

How hard is it to know whether care is safe?

Mary Dixon-Woods,
University of Leicester

Performance Indicators in healthcare:
development, validation and use
Copenhagen, 30 September 2013

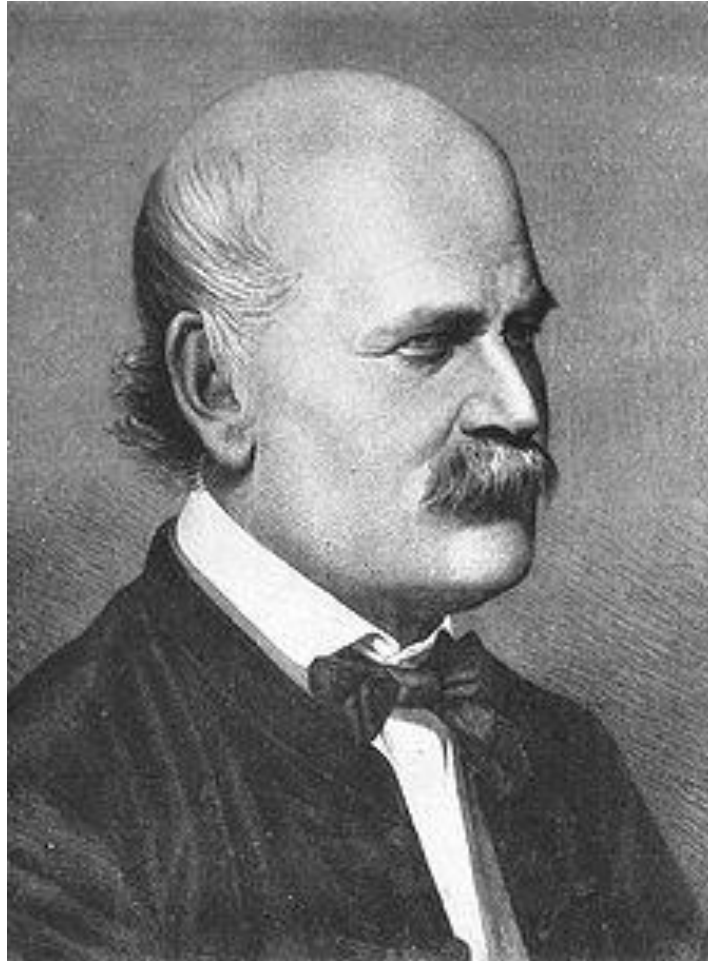
William Petty 1623-1687



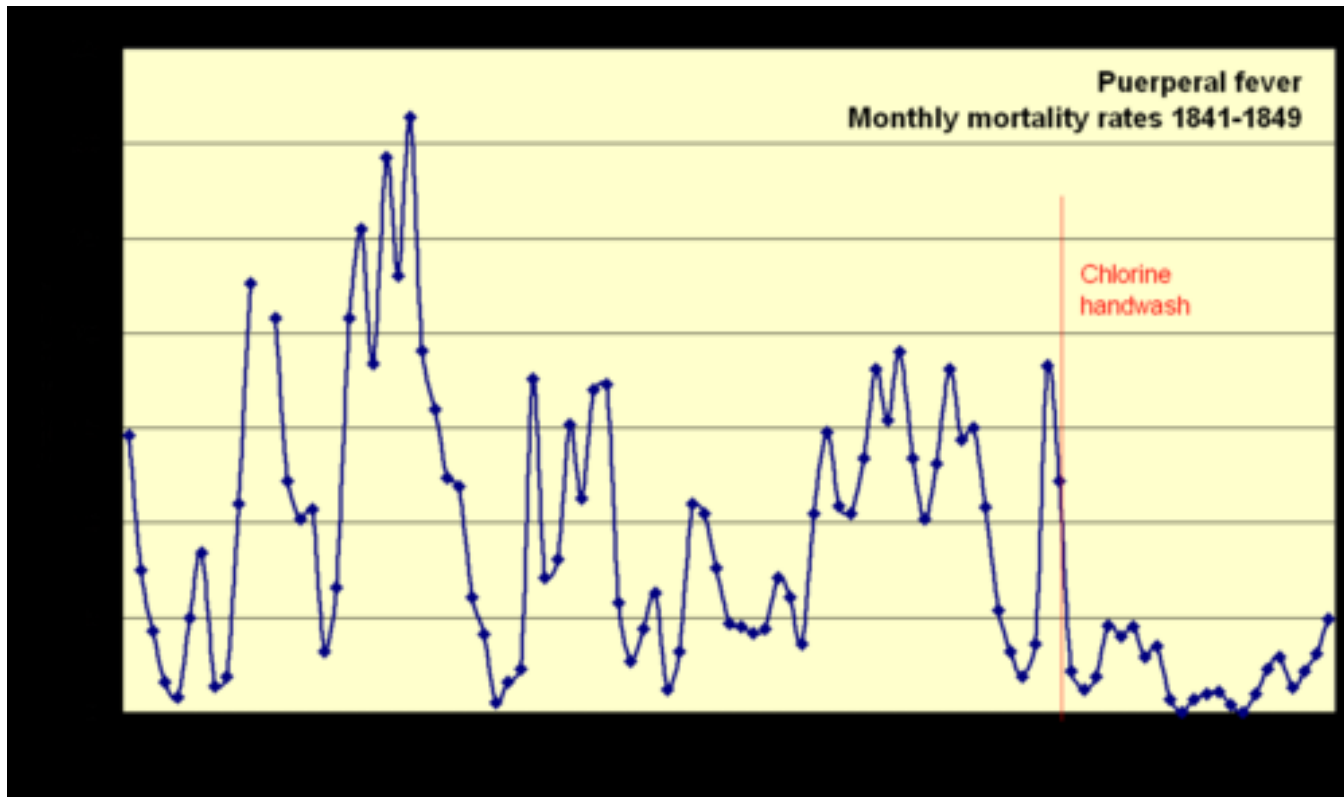
17th century performance management

- Mortality 1678-1679, Paris and London
 - L'Hotel Dieu: 28%
 - La Charité: 14%
 - St Bart's and St Thomas's: 12%
- Petty concluded that 3,000 of those who died in L'Hotel Dieu
 - 'did not die by natural necessity, but by the evil administration of that Hospital'
 - <http://www.theactuary.com/archive/old-articles/part-6/the-works-of-william-petty/>

Ignaz Semmelweis 1818-1865



Vienna Maternity Institution: rates of puerperal fever 1841-1849

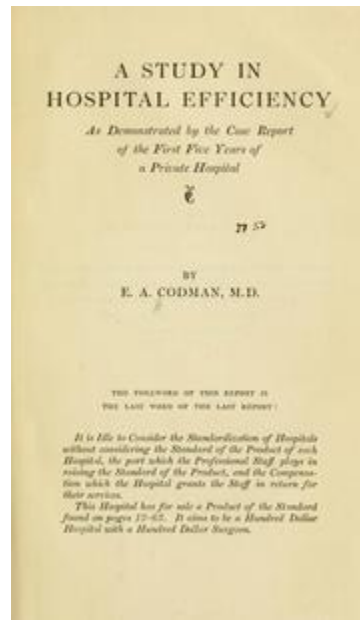


Ernest Codman 1869-1940



Codman's outcomes management

- Use of end-result cards



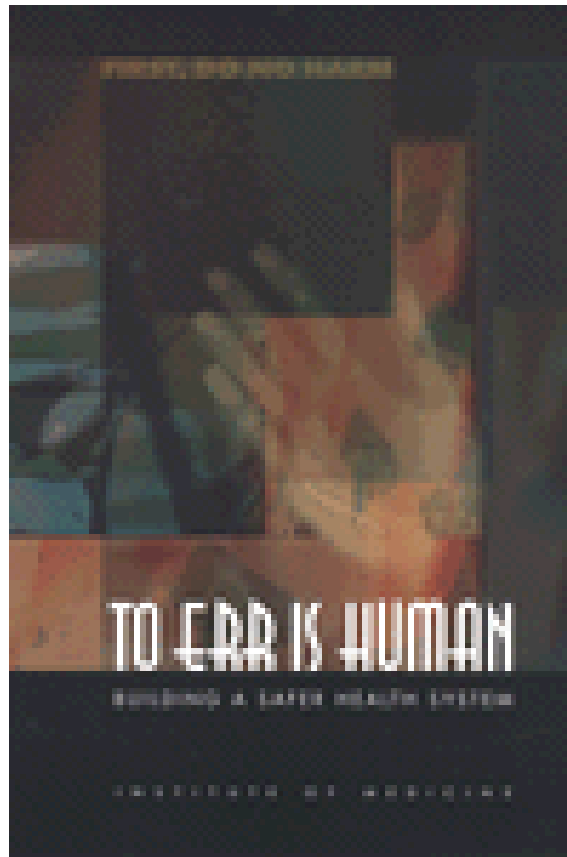
Dartmouth Atlas

NHS Atlas

Need for measurement

- So measurement is really important if we want to know about quality and safety of care
- But it's a lot harder than it looks





- 44,000 to 98,000 preventable deaths
- Based on one study from 1984 and another from 1992 using case note review
- Extrapolated figures using number of hospitalisations

- Used Global Trigger Tool; estimated 210,000 preventable adverse events annually that contribute to death of patients in US hospitals (34.4m hospitalisations)
- But we don't know how many are actually preventable

REVIEW ARTICLE

A New, Evidence-based Estimate of Patient Harms Associated with Hospital Care

John T. James, PhD

Objectives: Based on 1984 data developed from reviews of medical records of patients treated in New York hospitals, the Institute of Medicine estimated that up to 98,000 Americans die each year from medical errors. The basis of this estimate is nearly 3 decades old; herein, an updated estimate is developed from modern studies published from 2008 to 2011.

Methods: A literature review identified 4 limited studies that used primarily the Global Trigger Tool to flag specific evidence in medical

the national level. The amount of new knowledge generated each year by clinical research that applies directly to patient care can easily overwhelm the individual physician trying to optimize the care of his patients.¹ Furthermore, the lack of a well-integrated and comprehensive continuing education system in the health professions is a major contributing factor to knowledge and performance deficiencies at the individual and system level.² Guidelines for physicians to optimize patient care are quickly out of date and can be biased by those who write the

What is preventable?

- Many harms are in principle preventable
- But boundaries of preventability are often unclear
- And involve difficult trade-offs

- Marked differences in GTT harm rates in 5 Danish hospitals
- Training, experience, procedures

Open Access

Research



BMJ
open
accessible medical research

Experiences with global trigger tool reviews in five Danish hospitals: an implementation study

Christian von Plessen,¹ Anne Marie Kodal,² Jacob Anhøj³

To cite: von Plessen C, Kodal AM, Anhøj J. Experiences with global trigger tool reviews in five Danish hospitals.

ABSTRACT
Objectives: To describe experiences with the implementation of global trigger tool (GTT) reviews in five Danish hospitals and to suggest ways to improve

ARTICLE SUMMARY

Article focus
■ To describe experiences with the implementation

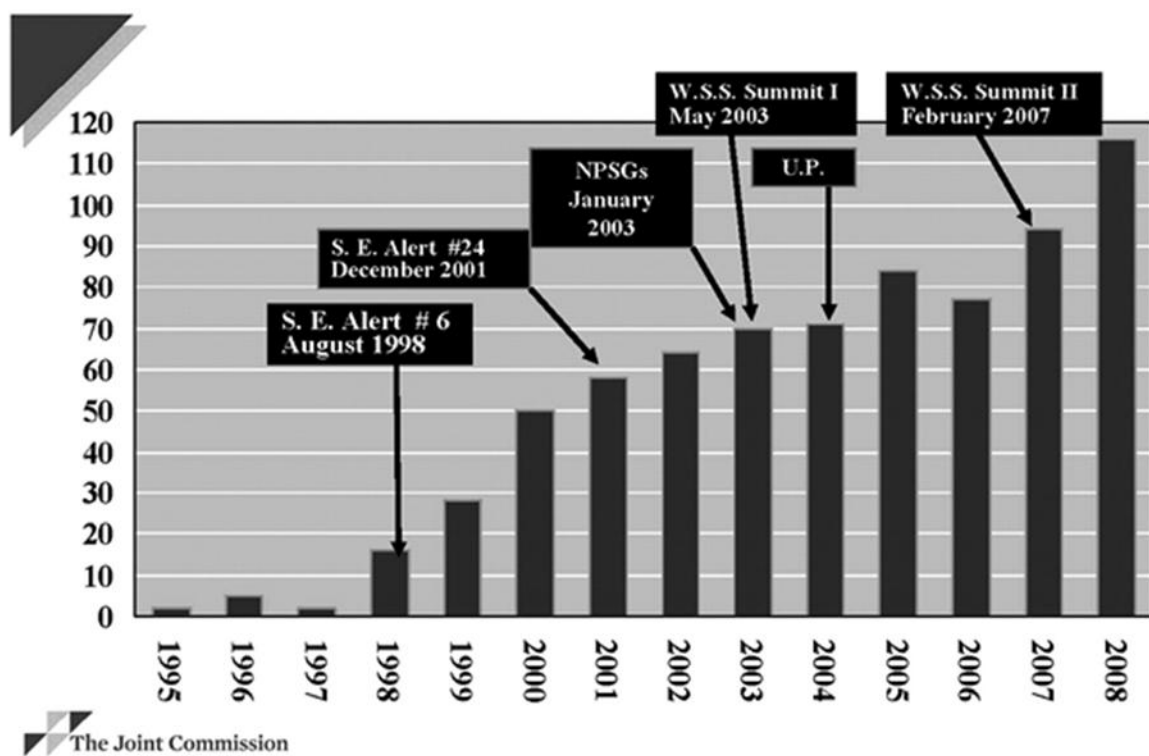
Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study

Helen Hogan,¹ Frances Healey,² Graham Neale,³ Richard Thomson,⁴
Charles Vincent,³ Nick Black¹

- 5% of deaths deemed preventable
- Most problems related to quality of clinical monitoring,
- Most patients whose death was preventable were *older people*



Wrong-site surgeries reported by year.



Pham J C et al. Qual Saf Health Care 2010;19:446-451

“Data for improvement”



Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed



Studying large-scale programmes to improve patient safety in whole care systems: Challenges for research

Jonathan Benn*, Susan Burnett, Anam Parand, Anna Pinto, Sandra Iskander, Charles Vincent

Department of Biosurgery and Surgical Technology, Imperial College London, St Mary's Campus, QEOM Building Praed Street, London W2 1NY, UK

- Insufficient data points
- Lack of sufficient baseline periods
- Changing samples and sampling strategies
- Inadequate annotations of changes

- Hospital-specific rates highly sensitive to adjustment for confounders and unit of analysis (admission or patient)

CMAJ

RESEARCH

Comparing methods to calculate hospital-specific rates of early death or urgent readmission

Carl van Walraven MD MSc, Jenna Wong BSc MSc, Steven Hawken MSc, Alan J. Forster MD MSc

ABSTRACT

Background: Hospital readmissions are important patient outcomes that can be accurately captured with routinely collected administrative data. Hospital-specific readmission rates have been reported as a quality-of-care indicator. However, the extent to which these

ratios adjusted for age and sex alone had the greatest variation. Within hospitals, ranges of the 4 ratios averaged 31% of the overall estimate. Readmission ratios adjusted for age and sex showed the lowest correlation (Spearman correlation coefficient 0.48–0.68).

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Carl van Walraven, carl.vanwalraven@utoronto.ca

Judging quality and safety

- Three major rankings of US hospitals
- MGH gets A from Leapfrog, ranked top by US News and Word report, but gets 45 out of 100 from Consumer Reports
- Bottom six in the CR ranking all got A from Leapfrog
- <http://blogs.sph.harvard.edu/ashish-jha/hospital-rankings-get-serious/>

How Well Can We Identify the High- Performing Hospital?

**Michael Schwartz^{1, 2}, Alan B. Cohen^{1, 2},
Joseph D. Restuccia^{1, 2}, Z. Justin Ren²,
Alan Labonte^{1, 2}, Carol Theokary²,
Raymond Kang³, and Jedediah Horwitt²**

Medical Care Research and Review

68(3) 290–310

© The Author(s) 2011

Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/1077558710386115

<http://mcr.sagepub.com>



Only a handful of consistently high performing hospitals,
and may be a chance finding

The story of one UK hospital

- Regulator rated this hospital as one of four “most improved” hospitals in 2006/7
- Based on self-assessment against core indicators, provisionally rated “good” in 2007/8
- Dr Foster’s Good Hospital Guide (2009) identified it as among 5 most improved over last three years
- November 2009 – ranked in best 10 in league tables for HSMR

The story of one UK hospital

In 2009, the Care
Quality Commission
said it was “appalling”

The Inquiry

“Some of the treatment of elderly patients could properly be characterised as abuse of vulnerable persons.”



The story of one UK hospital

“It soon became clear that the real position of the hospital in the national league of awfulness did not matter. What did matter was that many patients had received poor care and, for some, their treatment was appalling.”

- Dr Paul Woodmansey
<http://www.hospitaldr.co.uk/blogs/tag/mid-staffordshire>

What about process measures instead?

- Be careful what you measure
- Medicare policy on tight blood glucose control in ICU patients turned out to be wrong
- Health Affairs (March 2009 issue) – conforming to quality guidelines had no impact on outcomes



Problems of performance measurement

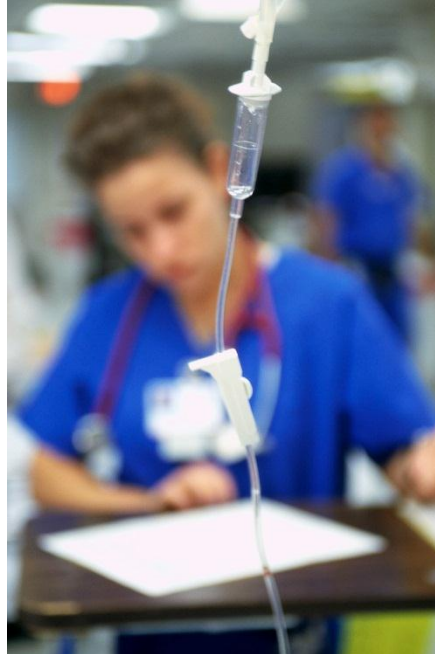
- Kelman and Friedman (2009):
 - Effort substitution
 - Gaming



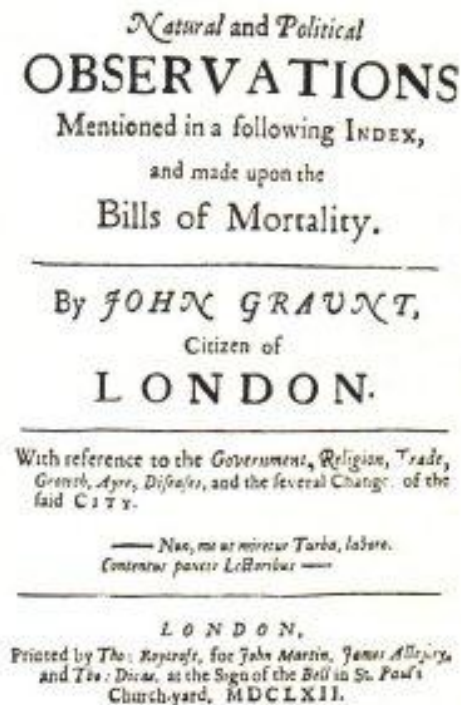
- Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service *J Public Adm Res Theory* (2009) 19 (4): 917-946

Effort substitution

- When people direct their attention to the thing being measured at the expense of other valuable activities that are not measured



Gaming: deliberate attempts at manipulation. It goes back a long way



Under-counting of plague deaths to appease municipal authorities

But is there more to the problem of measuring quality and safety than gaming and effort substitution?

The Health Foundation's Lining Up Research project

- An ethnographic study of interventions to reduce central line infections
- What happens when organisations are asked to interpret data definitions, collect data and report on CVC-BSIs?



The NEW ENGLAND
JOURNAL *of* MEDICINE

ESTABLISHED IN 1812

DECEMBER 28, 2006

VOL. 355 NO. 26

An Intervention to Decrease Catheter-Related Bloodstream
Infections in the ICU

Peter Pronovost, M.D., Ph.D., Dale Needham, M.D., Ph.D., Sean Berenholtz, M.D., David Sinopoli, M.P.H., M.B.A.,
Haitao Chu, M.D., Ph.D., Sara Cosgrove, M.D., Bryan Sexton, Ph.D., Robert Hyzy, M.D., Robert Welsh, M.D.,
Gary Roth, M.D., Joseph Bander, M.D., John Kepros, M.D., and Christine Goeschel, R.N., M.P.A.

Not even that programme eliminated CVC-BSIs

- **Mean rate of CVC-BSIs**
 - **7.7** infections per 1,000 CVC-patient days at baseline
 - **2.3** at 0 to 3 month after implementation ($p \leq 0.002$)
 - **1.4** during 18 months of follow up (median = 0)
 - Interquartile range 0-2.4

Measurement

THE
MILBANK QUARTERLY
A MULTIDISCIPLINARY JOURNAL OF POPULATION HEALTH AND HEALTH POLICY

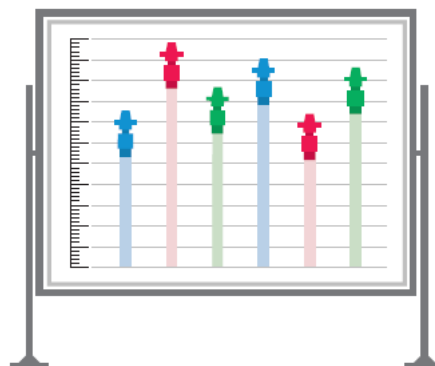
What Counts? An Ethnographic Study
of Infection Data Reported to a Patient Safety
Program

MARY DIXON-WOODS,¹ MYLES LESLIE,² JULIAN BION,³
AND CAROLYN TARRANT¹

- Units were not counting either denominators or numerators consistently
- Wide variability in underlying clinical practices and laboratory support

Lining Up: How is harm measured?

Lessons from an ethnographic research study
of interventions to reduce central line infections



Learning report
February 2013

DEFINITIONS FOR BLOOD STREAM INFECTION, CATHETER-LINKED INFECTION, AND CENTRAL VENOUS CATHETER

LABORATORY-CONFIRMED BLOOD STREAM INFECTION¹ *must meet at least one of the two criteria below*

Criterion 1	<ul style="list-style-type: none"> • Patient has one or more recognized pathogens cultured from ≥ 1 blood culture
Criterion 2	<p>If the microorganism is a common skin organism (ie, diphtheroids [<i>Corynebacterium</i> spp], <i>Bacillus</i> [not <i>B anthracis</i>] spp, <i>Propionibacterium</i> spp, coagulase-negative staphylococci [(CNS), excludes sensitive <i>Staph aureus</i>], viridans group streptococci, <i>Aerococcus</i> spp, <i>Micrococcus</i> spp), <i>then...</i></p> <ul style="list-style-type: none"> • It must have been cultured from 2 or more blood cultures drawn on separate occasions, or from one blood culture in a patient in whom antimicrobial therapy has been started, <u>and</u> • Patient has ≥ 1 of the following: fever of $>38^{\circ}\text{C}$, chills, or hypotension²

CATHETER-ASSOCIATED BLOOD STREAM INFECTION (CABSI)²

Criterion	<ul style="list-style-type: none"> • One of the criteria for BSI above, <u>and</u> • The presence of one or more central venous catheters at the time of the blood culture, or up to 48 hrs following removal of the CVC <u>and</u> • The signs & symptoms & positive laboratory results including pathogen cultured from the blood are not primarily related to an infection at another site.
-----------	---

CATHETER-RELATED BLOOD STREAM INFECTION (CRBSI)²

Criterion	<ul style="list-style-type: none"> • One of the criteria for BSI above, <u>and</u> • The presence of one or more central venous catheters at the time of the blood culture, or up to 48 hrs following removal of the CVC, <u>and</u> • One of the following: <ol style="list-style-type: none"> i. a positive semiquantitative (>15 CFU/catheter segment) or quantitative ($>10^3$ CFU/ml or $>10^3$ CFU/catheter segment) culture whereby the same organism (species and antibiogram) is isolated from blood sampled from the CVC or from the catheter tip, and peripheral blood; ii. simultaneous quantitative blood cultures with a $>5:1$ ratio CVC versus peripheral.
-----------	--

CATHETER-SUSPECTED BLOOD STREAM INFECTION

Criterion	<ul style="list-style-type: none"> • NEGATIVE blood cultures in the presence of parenteral antimicrobials, <u>and</u> • Clinical evidence of a systemic response to infection, <u>and</u> • Clinical condition improves following removal of CVC, <u>and</u> • No other likely source of infection
-----------	--

CENTRAL VENOUS CATHETER (CVC)

Criterion	<ul style="list-style-type: none"> • An intravascular device terminating in one of the great veins or pulmonary artery, including those in, or near, the right atrium, and those inserted via a femoral vein. • Includes PICCs, haemodialysis catheters, parenteral nutrition catheters
-----------	---

Denominators

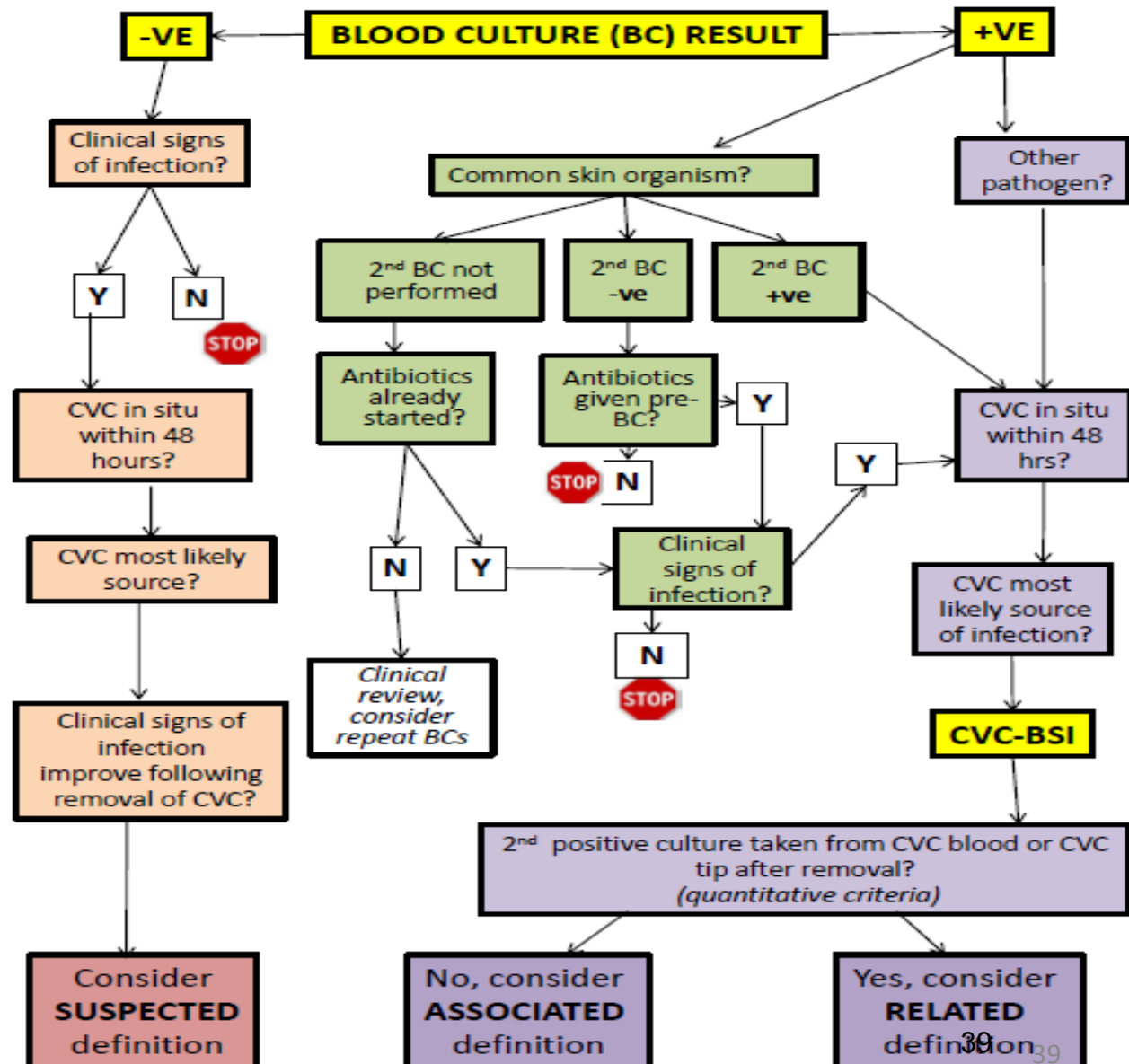
- Staff perceived that patients with different risk profiles were lumped together
- Sometimes excluded patients thought to be “low risk” or “high risk”
- Perceptions of fairness were very important



Clinical practices

- Some physicians started anti-microbial therapy without a blood sample
- What ICUs sent to the lab varied enormously
- Organisational systems meant samples were not always matched up

Numerators



Data collection systems

Controller centered	Track-trigger Track	Patrol
11	3	3
Highly fallible	Highly reliable	Reasonable reliability
Variable local credibility	High local credibility	Low local credibility



Differences in microbiology support

- Many could not support catheter-related definition
- Microbiology involvement in rounds varied
- Contribution to decision-making about what counted varied



Link between measurement and improvement

- High rates could motivate action – but only if credible
- Low rates sometimes induced unjustified complacency
- Credible data is a must
 - *If I'm honest right before we started, we didn't think we were that bad. [...] We thought, you know, [we] don't really have a problem with central line infections. But I think what it was, nobody ever looked to see whether we were any good [...] and when we compared our infection rates, actually they were far worse than any of us ever realised. (Senior nurse, participant 43)*

Data collection

- Data collection risked goal displacement
 - Effort to get the numbers risked displacing goal of cultural and systemic change to improve infection control
- Numbers could be a wake-up call *if done well and if there was a problem*
- Numbers could also used to reinforce status quo and apathy

COMMENTARY

When Counting Central Line Infections Counts

Mary Dixon-Woods, MSc, DPhil;¹ Eli N. Perencevich, MD, MS²

(See the article by Wise et al, on pages XXX–XXX.)

In this issue of the journal, Wise et al¹ report encouraging news of significant declines in reported rates of central venous catheter (CVC) bloodstream infections (BSIs) in critical care units over the past 2 decades. Using data from the US Centers for Disease Control and Prevention's (CDC's) National Nos-

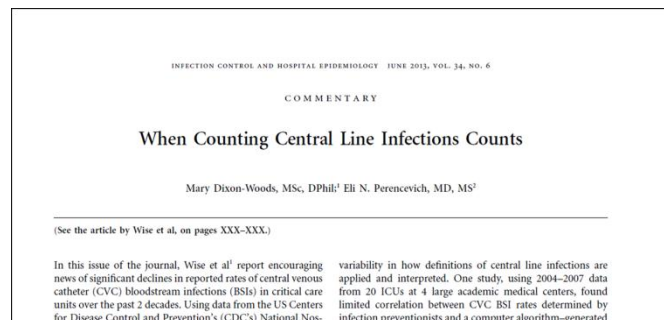
variability in how definitions of central line infections are applied and interpreted. One study, using 2004–2007 data from 20 ICUs at 4 large academic medical centers, found limited correlation between CVC BSI rates determined by infection preventionists and a computer algorithm-generated

Counting counts

- CDC definitions aimed at maximising sensitivity
- Financial and reputational penalties changing rules of game
- But accuracy of reported rates in question – one study found external validation raised reported rates by 27%

Counting counts

- Goodhart's law – any observed statistical regularity will tend to collapse once pressure is placed on it for control purposes
- Infection rates – willingness to report or real rate?



Targets and terror

- Centralised performance management of NHS
- “Targets and terror” regime for NHS in England throughout 2000s
- People become adept at working out what they need to do to survive performance management

WHAT'S MEASURED IS WHAT MATTERS: TARGETS AND GAMING IN THE ENGLISH PUBLIC HEALTH CARE SYSTEM

GWYN BEVAN AND CHRISTOPHER HOOD

In the 2000s, governments in the UK, particularly in England, developed a system of governance of public services that combined targets with an element of terror. This has obvious parallels with the Soviet regime, which was initially successful but then collapsed. Assumptions underlying governance by targets represent synecdoche (taking a part to stand for a whole); and that problems of measurement and gaming do not matter. We examine the robustness of the regime of targets and terror to these assumptions using evidence from the English public health service on reported successes, problems of measurement, and gaming. Given this account, we consider the adequacy of current audit arrangements and ways of developing governance by targets in order to counter the problems we have identified.



CORRESPONDENCE

Research Correspondence

Eroding the Denominator

The Incomplete Story of Door-to-Balloon Time Reporting

To the Editor: Door-to-balloon time (DTB) is the focus of national guidelines, and intense efforts to reduce it have been a core component of many national quality improvement campaigns. Recently, it has been the high point of advertisement campaigns by hospital systems and ultimately might be used to reward performance by payers. It is now publicly reported by the Centers for

Medicare and Medicaid, and the U.S. Department of Health and Human Services. However, not all patients are eligible for DTB reporting. Patients who die before gaining vascular access, and 0.7% ($n = 1$) would be excluded for delay in obtaining consent from the patient. Finally, 7.3% ($n = 11$) of in-hospital deaths would be excluded from reporting for "other" non-system delays. In comparison, only 44% of all deaths ($n = 66$) were outside of the current parameters for exclusion and thus eligible for DTB reporting (Fig. 1).

Eroding the denominator

- *Exclusions for reporting [may be] ...obscuring the current drivers of in-hospital mortality instead of helping focus attention on them. A key consideration for the future will be to dissociate these measures of performance from reimbursement, so as to allow for comprehensive reporting and data collection without the threat of punishment.*

Performance management by numbers

- Whether performance is enhanced or obstructed depends on the purpose to which numbers are put:
 - Targets
 - Rankings
 - Intelligence

<p>Christopher Hood All Souls College, Oxford University, United Kingdom</p>	
<p>Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses</p>	
<p><i>"Public management by numbers" has experienced an international policy boom in recent decades, and big claims have been made about its performance-enhancing effects. But it is hard to assess such claims systematically, even though we can find dramatic anecdotes of</i></p>	<p>and how or when they do it? (White 2012, 115). Such questions will always be central to the analysis of performance measurement, but it can be argued both that performance enhancement can sometimes be achieved with numbers that fall short of perfect</p>
<p>Christopher Hood is Gladstone Professor of Government and a fellow of All Souls College, Oxford University, and has studied government and public administration for four decades. He is a fellow of the National Academy of Public Administration and won the 2011 William E. Mosher and</p>	

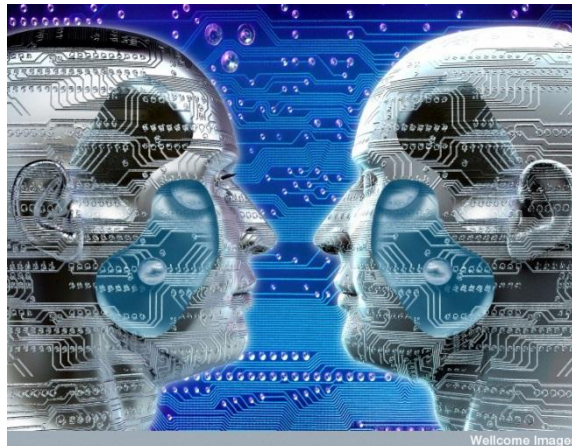
Measurement done badly

- Illusion of control
- Blindsight



Intelligence

- If you're not measuring, you're not managing
- If you're measuring stupidly, you're not managing
- If you're only measuring, you're not managing



Intelligence

- No single indicator will tell you whether care is safe
- Need multiple methods for problem-sensing
- Need ways of discovering fugitive knowledge in organisations
- More use of pro-active diagnostic tools

Conclusions

- Measurement is essential but we have to get better at it
- Current methods very limited in telling us
 - whether care is safe
 - how safety compares between organisations
 - where and how to intervene