

# What Counts? An Ethnographic Study of Infection Data Reported to a Patient Safety Program

MARY DIXON-WOODS,<sup>1</sup> MYLES LESLIE,<sup>2</sup> JULIAN BION,<sup>3</sup>  
AND CAROLYN TARRANT<sup>1</sup>

<sup>1</sup>*University of Leicester;* <sup>2</sup>*Johns Hopkins University School of Medicine;*

<sup>3</sup>*University of Birmingham, College of Medical & Dental Sciences*

**Context:** Performance measures are increasingly widely used in health care and have an important role in quality. However, field studies of what organizations are doing when they collect and report performance measures are rare. An opportunity for such a study was presented by a patient safety program requiring intensive care units (ICUs) in England to submit monthly data on central venous catheter bloodstream infections (CVC-BSIs).

**Methods:** We conducted an ethnographic study involving ~855 hours of observational fieldwork and 93 interviews in 17 ICUs plus 29 telephone interviews.

**Findings:** Variability was evident within and between ICUs in how they applied inclusion and exclusion criteria for the program, the data collection systems they established, practices in sending blood samples for analysis, microbiological support and laboratory techniques, and procedures for collecting and compiling data on possible infections. Those making decisions about what to report were not making decisions about the same things, nor were they making decisions in the same way. Rather than providing objective and clear criteria, the definitions for classifying infections used were seen as subjective, messy, and admitting the possibility of unfairness. Reported infection rates reflected localized interpretations rather than a standardized dataset across all ICUs. Variability arose not because of wily workers deliberately concealing, obscuring, or

---

*Address correspondence to:* Mary Dixon-Woods, Department of Health Sciences, University of Leicester, Adrian Building, Leicester LE1 7RH, United Kingdom (email: md11@le.ac.uk).

deceiving but because counting was as much a social practice as a technical practice.

**Conclusions:** Rather than objective measures of incidence, differences in reported infection rates may reflect, at least to some extent, underlying social practices in data collection and reporting and variations in clinical practice. The variability we identified was largely artless rather than artful: currently dominant assumptions of gaming as responses to performance measures do not properly account for how categories and classifications operate in the pragmatic conduct of health care. These findings have important implications for assumptions about what can be achieved in infection reduction and quality improvement strategies.

**Keywords:** Patient safety, infection control, intensive care units, qualitative research, implementation science.

**M**EASURES OF PERFORMANCE ARE AN INCREASINGLY prominent feature of health care systems. Powerful arguments for setting explicit standards and assessing performance against those standards lie in the potential for enhancing accountability and transparency, exposing practice to critical scrutiny, identifying variations in quality of care, and creating opportunities for remedial action (Spertus et al. 2005). Targets accompanied by monitoring may act as missions around which organizations can mobilize and cohere (Besley and Ghatak 2005). They can also allow the activation of a range of behavioral mechanisms: giving people goals may motivate better performance; setting a target can signal the priority to be given to an activity; and providing feedback may promote learning (Kelman and Friedman 2009). In practice, targets and standards for performance typically operate as organizational incentives to which people respond in characteristic ways, and designing incentives so that their overall effects are optimal has often proved difficult (Prendergast 1999).

The behavioral impacts of performance measures are widely discussed in the economics, political science, and public administration literatures (Bevan and Hamblin 2009; Kelman and Friedman 2009; Propper and Wilson 2003). Much of this work is focused on identifying possible dysfunctional reactions to performance measurement (Mannion and Braithwaite 2012). Writing from an economics perspective, Kelman and

Friedman (2009) divide possible distortions associated with performance measurement into two heuristic types: effort substitution and gaming. *Effort substitution* refers to the problem that when people's efforts are judged according to a specific standard, they tend to direct their attention to what is being measured, to the possible exclusion of other activities (Holmstrom and Milgrom 1991). For example, a target for completing risk assessment forms for venous thrombo-embolism might direct attention toward completing the forms and away from tasks not monitored. This effect in a health care context might usefully be dubbed *effort redirection* and is the term we use in this article.

*Gaming* describes activity that produces an apparent change in the measure but no actual change in the underlying performance (Kelman and Friedman 2009). It thus involves some manipulation of accounting practices. Gaming may sometimes entail cheating or falsifying data, but much more commonly it involves adjusting or modifying terms or their interpretation in order to suggest improvement. For example, changing the classification of an "emergency call requiring urgent response" may enable ambulance services to show an improvement in response times to emergency calls, even though no real change has occurred (Bevan and Hamblin 2009).

As Kelman and Friedman point out, scholarly writing on performance measurement has been concerned with these kinds of dysfunctional effects since the 1950s. Empirical studies of performance measures have repeatedly sought to detect, using statistical analyses, various kinds of apparently anomalous behavior and have then assessed the evidence for or against effort redirection and gaming. One consequence is that a picture has emerged of artful workers, sometimes seized with conspiratorial intent, who find ways of evading the spotlight or outwitting managerial or external control. "An important lesson is that incentive designers should beware that performance measures elicit dysfunctional and unintended responses because line workers acquire in their daily routine a superior understanding of how the measurement systems work, and how performance outcomes can be manipulated" (Courty and Marschke 2003, 269).

These behaviors are usually inferred rather than directly observed. We know relatively little about how health care organizations set up data collection systems, how they collect data, or how they interpret definitions of data for purposes of reporting. Field studies of the production of numbers specifically for performance measures have remained

rare, despite calls for more such work (Bird et al. 2005), particularly in health care. The need for better understanding how performance measures function in high-stakes areas such as patient safety is now pressing (Pronovost, Miller, and Wachter 2007). In this article, we focus on the production of one widely used measure of patient safety: rates of central venous catheter bloodstream infections.

### Central-Line Infections as a Patient Safety Performance Measure

In recent years, health care–acquired infections have emerged as a high-visibility, high-consequence performance measure for health systems worldwide, with central venous catheter bloodstream infections (CVC-BSIs) a key focus of interest. CVCs—often known as *central lines*—are narrow tubes inserted into large veins, with the tip lying close to the heart. They are most frequently used in intensive care units (ICUs) in the care of seriously ill patients. Some patients may have several CVCs. Although fungal and bacterial pathogens introduced into the bloodstream by these devices may cause infections that can result in death, serious complications, prolonged hospitalization, and increased cost, research evidence strongly indicates that many CVC-BSIs can be avoided (Frasca, Dahyot-Fizelier, and Mimoz 2010). The Michigan-Keystone project in particular attracted worldwide attention for its demonstration, in a *New England Journal of Medicine* article published in 2006, of a dramatic reduction in rates of CVC-BSIs. This followed implementation of an extensive program of systemic and cultural change to support compliance with specific evidence-based standards of practice (Dixon-Woods et al. 2011; Pronovost et al. 2006).

CVC-BSIs have now become a high-stakes metric in many health systems. In the United States, they are increasingly tied to consequential sanctions and are now a core measure of quality used by the Joint Commission, the U.S. accreditation agency. In 2008, CVC-BSIs were also deemed a Medicare “never event,” meaning that they should never occur and that organizations cannot claim reimbursement for any costs associated with their treatment. Since 2011, CVC-BSI rates have been used in calculating payments to hospitals by the Center for Medicare and Medicaid Services (CMS), and since 2012 they have been made publicly available on the *Hospital Compare* website. Beginning in 2013,

hospitals are required to submit data to the benchmarked reporting system of the U.S. Center for Disease Control and Prevention's (CDC) National Healthcare Safety Network (NHSN) or face a reduction in their annual Medicare inpatient payment update (Magill and Fridkin 2012). In simple terms, such policy moves have two important implications. First, they assume that the rates reported by different organizations are valid, reliable, and directly comparable. Second, they assume that the CVC-BSI rate is a direct outcome of the quality of care provided and that there is a near-perfect relationship between the rates of these infections and staff practices and behavior. Any rate above zero is then taken as evidence of a failure of care that can reasonably be subjected to financial, reputational, or other sanctions.

The research evidence, however, does not support such a strongly presumed relationship between observed outcomes and underlying actions. Although the Michigan-Keystone study is often read as meaning that all CVC-BSIs can, in principle, be prevented, this reading does not survive examination of the reported data. The study reported that the median infection CVC-BSI rate in participating ICUs fell from a baseline rate of 2.7 infections per 1,000 catheter days to a median of 0 and that the baseline mean fell from 7.7 per 1,000 catheter days to 1.4. However, considerable heterogeneity between participating ICUs was evident (Davidoff 2009). The median interquartile range was 0–2.4 BSIs per 1,000 catheter days by the end of the project, suggesting that some ICUs were still reporting relatively high rates of infection. Around half the ICUs were unable to control infections completely in the period after implementing the interventions. Why some ICUs were unable to report that they had eliminated infections, even in an otherwise highly successful project, is unknown. Among the possibilities are less effective implementation of the interventions, aspects of case mix that meant some settings were less tractable to the interventions, or measurement error.

Joining this ambiguity about whether a zero rate of CVC-BSI rate actually can be attained in all settings is an emerging body of evidence indicating substantial variability in how the definitions of these rates are interpreted and applied. The CDC's definitions are the most widely used (O'Grady *et al.* 2011) and form the basis of the CMS and NHSN reporting. Substantial inconsistencies in reported CVC-BSI rates when using these definitions have now been found in both adult ICUs (Aswani *et al.* 2011; Lin *et al.* 2010) and pediatric ICUs (Niedner and 2008

National Association 2010). One study of thirty adult and three pediatric ICUs over a thirty-day period estimated that 52 percent of infections that met the criteria had not been reported to the NHSN, but it also identified some overreporting (Backman, Melchreit, and Rodriguez 2010). The authors of this study estimated that the infection rate found by their independent reviewers was 78 percent higher than that reported to the NHSN.

What is not clear is what causes such variability in CVC-BSI reporting. Current scholarly work on performance measures, as discussed earlier, would propose a role for effort redirection and/or gaming and would see variations in reporting as attributable to underlying variations in these behaviors. It is also possible, of course, that variations in the reporting of infections should not be explained as effort redirection or gaming, but in some other way. Without studies of what organizations are actually doing when they collect and report data on performance measures, however, these other explanations remain elusive.

An important opportunity to examine how data on CVC-BSIs are generated was presented by Matching Michigan, a program led by the United Kingdom's National Patient Safety Agency over a two-year period (April 2009–2011). Matching Michigan drew explicitly and directly on the original Michigan-Keystone project. As did the original, it sought to reduce rates of CVC-BSIs in ICUs using a program with a technical component, a nontechnical component, and a data collection component.

1. The *technical* component aimed to ensure consistent use of five evidence-based interventions known to reduce the risks of CVC-BSIs: observing appropriate hand hygiene, using chlorhexidine to prepare the patient's skin, using full-barrier precautions during CVC insertion, avoiding the femoral (groin) route, and removing unnecessary CVCs. The program sought to improve the reliable delivery of these elements of care by encouraging ICUs to use dedicated carts and a checklist for CVC insertion, providing information on the evidence base, and reinforcing adherence to preexisting guidelines.
2. The *nontechnical* component sought to intervene in cultures and systems, including the use of a cultural survey, education on the science of safety, monthly learning from defects, and partnering between units and executives.

3. The *data collection* component involved a centralized online system to which data on CVC-BSIs were reported by ICUs. The data were then fed back to ICUs so that they could identify their own rate of infections and also could see the anonymized rates for other ICUs. Before submitting the data, the centers were invited to attend a one-day training event on data collection, which included education on the techniques and definitions used to identify CVC-BSIs and the data required to determine exposure and infection rates.

Although the technical interventions to reduce CVC-BSIs had been widely promoted and were made part of the UK Department of Health's policy from 2007 onward, Matching Michigan offered the first formal surveillance of rates of these infections in England. Like the Michigan-Keystone project, Matching Michigan based its data definitions on those used by the CDC and was similarly promoted as a patient safety improvement initiative, not as a target-led program tied to sanctions (financial or otherwise). While it set a goal of minimizing the mean rate of CVC-BSI in adult and pediatric ICUs in England to at least the mean level in Michigan (1.4 per 1,000 catheter days), no formal target was set for individual organizations. It was not intended that organization-level data would be made public.

In this article, we describe how the ICUs participating in Matching Michigan collected and reported their data on CVC-BSIs, using data we collected as part of a wider ethnographic study of interventions to reduce CVC-BSIs in ICUs. We seek to provide an empirical account of what happened when ICUs were asked to set up and use data collection and reporting systems, and to examine the consequences for the rates reported to the program, including the comparability of rates across different organizations. In particular, we aim to assess the adequacy of effort redirection and gaming as explanations for what was reported to the program by the participating units.

## Methods

Given the nature of our research questions, ethnographic methods using observations, interviews, and documentary analysis were especially appropriate (Dixon-Woods and Bosk 2010). Ethnography enables

detailed, contextualized descriptions of behavior and of how people make sense of the situations in which they live and work and, consequently, why their own actions make sense (Hammersley and Atkinson 2007). Ethnography is an especially useful approach to studying patient safety, as it provides an opportunity to observe firsthand how events are classified and communicated in particular ways, as well as the social, cultural, and organizational influences on such classificatory work (Bosk 2003; Waring 2009). In particular, ethnography enables insights into how professionals in health care settings account for patient safety issues (Bosk 2003) in ways that other methods may not detect.

We used a purposive strategy to generate a diverse sample from the population of adult ICUs in England, including those in different locations (north, south, east, west, Midlands, and London, urban/rural), size, and types of patients served (specialty/general). Of the 196 adult ICUs in England that participated in Matching Michigan, we recruited seventeen for inclusion in the study reported here. Research ethics committee approval was obtained.

Ethnographic fieldwork totaling ~855 hours, averaging around 48 hours per ICU, was conducted across the seventeen ICUs by Carolyn Tarrant and Myles Leslie. They observed care on the units, including, but not confined to, CVC insertion. They also conducted face-to-face interviews with ninety-three nurses and doctors of varying grades in the ICUs and also, when possible, with microbiology staff. In one ICU, only one interview and only a few hours of observation were conducted. In addition to this on-site data collection, twenty-nine telephone interviews were conducted by Janet Willars (see the acknowledgments) with staff who had attended Matching Michigan training events. These interviews covered a wide range of hospitals and grades of staff, such as senior managers and executives, and thus provided access to settings beyond the seventeen ICUs in the ethnographic study. In addition, we attended all training events and a selection of the program's team and external reference group meetings. We obtained signed consent from all those who took part in an interview. Interviews were recorded using a digital recorder, transcribed, and anonymized. The data excerpts given to illustrate our analysis are labeled with a number for each participant. We have also indicated those individuals who had particular local roles in the program—such as the local Matching Michigan medical lead or nursing lead for their unit—along with their (numbered) participant identifier.

We based our analysis of the qualitative data on the constant comparative method (Glaser and Strauss 1967). The research team initially generated open codes based on transcripts and fieldwork notes, which were then grouped into higher-order organizing themes. We also used some sensitizing concepts (Charmaz 2006), including the literature on performance measurement described earlier. Although these sensitizing concepts suggested directions for where to look in our data, they were used as points of departure for analysis rather than as a rigidly applied set of constructs. Our analysis was recursive, constantly moving from the specific, at the level of individual interviews or observations in particular settings, to the more general, with the aim of producing more generalizable categories and explanations for our findings. This enabled us to identify commonalities and patterns across the large number of settings in which we conducted our study. We actively sought disconfirming cases to enable us to check our emerging constructs. Myles Leslie coded the transcribed field notes and interviews using NVivo software, with Carolyn Tarrant and Mary Dixon-Woods checking the coding and interpretation.

## Findings

Matching Michigan organized ICUs in England into four groups, which joined the program as successive cohorts staged over time. Each cohort except the last one was based on geographical location, and all the ICUs belonged to organizations providing acute care, known in the NHS as *acute hospital trusts*. These organizations could be home to one or more ICUs, distributed across one or more hospitals. All participating ICUs were invited to two training sessions, one on data collection and one on program interventions. In addition, each participating organization was asked to appoint a dedicated team for the program, made up of a Matching Michigan medical lead, a nursing lead, and an executive lead. The medical lead was a consultant, that is, a senior specialist physician equivalent to an attending physician in a U.S. hospital. In each organization, this team was responsible for organizing the implementation of the technical and nontechnical interventions and for submitting monthly data on CVCs and CVC-BSIs. For all these individuals, these tasks were on top of their existing work.

## Responses to the Program and Effort Redirection

Response to the program varied. The staff of many ICUs reported initiative fatigue, and many questioned the need for a new program targeting CVC-BSIs, arguing that the technical interventions recommended by the program were already being implemented. Our direct observations largely supported these claims. Practices were generally (though not completely consistently) compliant with the technical interventions recommended by the program; mostly, hand hygiene was good, full-barrier drapes were used, patients' skin was prepared with chlorhexidine, and the femoral site was avoided. On ward rounds and during routine care, patients were generally carefully monitored and assessed for any signs of infection. Our interviews suggested that on most ICUs, Matching Michigan had relatively little role in promoting this level of compliance; rather, the explanation lay in a much longer history of improving practice in response to previous policy pushes and the emergence of scientific consensus on how to avoid CVC-BSIs.

I can't, off the top of my head, think of any major difference to the best practice we should have been observing beforehand compared to what we're doing with Matching Michigan. (087: consultant)

We found evidence on four ICUs that the implementation of the nontechnical interventions did increase in response to the program. On one ICU implementation of the interventions was already so high that there was little room for improvement. For twelve ICUs, however, the most salient feature of the program was not so much its technical and nontechnical interventions as its introduction of a requirement to report to an external agency their rates of CVC-BSIs.

Responses to this requirement for reporting varied. Although few ICUs knew their rates of CVC-BSIs before the program began, many were convinced, because they had seen or heard of very few during routine clinical practice, that their local rates were low. Four units used the program's requirement to introduce formal data collection as an opportunity to learn more about how well they were doing and to improve their practice. But a more common response was apathy: the program was seen as having little potential to make a difference, given

the gains that had already been made. This was perhaps reinforced by the finding that at the beginning of the program, 65 percent of all units in the program (not just those in our ethnographic study) were reporting zero infections in any one month, suggesting that CVC-BSIs were rare. Many individuals saw the program primarily as an externally imposed performance management audit that was tinged with threat. The reasons for this lay in the perception that any government-led program collecting data on performance should properly be regarded with suspicion and discomfort.

My worry about having a national [audit]—and this is where my cynicism comes in—is people are always concerned that the data will be used against them. (006: nurse)

The historical context was important to the pervasiveness of this view. In the years immediately preceding the program's launch in 2009, performance management was intensively used in the English NHS. Indeed, performance targets were attached so often to severe punitive sanctions and damaged reputations for organizations that the regime was described as one of "targets and terror" (Bevan and Hood 2006, 517). Infection control was no exception. Beginning in 2004, an intense, target-led government effort to reduce health care-acquired infections was followed in 2006 onward by new regulatory measures that made the control of infection a legal responsibility of NHS care providers. An inspectorate role was given to the Care Quality Commission, a regulatory agency with the power to remove accreditation from providers found to be in breach and to apply other penalties.

ICU staff had thus learned that infection data could be used both internally and externally to punish hospitals, and they worried that Matching Michigan was the thin end of a "targets and terror" wedge. Although the program team repeatedly emphasized that none of these effects was intended, people's previous experiences of tough sanctions for hospitals that failed to meet externally imposed accountability requirements made such appeals largely unconvincing.

People are extremely sensitive here about sharing data outside the [organization]. [The nurse] says, "It's a shame: the whole point of collecting this data is so that it can be used for shared learning."

But it seems that “everything you tell people is used for performance management.” . . . This is threatening, and “we miss opportunities to share and learn.” (fieldwork notes)

Despite these fears about the possible uses of the data collected by the program, we found little evidence that staff engaged in any of the activities that might be hypothesized to occur if effort redirection were a feature of their behavior. We found no evidence that people were trying to prevent CVC-BSIs to the exclusion or detriment of other activities, nor were they engaging in effort redirection in regard to use of CVCs. Our observations found that CVCs were inserted when physicians believed there was clinical need and were generally removed only when that need had expired or there was suspicion of infection. Our findings therefore suggest (but do not prove) that effort redirection was not a significant impact of the program.

### Counting for the Program

The ICUs were required to submit monthly data to the program’s online reporting system. Two-thirds (147) of all the ICUs participating in the program submitted a full data return for every possible month, but we found considerable variability in what the seventeen ICUs in our ethnographic study counted as reportable to the program. This variability could not be adequately explained as a simple matter of gaming, however. To understand why requires, first, some technical background; second, a description of how the ICUs did the counting; and third, an understanding of the multiple kinds of work involved in counting.

### *Calculating a Rate*

In both the scientific literature and surveillance programs, CVC-BSI rates are conventionally reported as the monthly number of bloodstream infections per 1,000 CVC patient days. An infection rate is calculated by dividing the number of bloodstream infections (the numerator) by the number of CVC patient days (the denominator) and multiplying by 1,000. For example, an ICU with three CVC-BSIs and 625 CVC patient days in that month would have a rate of  $3/625 \times 1,000 = 4.8$  BSIs

per 1,000 patient CVC days. To comply with the program's data requirements for calculating these rates, each ICU needed to find a way to

- Count all patients with CVCs in situ, and also to count how many CVCs each patient had, every day at the same time.
- Collect data on infections that were possible candidates for reporting to the program.
- Compile microbiology test results linked (locally) to patients' records and to determine definitively which candidate infections satisfied the program's definitions.
- Submit the infection data two weeks after the end of each month.

### *How ICUs Operationalized the Task of Counting*

Only a few ICUs had established systems for counting their CVC-BSIs before Matching Michigan. The program therefore required that the ICUs organize to make themselves auditable. The program did not prescribe how this should be done, instead allowing ICUs to develop locally appropriate systems. We found three different types of data collection systems in operation across the seventeen ICUs, which we characterized as Track-Trigger-Track (TTT), Patrol, and Controller-Centered.

These systems had a profoundly mundane character involving material objects (e.g., laboratory tests, forms, and medical records) as well as an evaluative character requiring judgments about the likely source and significance of positive blood cultures. All three systems featured an individual, whom we called a *controller*, who was responsible for submitting his or her unit's data to the program's online database. Next we describe each system before showing how multiple sources of variability undermined the possibility of a standardized, comparable data set across the ICUs we studied.

*System 1: Track-Trigger-Track.* We identified a Track-Trigger-Track (TTT) system in operation in three of the seventeen ICUs we studied ethnographically; this system was also described in telephone interviews. One ICU in the ethnographic study had already introduced its system before Matching Michigan, as it had previously participated in a patient safety program that required a CVC-BSI surveillance system. The other two ICUs introduced TTT in response to Matching Michigan.

The defining characteristic of TTT was that it integrated the routine monitoring of patients' infection status and treatment with the generation of data for counting purposes. *Tracking* refers to capturing data for each patient on a dedicated form documenting how many CVCs that patient has had inserted and recording a number of data points, including any suspicion of infection. On two units, the forms were completed by a nurse at the same time each day. On the third unit, the data were collected by junior doctors as part of the ward round. In all three units, the data captured on the form integrated evidence from bedside records, direct observations of patients, and discussion with the staff directly involved in the patient's care. As part of their routine work, these staff continually checked and validated CVC data and suspicions about infection. They were given multiple prompts and reminders, such as sending blood samples to the microbiology laboratory for analysis.

At the end of each month, the forms produced as part of this routine monitoring were reviewed by the unit's controller. Any patient flagged as indicating a possible CVC-BSI *triggered* a full investigation by the controller, who *tracked* backward by reviewing the microbiological test results and the patient's clinical charts and consulting with clinical colleagues, including microbiologists. The controller then determined which, if any, of the program's definitions had been met and submitted data accordingly to the online system.

*System 2: Patrol.* The Patrol system, by contrast, removed from the ICU's staff the responsibility for counting both CVCs and infections. Two ICUs in our ethnographic study used this system, although one stopped collecting data after some months owing to a shortage of resources. A third ICU had planned to use this system but was unable to do so, again for resource reasons. In addition, some participants in the telephone interviews described systems that had Patrol features.

Under the Patrol system, infection control nurses from outside the ICU visited the ICU daily. They counted CVCs and identified possible CVC-BSIs using medical and nursing treatment charts and observation of patients, and they occasionally discussed cases with the unit staff. The patrolling nurses did not generally prompt for samples to be sent to microbiology, relying instead on the existing information in clinical records. At the end of each month, the controller reviewed both the data collected by the patrolling nurses and the microbiology test results for patients identified from the Patrol records as possibly having a CVC-BSI. The controller then made a final decision as to which, if any,

of the program's definitions had been satisfied. These decisions were usually made without direct input from the clinicians caring for the patients.

*System 3: Controller-Centered.* The Controller-Centered system was by far the most common system we observed, with eleven ICUs deploying it in several variants. Some participants in the telephone interviews also described its use. Counting of CVCs and infections was, unlike TTT and Patrol units, usually split into two separate tasks. Generally, ICU nurses were responsible for counting the number of patients with CVCs, although sometimes this task was assigned to a junior doctor. Typically, the CVC counter went around the ICU beds at a particular time of day using direct observation, review of treatment notes, and consultation with nursing staff to identify patients with CVCs. These data were transcribed onto a form and then given to the unit controllers at the end of the month.

Each unit's controller—usually the local Matching Michigan medical lead—was responsible for generating the infection counts, identifying which patients had tested positive for infections that might possibly be attributed to CVCs over the previous month, and then deciding whether a CVC-BSI should be reported to the program. How the controllers did this varied. In two of the eleven ICUs using this system, the controllers relied on their own memory, treatment chart analysis, occasional prompts from colleagues, and working through batches of microbiological test results to identify positive results and any evidence that the treating clinicians had perceived any infections to be CVC-BSIs. These units did not engage other members of staff in this process, so on these ICUs it was not uncommon to find staff unaware that CVC-BSIs were being counted and reported.

The other nine of the eleven controller-centered ICUs did use their clinical staff to find cases. On these units, the clinical staff were asked to identify patients who might be candidates for a CVC-BSI, record their suspicions on a form or otherwise notify the controller, and ensure a microbiology follow-up. The mechanics of this distributed system varied. On some units, when the staff found a patient with a likely CVC-BSI, they were expected to complete a special notification form and file it in a dedicated binder. This was supposed to ensure that the controller would have a comprehensive set of prompts for investigation regarding possible cases when reviewing the evidence at the end of the month. On other ICUs, the clinical staff were expected to notify the controller in person

or by email, or through a mixture of forms and personal notification. These notifications (by whatever route) were usually supplemented by the controller with review of records and microbiology reports to make a final decision on whether a given infection met the program definitions and to submit the data.

## Work Involved in Counting

The three systems faced different challenges in their operation, some specific to each system and others common to them all. What was clear across all systems was that the variability was not because of wily workers deliberately concealing, obscuring, or deceiving, but because counting for the program was as at least as much a social as a technical practice. Threats to the comparability of the data produced by the ICUs could be explained not by invoking a conspiracy to game or manipulate the data but by the logistical challenges of designing and operating reliable data collection systems; highly localized variability in underlying clinical practices; contestations about the legitimacy of both counting and methods of counting; and a widespread perception that the program's definitions of how to count both the denominator and the numerator, far from providing objective and clear criteria, were subjective and messy and admitted the possibility of unfairness.

### *The Denominator: Determining the Number of CVC Patient Days*

We begin by examining what happened when the ICUs were asked to report a census of the total number of patients and CVCs present on their ICUs each day. Our observations suggested that regardless of which system of data collection they used, most ICUs (15/17) counted daily. Two Controller-Centered units were less consistent in their counting and sometimes used medical records retrospectively to reconstruct which and how many patients had CVCs. However, even those ICUs that counted every day did not always count reliably or at the same time each day, nor did they always clearly determine which and how many CVCs patients had. For example, sometimes the patients' catheters were invisible under blankets; sometimes the patients were undergoing a procedure and could

not be observed when they counted; and sometimes the patients were moved to another ward before being counted.

Added to these mundane problems were more profound definitional issues about what was being counted. Both the research literature and surveillance programs often are vague or inconsistent in what they are counting as the denominator. Some studies count each *patient* with one or more CVCs as generating a CVC patient day, but others (e.g., Longmate et al. 2011) count each *CVC* (rather than each *patient*) as generating a CVC patient day. Matching Michigan asked ICUs to collect both denominators, but for reporting purposes deliberately decided to use the patient (not the catheter) as the unit of measurement, consistent with the CDC's guidelines (Mermel et al. 2009). Thus, for purposes of comparisons and public reporting, one patient was counted as generating one CVC patient day, regardless of how many CVCs he or she had.

Bowker and Star (1999, 6) noted that “for any individual, group or situation, classifications and standards give advantage or they give suffering . . . some regions benefit at the expense of others.” The ICUs in the program quickly recognized the consequences of the classificatory scheme of reporting a low rate of infections. Defining the patient rather than the CVC as the basis for the denominator meant that patients with potentially different risk profiles were mixed together, since patients with more than one CVC may be more susceptible to infection. This lumping together of all risk profiles implied by the definitions was perceived to obscure the ICUs' very different challenges in controlling infection in different kinds of patients.

Our participants saw this failure to recognize these differences as violating principles of fairness, which led them to various attempts to “even up.” Some units caring for what they saw as high-risk patients were concerned that they could be made to “look bad” if they were compared with other units whose case mix was less prone to infection. Some units therefore decided locally that those patients who were deemed unusual in some way should not be counted. For example, patients who had had particular kinds of surgery, had particular conditions, or had arrived from outside the unit's usual catchment were not consistently included by all ICUs in the daily CVC counts. Although our methods were unable to determine the number of these exclusions, we estimated that except for one unit caring for a particularly distinctive group of patients, it was small.

Those patients with only one CVC and only for a short time (such as those undergoing cardiac surgery), however, have lower risks of infection. The ICUs again recognized that *including* patients at very low risk of infection in their counts would give a more flattering rate of infection for their unit by increasing the size of the denominator. Perceptions of fairness again appeared to influence the ICUs' decisions about whether to count these patients, though not in a consistently predictable way. For example, some ICUs deliberately excluded cardiac patients from their counts because they felt it would give a misleading and overly favorable picture.

We decided not to count cardiac patients . . . so we don't flaw the data.  
(035: infection control nurse)

Other ICUs, having sought advice from the program's organizers, did include cardiac patients but noted that it was not "fair to compare."

I said [to the program organizers] "I'd be delighted to return data on [our cardiac] patients, but actually the incidence [of CVC-BSIs] is gonna be zero!" I don't have a problem with that. But [those numbers] will be compared to national averages and infection levels across the region, and [I asked,] "Was it really fair to compare?" It would make us look good but make [other units] feel pretty bad when actually our patient cohort was hugely different . . . actually it's dirty data. (011: infection control nurse)

Neither of these responses to the program's request that a census of patients be returned can be appropriately described as gaming; in different ways, they both were honest attempts to provide meaningful information. They illustrate the consequences of a perceived attempt to impose sameness even when at the sharp end of practice, difference abounded. The practical effect of these local interpretations was that what counted as the denominator was not standardized across the participating ICUs, but instead was based on highly localized decisions about what counted as an evaluable patient.

### *The Numerator: Determining Whether the CVC Was the Primary Source of Infection Organisms*

What counted as a central venous catheter bloodstream infection for purposes of reporting to the program was even more fraught. Determining whether a CVC is to blame for any bloodstream infection detected in a patient is not straightforward; it is not a matter of mechanically applying simple technical criteria. A positive blood culture taken from a CVC or from the tip of a CVC removed from the patient, for example, is not enough to conclude definitively that the CVC is the source of the infection. Microorganisms can travel through the bloodstream from a remote site of infection (e.g., the urinary tract) and then be sampled from an uninfected CVC, or they can lodge on and colonize a CVC. Catheter tips can also be contaminated when being removed, perhaps by picking up microorganisms on the skin. Thus, the CVC may be only one of a number of competing suspects to blame for an infection.

These uncertainties clearly pose problems when the task is to decide whether or not a CVC is the culprit, and when blaming a CVC is to suggest a potential deficiency of care. To guide such rulings, Matching Michigan, like many other surveillance programs and studies, used the CDC's definitions. An important but often neglected feature of these definitions is that they distinguish between catheter-related infections (CR-BSIs) and catheter-associated infections (CA-BSIs). The difference between a CR-BSI and a CA-BSI pertains to the standard of proof used to determine whether the CVC is the source of any infection detected. Again, some technical detail is necessary.

#### *Catheter-Related Bloodstream Infections*

The definition of a CR-BSI demands a higher standard of proof than the definition of a CA-BSI to rule that a CVC is culpable for any bloodstream infection detected. The determination of a CR-BSI maximizes *specificity*. It seeks to establish beyond a reasonable doubt that the CVC is the source of the infection. It requires *two* positive blood cultures (samples) indicating infection:

- One from the tip of the catheter (the part lying closest to the heart, cut off after the CVC has been withdrawn) *or* from a blood sample taken through the catheter, *and*

- A second separate positive culture drawn from a peripheral site, such as a blood vessel in the arm or the leg.

For an infection to be attributed to the CVC using this definition, both cultures must test positive for the same microorganism, determined using semiquantitative or quantitative techniques. These techniques refer to the microbiology laboratories' different kinds of analyses, which vary in cost, complexity, validity, and the laboratories' ability to undertake them. Hospital microbiology departments (in the United States, the United Kingdom, and elsewhere) are not universally able to perform semiquantitative analyses of blood, in which case the diagnosis of a CR-BSI can be made only by a special technique involving roll-tip cultures of the CVC tip. Obtaining the tip, however, requires removing the catheter, something clinicians may be reluctant to do if the patient still requires it.

### *Catheter-Associated Bloodstream Infections*

The definition of a CA-BSI has less rigid requirements for attributing the infection to the CVC and is often more feasible in clinical settings because it involves fewer resources than the definition of a CR-BSI. It requires only a single blood culture indicating infection in a patient with a CVC in situ, taken from peripheral blood or directly through the CVC or from the catheter tip following removal, together with the clinical judgment that no other source is responsible for the infection. A physician might consider, for example, whether a patient's condition improved after the CVC was removed, which might suggest that the CVC was a plausible (though not a definite) source of infection. The definition of a CA-BSI maximizes sensitivity, but at the cost of specificity. That is, it identifies most infections that originate in the CVC, but it increases the risk that a CVC will be blamed for an infection whose source is actually elsewhere (Sihler et al. 2010). Matching Michigan, somewhat unusually, required ICUs to specify which definition they used for any infection reported to the program (see box 1). It thus sought to address the pervasive tendency in both the research literature and practice to use the terms *catheter-related* and *catheter-associated* interchangeably (O'Grady et al. 2011).

Despite the training provided by the program, our interviews and observations found considerable variability within and across ICUs in

what they counted as a CVC-BSI for purposes of reporting to the program. We found nothing to suggest that this variability was linked to deliberate efforts to deceive. Rather, the formal program definitions were poorly aligned with the clinicians' intentions and priorities in their routine management of patients, and that variability arose because of inconsistencies across and within the units in

- Practices in sending blood samples for analysis.
- Microbiological support and laboratory analysis for submitted samples.
- Systems for collecting and compiling data and other supporting evidence of possible candidate infections.
- How controllers made decisions about qualifying infections.

## BOX 1

### Definitions Used in Matching Michigan

Laboratory-Confirmed Bloodstream Infection must meet at least one of the following two criteria:

- Patient has one or more recognized pathogens cultured from  $\geq 1$  blood culture, *or* if the microorganism is a common skin organism, then . . .
- It must have been cultured from two or more blood cultures drawn on separate occasions or from one blood culture in a patient in whom antimicrobial therapy has been started, *and*
- Patient has  $\geq 1$  of the following: fever of  $>38^{\circ}\text{C}$ , chills, or hypotension.

Catheter-Associated Bloodstream Infection (CABSI)

- One of the criteria for Laboratory-Confirmed Blood Stream Infection *and*
- The presence of one or more central venous catheters at the time of the blood culture or up to forty-eight hours following removal of the CVC *and*

- The signs, symptoms, and positive laboratory results, including pathogen cultured from the blood, are not primarily related to an infection at another site.

#### Catheter-Related Bloodstream Infection (CRBSI)

- One of the criteria for Laboratory-Confirmed Blood Stream Infection *and*
- The presence of one or more central venous catheters at the time of the blood culture or up to forty-eight hours following removal of the CVC *and*
- One of the following:
  - A positive semiquantitative ( $>15$  CFU/catheter segment) or quantitative ( $>10^3$  CFU /ml or  $>10^3$  CFU/catheter segment) culture in which the same organism (species and antibiogram) is isolated from blood sampled from the CVC, or from the catheter tip, and peripheral blood, *or*
  - Simultaneous quantitative blood cultures with a  $>5:1$  ratio CVC versus peripheral.

#### *Alignment between Clinical Practices and the Formal Definitions*

Our observations suggested that the quest to make a precise and definitive judgment about whether a CVC was to blame for an infection was poorly aligned with the routine behavior of clinicians confronted by a sick patient who might have a bloodstream infection. Although the physicians wanted to take action to help the patient, they were often faced with substantial uncertainty about the role of the CVC:

The problem is that sometimes when you get a patient that's spiking a temperature and their white cell [counts] are going off, it's obvious that they have an infection. . . . You can sometimes narrow that down to, let's say a blood-borne infection, as opposed to a chest or abdominal sepsis. You will send off, say, peripheral [blood cultures] or central line [blood cultures], and it may come back that it is a blood-borne infection of unknown origin. It may have come from the [patient's]

arterial line, let's say, or the central line, or one of any of the [peripheral lines]. So sometimes it's difficult to actually pinpoint that it was the CVC line that was the main culprit. (116: senior nurse)

Given the uncertainty, physicians faced a dilemma. They did not want to risk leaving an infected catheter in situ and increase the risk of sepsis. They also, however, wanted to avoid exposing patients to the risk from prematurely removing the catheter (and possibly the risk from inserting a new catheter) and unnecessary antibiotics. Yet they often needed to decide quickly, perhaps before the results of any blood tests had returned from the microbiology lab. Physicians thus often treated the patient "empirically" (to use their term) by removing the catheter and/or administering antibiotics, rather than delaying their decision while awaiting the microbiology results.

We cut the tip off [the suspected CVC] and send it to micro. . . . [By the time we] get a report [back from the microbiology lab], the patient [will] already have gone to the ward, so no one will really look at it. (073: junior doctor)

Clinical judgment was therefore constantly engaged in determining what to do in the event of a suspected CVC-BSI, and physicians' priority was finding a way of treating the patient rather than making an absolutely defensible decision about the source of any suspected infection for purposes of external reporting.

### *Variability in Sending Blood Samples for Microbiology Analysis*

The program's definitions relied on having at least one blood sample available. Our observations and interviews found that taking blood cultures or catheter tips in cases of suspected infection was not standardized across or within units. Their policies and practices regarding blood cultures varied depending on what local microbiology departments could support, the local unit's policy on what was appropriate, and the particular physician.

The least variability was found on the three units using the TTT system, where the routine care of patients and the generation of countable data on infections were fully integrated. On these units, systems for

tracking patients at risk of, or diagnosed with, infection were highly explicit and standardized, and they engaged the multidisciplinary team through multiple, real-time, inter- and intraprofessional checks. This degree of integration of clinical management and data generation was not found in the units using the Patrol or Controller-Centered systems, which constituted the majority (14/17) of those in our study. Both formal and actual practices for obtaining blood and other cultures on these ICUs varied widely. For many, the main reasons for sending off a blood sample and seeking a microbiology opinion were

1. To make a more definitive determination of whether an infection was caused by the CVC so that a more confident decision could be made on whether the CVC should be removed.
2. To identify which microorganism was implicated and so target appropriate antibiotic therapy.

The physicians also varied widely in their preferences for dispatching blood cultures and CVC tips to the lab, in part because they often could treat a patient with a suspected CVC-BSI without ever sending a blood sample to the microbiology lab. Patients might visibly improve in response to such therapy, and they were discussed during ward rounds as likely having a CVC infection. But without a blood culture, the criteria for the definition of either the catheter-associated or the catheter-related definition could not be met. Without any formal, recordable trace, these infections were not candidates for reporting to the program. To use Atkinson and Coffey's (2004, 56) phrase, they had no "documentary reality" and therefore did not produce recorded information that a controller could count when trying to identify reportable infections at the end of the month.

We also found considerable variability in what the staff submitted for microbiology analysis. Some submitted CVC tips, samples taken from the CVC, and a peripheral blood sample; others routinely submitted only one or two of these items. On some ICUs, all the CVC tips were supposed to be sent off for analysis following removal from the patient; others sent only the suspicious tips; and one unit had a policy of not routinely sending off tips on grounds that the benefits did not justify the costs. Units that submitted all tips could come up with a larger population of possible infections, certainly larger than would those ICUs that sent off tips only occasionally. The variability in what was submitted

to the microbiology lab appeared to be explained by local cultures and individual clinician preferences and the contingencies of particular cases, not by attempts to conceal possible infections.

Our observations suggested that losses of samples and information were not unusual. Even on ICUs with an “all tips” policy, we found considerable ambiguities and uncertainties over the existence of a rule and how it ought to be applied. For instance, on one “all tips” unit, we saw some tips being discarded without being sent for analysis because of nurses’ confusion over the requirements. When physicians asked during ward rounds for samples or tips to be sent off for analysis, they were not always taken or sent (they were forgotten, or the patient moved before they were taken); or they went missing or were not matched up with other samples from the same patient. Again, there was no evidence that such losses were the result of deliberate attempts to suppress information about possible infections; instead, they were linked to the realities of caring for patients in busy, stressed environments. The practical effect, nonetheless, was evident: that the data necessary to make a confident decision about whether a CVC was the likely source of an infection had been lost.

Getting all of those pieces of evidence together doesn’t necessarily happen. You might just have the tip culture on its own. You might have the tip culture but no blood culture. (090: consultant microbiologist)

### *Variability in Microbiology Laboratory Analysis and Support*

All the microbiology laboratories in our study were based in the same organizations as the ICUs. They varied greatly in what analyses they undertook and the extent to which they were prepared to support decisions about whether an infection could be attributed to a CVC. Generating data for the standard required to meet the catheter-related definition for the program was often not seen as a clinical priority, since patients could be treated successfully without making such a determination. In particular, many microbiology labs did not see quantitative or semiquantitative analyses, as required by the CR-BSI definition, as appropriate.

{It's} not something we do routinely. . . . We've taken the view that it's too time-consuming for us to do, and in {some cases when it's been requested} we didn't deem that it was clinically important to do it anyway. So quantitation—and even semiquantitation [*sic*]—isn't somewhere where we've really had any enthusiasm for going. (013: consultant microbiologist)

The role of microbiologists in identifying, and prompting the identification of, CVC-BSIs was also highly variable. In some ICUs, their role was confined to advising on the results of tests and correct antibiotics for patients, but otherwise they had little contact with the ICU clinical teams. On other ICUs, microbiologists were full members of the ward rounds, and some were active in prompting the ICU staff that a patient might qualify for inclusion in the program and in alerting the controllers accordingly.

We are simply the bearers of results, and part of our deal is that we would prod them to consider whether this is a patient who needs to be included in Matching Michigan or not. (077: consultant microbiologist)

We tell {the ICU consultants} about {positive blood cultures}, and they sort of assess the patient. I don't know what the final paperwork {for Matching Michigan} actually says. (112: consultant microbiologist)

### *Variability in Collecting and Compiling Data on Possible CVC-BSIs*

Finding a way of collecting information on possible CVC-BSIs that the controller could assess for purposes of reporting a rate was complicated. The controller needed both information on suspected BSIs and supporting evidence (such as microbiology reports and medical records) to enable a decision to be made on which infections could be reported to the program. The prospect of setting up and running a system of data collection that could serve this purpose was not a welcome one for most staff, especially those in middle management positions, as it was hard to find the necessary time, energy, and personnel, given their many competing demands.

Resistance does come in as well when things start to get bigger and bigger, and nothing drops off [your list of things to do.] We've run into quite a lot of negativity of late, about that, because there are just so many things that need auditing. (109: senior nurse)

The three systems we identified differed in what they provided to the controller and how easy they made it for the controller to compile relevant information. These variations held even when ICUs were using electronic health records. Often the electronic record-keeping systems ran alongside other manual systems, including nursing notes containing important clinical observations, such as temperature and wound appearance, and microbiology reporting systems that were not fully integrated with the unit-based electronic systems. The computerized systems we observed were not programmed to allow controllers to easily capture and compile all the data relevant to Matching Michigan.

The TTT units collected information on suspected (and treated) infections in a standardized form that was fully integrated with the routine care of patients. A TTT controller therefore had ready access to the information needed for making decisions, plus assurance that the information was largely congruent with clinical decision making as it happened. In contrast, most ICUs operating Patrol or Controller-Centered systems had many difficulties collecting and compiling such evidence. In these ICUs, records of line insertion, signs of infection, and microbiology results tended to be patchy and disorganized. On Patrol units, the nurses collecting the data often had to riffle through electronic records or multiple documents, stored in different places, to determine whether the patient might have an infection, or even the date on which the CVC was inserted. As a result, on one field visit, a patrol nurse was observed missing an infection that had been discussed earlier in the ward round and not recording it.

The Controller-Centered units varied greatly in how well their systems were designed and how well they worked. On nine of these units, staff were supposed to record suspected CVC-BSIs on a standardized notification form that they had to fetch from a centralized location (such as the nursing station). These notification forms were intended to be used by the controller to identify possible infections on which he or she could make a judgment on the reporting. However, staff were routinely seen to discuss and initiate treatment for a suspected infection, but not to complete a notification form recording this suspicion. One reason for

this was that their early decisions lacked the certainty that physicians felt was needed to produce a formal record. Although physicians could conclude whether an infection might be linked to a CVC if the only reason for so doing was to formulate a treatment plan, they were much less comfortable with turning an intuitive decision, perhaps made in the absence of laboratory tests, into an auditable moment. Physicians also did not complete forms because they were unaware of or forgot the need to complete them. If they did know and remember, they might have little enthusiasm for completing a document that some saw as having no use other than satisfying an externally imposed and illegitimate demand and as being of no benefit to patients or staff.

My approach to Matching Michigan is that it's just another political buzz word. I'm sorry, but I'm busy, and I haven't had the time to fill in the data sheets, and I haven't had the energy to continue to nag my colleagues to fill the sheets in either. (028: medical lead)

The more enthusiastic and committed controllers sought to counter these problems by actively seeking information about infections and encouraging their colleagues to notify them of any suspected cases, but with mixed results.

I'm only there one week in [every six or seven]. And the other weeks, you know, I can't really be sure what's going on a lot of the time. Enthusiasm varies amongst my colleagues quite a lot as well. (110: medical lead)

Much depended on the quality and number of prompts notified to the controller, but the notification forms were, of course, only part of the evidence required. The job of compiling comprehensive evidence from multiple sources to support decisions about what to report was far from trivial. Controllers varied in their ability to access patients' records, microbiology reports, and any relevant clinical information; mundane problems of coordination and organization were often evident.

The problem is with the pieces of paper as well. They were going missing, or I wasn't picking them up. They were just getting left lying around. (089: clinical quality and risk manager)

[She] does the reports at the end of the month. Not when the [microbiologists] come around. Not when she's got access to the patient and the file. A lot of the time it's done from memory, and so it's even less accurate. I've tried on a couple of occasions to suggest she start filling the form in during the [microbiology] rounds but, to be honest, she's not very proactive about it. (109: nursing lead)

The problem we have with the numerator data is [our] microbiology lab changed their computer—which has thrown everything into chaos from their point of view—and they're finding it very very difficult to maintain [a] full clinical service, let alone contribute to data requests [like] "Can we have our last three months' possible [CVC-]BSIs? And can a consultant microbiologist spend two hours and sit down and go through them with me?" (098: medical lead)

### *Making Judgments about What Counted as a CVC-BSI*

Our analysis thus far suggests that the information available to controllers about both the numbers of CVCs in situ and the possible infections varied from unit to unit, reflecting the ways in which the local data collection systems operated. Yet this was what the controllers had to use to determine whether any infections could be deemed CVC-BSIs to the program: they were clearly not making decisions about the same things. Not only did the information available to the controllers vary, so too did how they made their decisions and who the controllers were.

*How* the controllers made their decisions about what could be reported as an infection meeting the program's criteria was not consistent. The controllers did not report to the program at the end of the month all the patients who were discussed on the ward rounds as having infections linked to their catheters as having a CVC-BSI, whether or not they had received a notification form or other prompt about these patients. What controllers decided to report was not a matter of manipulation or fabrication of data, however. Rather, the standards used to judge what counted as an infection for purposes of initiating treatment for a patient differed from the standards used to determine whether an infection should be reported.

You've got a positive culture from [a patient's] tracheo-bronchial tree; you've got a positive culture from the central line; and the patient has been on antibiotics for the last couple of days. What's the most

plausible [source of infection?] Roll the dice! . . . I think [this] is the weakest link in the whole [Matching Michigan] protocol. It doesn't change clinical care. It doesn't have any impact on patient care, because you do what you do. The problem is how you count it. (103: medical lead)

At the heart of this difference was the view that evidence good enough for clinical care was not of the same character and quality as that needed for external reporting. For ICU physicians in day-to-day practice, the decision about whether there was enough suspicion of a CVC-BSI to pull a catheter and start a patient on antibiotics was situationally dependent. It could be made on the basis of incomplete information but could be explained by the patient's biography, the contingencies of the clinical context, and/or the physician's own experience and knowledge. This practical activity was thus accountable and justifiable within the norms of its local context. But converting the vernacular of infections in the clinical setting into the cold formality of a performance measure was a task that had to take account of the possible scrutiny of the performance of the unit reporting the data. Translating a clinical decision into information that could be used to judge the standard of care provided by the organization challenged the controllers because they lacked a form of what Power (2004, 768) termed "legitimate instrumentation." The challenge for controllers was that CVC-BSIs were not entities with fixed and immutable properties that yielded to straightforward measurement.

As we have noted, many microbiology laboratories could not undertake the analyses necessary for the catheter-*related* definition, which would provide more reliable evidence of whether the CVC was guilty. But the catheter-*associated* definition that controllers then had to use was seen as slippery and subjective, leaving considerable "wriggle room" for individual clinical judgment and lack of standardization. Staff pointed out that without a trusted technology that could secure the objectivity of their assessments, any decision about whether the CVC was to blame was inherently discretionary. Only rarely, it seemed, was it possible for absolute confidence about the catheter's guilt.

I think it's still—it could still be—subjective. I'm sure if two clinicians looked at the same data, they would possibly come up with two different answers. I still think [the definitions] are a little bit ambiguous. (056: nursing department head)

Nonetheless, all the reports of infections submitted to the program were “latently supervisory” (Freidson 1988, 148) and capable of being used to judge quality of care. The uncertainties meant that the threshold for a declarable infection was set in different places by different units. In three separate discussions of possible CVC-BSIs that we witnessed, all were eventually judged not to qualify as reportable under the program’s definitions. To the nonclinical observer, it seemed that the decisions could have gone either way but that all three decisions were defensible. Our interviews suggested that some controllers were unwilling to forgive the CVC but that others favored finding any reasonable explanation other than the CVC.

I think you know if you have got the same bug growing on your catheter as you have in the bloodstream and the patient is unwell. . . . If [the microbiological test results] fit the clinical picture and the clinical suspicion is there, [then] I think [regardless of] the definition of “related” or otherwise, I think we have to assume that the line is at least partly implicated if not [totally responsible]. (042: medical lead)

My personal feeling is that it’s very easy to—I’m not going to say manipulate—but to play with the data and look for another plausible source of infection. . . . You can get any number you want. (020: consultant)

These ambiguities about how the definitions would be applied and the institutional logics that pervaded staff’s understanding of performance measurement, as well as anxieties about the fairness of any comparisons, meant that some staff were concerned that “looking harder” for infections could be punished, even though the program had no formal sanctions.

People have learned that if you can wriggle out of a diagnostic criterion, you’ll do so. And if the diagnostic criteria are—[pauses] troublesome, to put it politely, or difficult or impractical—then it’s even easier. (093: consultant microbiologist)

We may have shot ourselves in the foot by having a better-quality data collection system and robust [counting] processes in place. People are not all going about this the same way in different units. You can see really there is no incentive to report accurately if you are going to be penalized for it (035: infection control nurse)

These sources of variability were joined by further variability in the controllers' backgrounds and skill sets, how they went about their tasks, and the degree of support they received from colleagues. In most cases, the controller was the local Matching Michigan medical lead and thus was usually an ICU consultant, although sometimes he or she was a senior nurse from the ICU or infection control. Organizations with multiple ICUs varied in their arrangements: some had one controller per ICU, and others had one controller for the whole organization. Several ICUs had trouble finding someone to take the job of controller, sometimes reflecting the priority given to the program and sometimes reflecting more mundane issues of available personnel, with the two reasons tending to interact. Three Controller-Centered units assigned the responsibility to individual junior doctors, presenting it as a bonus for a young physician's résumé. This strategy tended to cause losses of data and to signal that it was a low-priority role, and in two units it resulted in incomplete submissions over time.

Some controllers made their decisions about reportable infections alone. Others involved colleagues, such as infection control nurses and other physicians, particularly those from microbiology, although the microbiologists did not necessarily feel any less ambiguity about making determinations.

It's part of my working relationship with the ICU to help them with the task they've been given. And the task they've been given is to count the uncountable. [The ICU physicians] will say, "I don't know!" And the surveillance nurse will say, "I don't know!" And they'll [both] say, "Ask [the microbiologist]!" And I'll get a coin out of my pocket. And I'm not comfortable with putting a lot of effort into flipping coins. (093: consultant microbiologist)

### *Local Credibility of CVC-BSI Rates*

Making rates of CVC-BSIs visible was intended to provoke change where needed in local units. The quality of the data collection had a direct impact on the local credibility of the data and on the extent to which it was seen as actionable. The strengths and weaknesses of systems were often mutually reinforcing. The few ICUs that saw the program as a local opportunity to improve practices tended to adopt more robust data collection systems, to use the data to determine what change was

required, and to monitor progress. On TTT units in particular, the tight link between the care of the patients and the data on those patients provided the impetus for learning. One TTT unit dramatically drove down its infections from an initially high rate through full implementation of the program's technical and nontechnical interventions and through very carefully monitoring how their rates of infection responded. In contrast, some units that saw the program as an externally imposed audit of little relevance to clinical care devoted little effort to designing and operating a robust system, often seeing the data as a source of fear rather than learning or as a tedious distraction from the real work of caring for patients.

Where there were weaknesses in capturing data in a locally credible way, there usually was little change. Data indicating that infection rates were low had the potential to reassure (possibly inappropriately) that no action was needed. Data indicating that rates were high were, in contrast, prone to being dismissed by senior clinicians as being of poor quality and lacking in legitimacy. This was particularly true of the Patrol system. Even though it was operated by enthusiastic and dedicated personnel, its reliance on medical records and ICU outsiders opened it not just to data loss problems but also to controversy. High rates of reported infection were challenged by ICU insiders, who argued that the data collected by the patrol nurses and interpreted by the off-site controller did not reflect clinical realities. This resulting loss of credibility eroded clinicians' willingness to change their practices.

## Discussion

Our analysis of a patient safety improvement program that sought to measure rates of central venous catheter bloodstream infections demonstrates that despite being given explicit and widely used definitions, the participating units were not counting the same things in the same way. Reported rates of CVC-BSIs depended on where and by whom the measuring was done: neither the numerators nor the denominators reported to the program by the different ICUs were fully comparable. Our study shows that what counted as relevant to reporting reflected localized interpretations and differed from what was required to produce a standardized data set across all ICUs. This may help explain previous studies indicating variability in diagnosing and reporting CVC-BSIs

(Lin et al. 2010). It also challenges currently dominant assumptions in the economics and public administration literature that explain these different measurements as the result of either effort redirection or gaming. We propose that variability arises because CVC-BSIs do not simply wait to be discovered through the mechanical application of unambiguous technical criteria. Instead, determining what counts as a CVC-BSI requires different kinds of work. Rather than a single narrative of gaming and effort redirection, three rather more complex and interrelated stories emerged from our study of the organizational processing of data on CVC-BSIs and how classificatory judgments are made. All were important to undermining the comparability of rates of infection across units.

The first type of work needed to generate counts was logistical in character. Contradicting the literature on effort substitution and gaming, we did not find cunning workers engaged in deliberate manipulation. Instead, much of what explains variation in the infection rates generated by different ICUs was artless rather than artful. ICUs were charged with creating the conditions of their own measurability, but reliable data collection systems were challenging and tiresome to design and operate. Poor alignment of traditional clinical practices and the generation of data for auditing purposes was evident in most ICUs. Long established and deeply institutionalized local practices, including microbiology support and clinicians' preferences and behaviors, contributed to substantial variability in what was submitted for laboratory analysis and what analyses were undertaken. Information was prone to losses because of difficulties in coordinating large amounts of activity in highly pressurized, fast-moving environments with multiple competing priorities. The importance of the mundane challenges associated with data collection should not be underestimated; at every turn, the complexity of apparently simple things continues to confound organizational attempts at improving quality (Dixon-Woods 2010).

Legitimacy work (Spertus 1995) was the second kind of work required to produce the numbers. Health care staff have little attention and time. They therefore must choose what to prioritize, and persuading staff that complying with record keeping is "a respectable thing for them to be doing" (Garfinkel 1994, 94) is never easy. For ICU staff to value the process and outcome of counting CVC-BSIs, these activities needed to be seen as legitimate. Legitimacy work involves persuading staff that the count is relevant to their patients' care and convincing them that

the data themselves are valuable and are not simply a response to an illegitimate bureaucratic intrusion. The particular characteristics of a program led by a government agency, perceived by many as having little relevance to the realities of clinical work and as latently punitive, made legitimacy work especially difficult.

The third form of work was classificatory in character. The ICUs classified both eligible patients and eligible infections in ways that were sensitive to the purposes for which any resulting information might be used. Staff asked to determine CVC-BSIs to be reported to the program understood that to some extent they were constructing visible signs of reasonable practice (Power 2003). Given the recent history of targets and terror, a tendency to absolve CVCs of possible guilt if other less shaming culprits could be found would be understandable. Yet not only was this tendency absent from some settings we studied, it would also be a mistake to treat such tendencies as illegitimate or as evidence of cheating where it was found. Counting and reporting was not the work of applying simple technical principles. Instead, precise and reliable instrumentation that could objectively determine the source of infections was lacking. Counting was best understood as having some cognitive and some technical components, as well as many of the features of a social practice influenced by organizational and institutional contexts.

To some extent, our analysis follows a long tradition of sociological work investigating the social organization and production of official statistics in areas ranging from crime to death certification (Atkinson 1978; Bloor 1991; Cicourel 1964; Haggerty 2001; Timmermans 2005) and the more recent work on practices of making visible (and hence evaluable) aspects of societal, organizational, and individual behavior and performance (Berg and Mol 1998; Hacking 1990; Power 1997). Our concern, however, is not simply reaffirming the micropolitical character of measurement practices or demonstrating that the technical design of an accounting system is never independent of the organizational environment in which it is deployed (Bowker and Star 1999). We are also concerned with identifying the practical and policy implications of our findings, particularly given the high stakes of using rates of CVC-BSIs as a measure of quality.

Our study suggests that differences in reported rates of CVC-BSIs may relate at least in part to underlying variations in the social practices of data collection and reporting rather than to objective differences in incidence. Assumptions of the comparability of CVC-BSI rates across

different institutions and contexts may therefore be ill founded. Reported rates may not provide a sound basis for judging the relative performance of organizations or for the use of rewards and sanctions. The lack of agreement in classificatory practices that we found is, of course, a common feature of scientific fields as far apart as astronomy, fisheries, and museum artifacts (Winiiecki 2008) and is not unusual in fields of medicine, such as pathology (Harris et al. 2008). But when pressure is put on such classificatory practices as a means of judging quality of practice, there is a substantial risk that the standards used to hold organizations or individuals to account will be perceived as unfair or capricious.

Inequity aversion—a dislike of unfair outcomes—is a very strong characteristic of any setting in which people are asked to work toward common goals (Poteete, Janssen, and Ostrom 2010). Our study showed that perceptions of inequity caused erratic counting behavior, including, for example, the inclusion or exclusion of certain kinds of patients from the denominator. Such exclusions of patients who are distinctive in some way are explicitly allowed in some performance measurement schemes, such as the Quality and Outcomes Framework used to reward primary care physicians in the United Kingdom. This framework asks physicians to determine whether patients are ineligible for counting for some indicators, on grounds such as frailty or terminal illness. Such “exception reporting” appears to deliver substantial benefits in this pay-for-performance program (Doran et al. 2008). Research evidence is now demonstrating that counting patients with multiple concurrent CVCs as generating one catheter day in the same way as a patient with one CVC inflates the CVC-BSI rate for ICUs caring for particular kinds of patients and fails to adjust for underlying illness (Aslakson et al. 2011), suggesting that some units with high-risk patients may be unfairly labeled as poor performers. There is now a strong case for revisiting the CDC’s definition of a patient catheter-day, not least in order to promote fairness and improve scientific understanding.

Definitions of bloodstream infections may be less tractable to easy resolution. Currently, U.S. hospitals may use either the Catheter-Related (CR-BSI) or the Catheter-Associated (CA-BSI) definition to classify their infections, but most, for resource reasons, use the CA-BSI definition (O’Grady et al. 2011). Using the CA-BSI definition to measure performance on health care-acquired infections is problematic (Fridkin and

Olmsted 2011). Known as a “surveillance” definition in public health terms, it was originally developed primarily to monitor populations and for organizations’ internal purposes. Its validity is demonstrably precarious (Sihler et al. 2010) because it overcounts CVC-BSIs, thus potentially penalizing organizations unable to adhere to the CR-BSI definition. It is also prone to inconsistent interpretation, as our study and others have shown. In the United States, where decisions about countable infections are usually made by infection preventionists (IPs), the model most similar to the Patrol system that we described, a recent study found poor interrater reliability in IPs’ decisions about whether CVC-BSIs were present when reviewing the same medical records (Mayer et al. 2012), even though these decisions were made in a research context that did not involve performance assessment.

Before CVC-BSIs were used as a performance measure, the data noise associated with the CA-BSI definition was of little consequence, and could be resolved locally (Sexton, Chen, and Anderson 2010). Rates based on this definition could be used by organizations to detect trends over time as long as they were internally consistent in their counting practices. The current use of these rates for performance measurement, pay-for-performance, and reputational sanctions, however, has converted a locally useful definition into a means of scrutiny and control, and could undermine its value for any purpose, as well as risking unfairness. The fallibilities of data collection and reporting systems also have important consequences for improvement efforts: poor practices may be reinforced; improvements may not be rewarded; or the search for cases may be less aggressive (Niedner and 2008 National Association 2010).

Even though automated surveillance (e.g., using computer algorithms) has been proposed as a way of rationalizing and standardizing decisions about the source of BSIs, our study suggests some reasons for caution about this kind of technical fix. First, such methods may not account for the underlying variation in clinical practices and organizational processing of data that we found. Second, some evidence from the United States suggests that, as we also found, clinical staff do not always agree with the decisions about qualifying infections that are made by those not at the scene of the clinical action (Sexton, Chen, and Anderson 2010). Some element of clinical discretion may therefore be an inescapable feature of counting infections. Our findings suggest that

systems with features similar to those we found in Track-Trigger-Track units may offer the best opportunity for aligning the goals of clinical practice and surveillance. If these features can be made part of electronic health records, they may prove especially beneficial. Third, if CVC-BSI rates are to be used for performance assessment, fairness and consistency may best be served if all microbiology departments can use quantitative and semiquantitative culture techniques in order to use the CR-BSI definition, which would improve diagnostic accuracy for clinical purposes at the same time. Moving to this standard would, however, require some time and investment and also standardization of other clinical and social practices relating to data collection and reporting. In the meantime, the justice of using CVC-BSIs as a performance measure open to sanctions may be questionable.

Our study also has important implications for current policies of classing a CVC-BSI as a “never event.” If the data produced by different settings are not comparable, then “getting to zero,” the standard implied by most targets and standards in the United States and elsewhere, may not always be possible for all units. The relationship between catheter care and infection outcomes may not be as stable as the current policy assumes. Recent research is beginning to show that those patients who develop CVC-BSIs may do so despite the implementation of technical best practices, and they also may have distinctive characteristics, including more severe illness and repeat abdominal operations (Lissauer et al. 2012). Some evidence is now also suggesting that some CVC-BSIs are the result of secondary infection and cannot be addressed by interventions focused on CVC insertion (Sihler et al. 2010). A rate of zero may be thus consistently achievable only for certain well-defined patient groups at low risk of infection. So, until the evidence is firmer, perhaps it should not be used as the standard against which all others are judged. This is important not only because organizations could be inappropriately penalized but also because of the implications for patient management in different risk groups. At present, it is not clear what proportion of CVC-BSIs can be eliminated for which patient groups. However, the current insistence on “zero” as the only acceptable standard may inhibit learning if organizations are discouraged from admitting that they have identified infections.

Our study has a number of limitations. We do not know the extent to which any reported changes in CVC-BSIs rates in Matching Michigan over time are “true” or an artifact of changing data collection and

reporting behaviors, since our observations were undertaken as one-off snapshots in each ICU rather than over time. This will be an important focus for future work. We also carried out our work in a setting where the incentives for gaming and effort redirection were (in principle) few, since no official penalties applied, and the generalizability of our findings to higher-stakes settings may therefore be questioned. This is important, given some evidence of effort redirection in some U.S. hospitals, such as the greater inappropriate use of peripheral lines in order to avoid infections attributable to CVCs (Lissauer et al. 2012).

## Conclusions

In designing improvement programs and using performance measures for the future, our study offers some important lessons and reinforces Wachter's (2012, 40) argument that "we have much to learn about which measures to target, how to collect the data, and how to promote improvement at a reasonable cost and with a minimum of unanticipated consequences." Effort substitution and gaming do not provide a full account of reporting behavior in relation to the rates of central venous catheter bloodstream infections in the settings we studied. However, unless they are carefully designed and fully integrated with clinical priorities, and impose minimum burden, data collection systems may produce incommensurable data that do not provide a fair basis for cross-institution comparison. Measuring and managing quality in health care requires the development of metrics and data collection systems that clinicians and organizations believe to provide fair and true comparisons, not least so that if sanctions are applied, they are based on good evidence.

## References

- Aslakson, R.A., M. Romig, S.M. Galvagno, E. Colantuoni, S.E. Cosgrove, T.M. Perl, and P.J. Pronovost. 2011. Effect of Accounting for Multiple Concurrent Catheters on Central Line-Associated Bloodstream Infection Rates: Practical Data Supporting a Theoretical Concern. *Infection Control and Hospital Epidemiology* 32(2):121-24.
- Aswani, M.S., J. Reagan, L. Jin, P.J. Pronovost, and C. Goeschel. 2011. Variation in Public Reporting of Central Line-Associated

- Bloodstream Infections by State. *American Journal of Medical Quality* 26(5):387–95.
- Atkinson, J. 1978. *Discovering Suicide: Studies in the Social Organization of Sudden Death*. London: Macmillan.
- Atkinson, P., and A. Coffey. 2004. Analysing Documentary Realities. In *Qualitative Research: Theory, Methods and Practice*, ed. D. Silverman, 56–73. London: Sage.
- Backman, L.A., R. Melchreit, and R. Rodriguez. 2010. Validation of the Surveillance and Reporting of Central Line–Associated Bloodstream Infection Data to a State Health Department. *American Journal of Infection Control* 38(10):832–38.
- Berg, M., and A. Mol. 1998. *Differences in Medicine: Unraveling Practices, Techniques, and Bodies*. Durham, NC: Duke University Press.
- Besley, T., and M. Ghatak. 2005. Competition and Incentives with Motivated Agents. *American Economic Review* 95(3):616–36.
- Bevan, G., and R. Hamblin. 2009. Hitting and Missing Targets by Ambulance Services for Emergency Calls: Effects of Different Systems of Performance Measurement within the UK. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 172(1): 161–90.
- Bevan, G., and C. Hood. 2006. What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration* 84(3):517–38.
- Bird, S.M., D. Cox, V.T. Farewell, H. Goldstein, T. Holt, and P.C. Smith. 2005. Performance Indicators: Good, Bad, and Ugly. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 168(1): 1–27.
- Bloor, M. 1991. A Minor Office: The Variable and Socially Constructed Character of Death Certification in a Scottish City. *Journal of Health and Social Behavior* 32(3):273–87.
- Bosk, C.L. 2003. *Forgive and Remember: Managing Medical Failure*. 2nd ed. Chicago: University of Chicago Press.
- Bowker, G.C., and S.L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Charmaz, K. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: Sage.
- Cicourel, A. 1964. *Method and Measurement in Sociology*. New York: Free Press.
- Courty, P., and G. Marschke. 2003. Dynamics of Performance-Measurement Systems. *Oxford Review of Economic Policy* 19(2): 268–84.
- Davidoff, F. 2009. Heterogeneity Is Not Always Noise. *JAMA* 302(23):2580–86.

- Dixon-Woods, M. 2010. Why Is Patient Safety So Hard? A Selective Review of Ethnographic Studies. *Journal of Health Services Research & Policy* 15(suppl. 1.):11–16.
- Dixon-Woods, M., and C.L. Bosk. 2010. Learning through Observation: The Role of Ethnography in Improving Critical Care. *Current Opinion in Critical Care* (epub ahead of print).
- Dixon-Woods, M., C.L. Bosk, E.L. Aveling, C.A. Goeschel, and P.J. Pronovost. 2011. Explaining Michigan: Developing an Ex Post Theory of a Quality Improvement Program. *The Milbank Quarterly* 89(2):167–205.
- Doran, T., C. Fullwood, D. Reeves, H. Gravelle, and M. Roland. 2008. Exclusion of Patients from Pay-for-Performance Targets by English Physicians. *New England Journal of Medicine* 359(3):274–84.
- Frasca, D., C. Dahyot-Fizelier, and O. Mimoz. 2010. Prevention of Central Venous Catheter–Related Infection in the Intensive Care Unit. *Critical Care* 14(2):212.
- Freidson, E. 1988. *Profession of Medicine: A Study of the Sociology of Applied Knowledge*. Chicago: University of Chicago Press.
- Fridkin, S.K., and R.N. Olmsted. 2011. Meaningful Measure of Performance: A Foundation Built on Valid, Reproducible Findings from Surveillance of Health Care–Associated Infections. *American Journal of Infection Control* 39(2):87–90.
- Garfinkel, H. 1994. *Studies in Ethnomethodology*. Cambridge: Polity.
- Glaser, B., and A. Strauss. 1967. *The Discovery of Grounded Theory*. Hawthorne, NY: Aldine.
- Hacking, I. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.
- Haggerty, K. 2001. *Making Crime Count*. Toronto: University of Toronto Press.
- Hammersley, M., and P. Atkinson. 2007. *Ethnography: Principles in Practice*. London: Routledge.
- Harris, E.I., D.N. Lewin, H.L. Wang, G.Y. Lauwers, A. Srivastava, Y. Shyr, B. Shakhtour, F. Revetta, and M.K. Washington. 2008. Lymphovascular Invasion in Colorectal Cancer: An Interobserver Variability Study. *American Journal of Surgical Pathology* 32(12):1816–21.
- Holmstrom, B., and P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization* 7(1):24–52.
- Kelman, S., and J.N. Friedman. 2009. Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortional Impacts of the Emergency Room Wait-Time Target in the English National Health Service. *Journal of Public Administration Research and Theory* 19(4):917–46.

- Lin, M.Y., B. Hota, Y.M. Khan, K.F. Woeltje, T.B. Borlowsky, J.A. Doherty, K.B. Stevenson, R.A. Weinstein, and W.E. Trick. 2010. Quality of Traditional Surveillance for Public Reporting of Nosocomial Bloodstream Infection Rates. *JAMA* 304(18): 2035–41.
- Lissauer, M.E., S. Leekha, M.A. Preas, K.A. Thom, and S.B. Johnson. 2012. Risk Factors for Central Line–Associated Bloodstream Infections in the Era of Best Practice. *Journal of Trauma and Acute Care Surgery* 72(5):1174–80.
- Longmate, A.G., K.S. Ellis, L. Boyle, S. Maher, C.J. Cairns, S.M. Lloyd, and C. Lang. 2011. Elimination of Central-Venous-Catheter-Related Bloodstream Infections from the Intensive Care Unit. *BMJ Quality & Safety* 20(2):174–80.
- Magill, S.S., and S.K. Fridkin. 2012. Improving Surveillance Definitions for Ventilator-Associated Pneumonia in an Era of Public Reporting and Performance Measurement. *Clinical Infectious Diseases* 54(3):378–80.
- Mannion, R., and J. Braithwaite. 2012. Unintended Consequences of Performance Measurement in Healthcare: 20 Salutary Lessons from the English National Health Service. *Internal Medicine Journal* 42(5):569–74.
- Mayer, J., T. Greene, J. Howell, J. Ying, M.A. Rubin, W.E. Trick, M.H. Samore, and CDC Prevention Epicenters Program. 2012. Agreement in Classifying Bloodstream Infections among Multiple Reviewers Conducting Surveillance. *Clinical Infectious Diseases* (epub ahead of print).
- Mermel, L.A., M. Allon, E. Bouza, D.E. Craven, P. Flynn, N.P. O'Grady, I.I. Raad, B.J. Rijnders, R.J. Sherertz, and D.K. Warren. 2009. Clinical Practice Guidelines for the Diagnosis and Management of Intravascular Catheter-Related Infection: 2009 Update by the Infectious Diseases Society of America. *Clinical Infectious Diseases* 49(1):1–45.
- Niedner, M.F., and 2008 National Association of Children's Hospitals and Related Institutions Pediatric Intensive Care Unit Patient Care FOCUS Group. 2010. The Harder You Look, the More You Find: Catheter-Associated Bloodstream Infection Surveillance Variability. *American Journal of Infection Control* 38(8):585–95.
- O'Grady, N.P., M. Alexander, L.A. Burns, E.P. Dellinger, J. Garland, S.O. Heard, P.A. Lipsett, H. Masur, L.A. Mermel, M.L. Pearson, I.I. Raad, A.G. Randolph, M.E. Rupp, S. Saint, and Healthcare Infection Control Practices Advisory Committee. 2011. Guidelines for the Prevention of Intravascular Catheter-Related Infections. *American Journal of Infection Control* 39(4 suppl. 1):S1–S34.

- Poteete, A.R., M.A. Janssen, and E. Ostrom. 2010. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton, NJ: Princeton University Press.
- Power, M. 1997. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Power, M. 2003. Auditing and the Production of Legitimacy. *Accounting, Organizations and Society* 28(4):379–94.
- Power, M. 2004. Counting, Control and Calculation: Reflections on Measuring and Management. *Human Relations* 57(6):765–83.
- Prendergast, C. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37(1):7–63.
- Pronovost, P.J., M. Miller, and R.M. Wachter. 2007. The GAAP in Quality Measurement and Reporting. *JAMA* 298(15):1800–1802.
- Pronovost, P., D. Needham, S. Berenholtz, D. Sinopoli, H. Chu, S. Cosgrove, B. Sexton, R. Hyzy, R. Welsh, G. Roth, J. Bander, J. Kepros, and C. Goeschel. 2006. An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU. *New England Journal of Medicine* 355(26):2725–32.
- Propper, C., and D. Wilson. 2003. The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy* 19(2):250–67.
- Sexton, D.J., L.F. Chen, and D.J. Anderson. 2010. Current Definitions of Central Line–Associated Bloodstream Infection: Is the Emperor Wearing Clothes? *Infection Control and Hospital Epidemiology* 31(12):1286–89.
- Sihler, K.C., C. Chenoweth, C. Zalewski, W. Wahl, R. Hyzy, and L.M. Napolitano. 2010. Catheter-Related vs. Catheter-Associated Blood Stream Infections in the Intensive Care Unit: Incidence, Microbiology, and Implications. *Surgical Infections* 11(6):529–34.
- Spertus, J.A., K.A. Eagle, H.M. Krumholz, K.R. Mitchell, and S.T. Normand. 2005. American College of Cardiology and American Heart Association Methodology for the Selection and Creation of Performance Measures for Quantifying the Quality of Cardiovascular Care. *Journal of the American College of Cardiology* 45(7):1147–56.
- Spertus, M.C. 1995. Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review* 20(3):571–610.
- Timmermans, S. 2005. Suicide Determination and the Professional Authority of Medical Examiners. *American Sociological Review* 70(2):311–33.
- Wachter, R.M. 2012. *Understanding Patient Safety*. 2nd ed. New York: McGraw-Hill Medical.

- Waring, J.J. 2009. Constructing and Re-constructing Narratives of Patient Safety. *Social Science & Medicine* 69(12):1722–31.
- Winiecki, D.J. 2008. An Ethnostatistical Analysis of Performance Measurement. *Performance Improvement Quarterly* 20(3–4):185–209.

---

*Acknowledgments:* We thank Janet Willars for conducting the telephone interviews. We thank Charles Bosk, Frank Davidoff, Carl Macrae, and Graham Martin for invaluable comments on an earlier draft. We thank three anonymous reviewers whose very thoughtful comments helped improve our manuscript. We thank the organizations, units, and staff who participated in this study. This article is based on work commissioned by the Health Foundation (registered charity number 286967). Professor Julian Bion held the advisory role of senior clinical lead for the Matching Michigan project, conducted in parallel with the Lining Up project on which this article is based. He did not have access to any primary data from Lining Up while Matching Michigan was running, and analysis of the Lining Up data was conducted independently of Professor Bion.