automated measures can be used to confirm findings of another infection measure. Investigators studying the use of chlorhexidine bathing among medical ICU patients used CLABSI as a primary outcome; because of the possibility of incomplete blinding, they used a computer algorithm for CLABSI determination to validate their primary outcome, demonstrating that there was no significant bias [39].

Miscellaneous Measures
Other miscellaneous methods can be used to combat assessment bias. A study can use multiple observers for infection determination (eg, use 2 different assessors plus an adjudicator in cases of disagreement). This consensus strategy is resource intensive and may lead to more conservative rates compared with single-observer surveillance, but it has been used to increase confidence in infection-related outcomes such as CLABSI [39] and VAP [40]. Lastly, with remote computer access available for electronic medical records, it is feasible for external assessors from a nonstudy location who are not intellectually invested in the study to perform chart reviews for infection assessment.

Caveats
Measures designed to minimize assessment bias still require study designs that protect against other types of bias. Any measure that relies on microbiological cultures can be susceptible to surveillance bias (ie, prior knowledge of an intervention can affect how intensively clinicians search for infection) [41]. For example, physicians who are unblinded to an intervention to decrease urinary tract infections may order urine cultures less frequently among one group of patients compared with another in response to fever, leading to a potentially false difference in culture-based infection rate. If masking of the intervention is not feasible, then standardizing microbiologic culturing practice may be advisable. Additionally, misclassification bias can occur if clinical cultures are obtained after (rather than before) the initiation of new antibiotics at the time that an infection is suspected, increasing the rate of falsely negative cultures. Investigators should be aware of secular changes in antimicrobial prescribing practice (such as initiatives for early antimicrobial therapy in response to suspected sepsis) that may differentially bias the ability of cultures to detect true infection.

CONCLUSION
Investigators, practitioners, and policymakers seek research that is as free from bias as possible. However, blinded assessments, which are a requisite part of modern study design to combat bias, are not routinely found in the infection control literature. This anomaly might be explained by factors that are unique to infection control research. It is challenging to mask group-level interventions in quasi-experimental studies, especially during infection outbreaks. Furthermore, infection control studies are often quality improvement studies that lack financial support for rigorous outcome measurement; as such, they often rely on surveillance infection definitions that have provided a false sense of objectivity. Importantly, infection control studies increasingly use grouped (“bundled”) interventions rather than single interventions, making masked allocation even more difficult and raising the risk of bias. Regardless of the type of intervention studied, awareness of modern study design tools to mitigate bias will allow the infection control community to raise the standard of research quality to a higher level.

Notes
Acknowledgments. We thank William E. Trick for his thoughtful review of the manuscript.
Financial support. This work was supported by a Centers for Disease Control and Prevention Prevention Epicenters grant (1U54CK000161-01 to M. Y. L) and the Netherlands Organization for Scientific Research (VICI NWO grant 918.76.611 to M. J. B).
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


19. Klompas M. Interobserver variability in ventilator-associated pneu-


