Information in practice

Surgical wound infection as a performance indicator: agreement of common definitions of wound infection in 4773 patients

A P R Wilson, C Gibbons, B C Reeves, B Hodgson, M Liu, D Plummer, Z H Krukowski, J Bruce, J Wilson, A Pearson

Department of Clinical Microbiology, University College London Hospitals, London WC1E 6DB
A P R Wilson *consultant microbiologist*
B Hodgson *registered general nurse*

Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London
C Gibbons *research fellow*
B C Reeves *senior lecturer in epidemiology*

Department of Medical Physics, University College London Hospitals
M Liu *physicist*
D Plummer *head of medical physics*

Department of Surgery, Medical School, Aberdeen
Z H Krukowski *consultant surgeon*

Department of Public Health, Medical School, Aberdeen
J Bruce *research fellow in epidemiology*

continued over

BMJ 2004;329:720–3

Abstract

Objective To assess the level of agreement between common definitions of wound infection that might be used as performance indicators.
Design Prospective observational study.
Setting London teaching hospital group receiving emergency cases as well as tertiary referrals.
Participants 4773 surgical patients staying in hospital at least two nights.
Main outcome measures Numbers of wound infections based on purulent discharge alone, on the Centers for Disease Control (CDC) definition of wound infection, on the nosocomial infection national surveillance scheme (NINSS) version of the CDC definition, and on the ASEPSIS scoring method.
Results 5804 surgical wounds were assessed during 5028 separate hospital admissions. The mean percentage of wounds classified as infected differed substantially with different definitions: 19.2% with the CDC definition (95% confidence interval 18.1% to 20.4%), 14.6% (13.6% to15.6%) with the NINSS version, 12.3% (11.4% to 13.2%) with pus alone, and 6.8% (6.1% to 7.5%) with an ASEPSIS score >20. The agreement between definitions with respect to individual wounds was poor. Wounds with pus were automatically defined as infected with the CDC, NINSS, and pus alone definitions, but only 39% (283/714) of these had ASEPSIS scores>20.
Conclusions Small changes made to the CDC definition or even in its interpretation, as with the NINSS version, caused major variation in estimated percentage of wound infection. Substantial numbers of wounds were differently classified across the grades of infection. A single definition used consistently can show changes in percentage wound infection over time at a single centre, but differences in interpretation prevent comparison between different centres.

Introduction

Surveillance of surgical site infection became mandatory in the NHS for orthopaedics in April 2004, and this will soon spread to other specialties.[1] The feedback of infection data to surgeons clearly reduces infection rates.[2 3] Given that the percentage of wounds classified as infected will probably be used as a performance indicator,[4] the new surveillance system must allow reliable comparisons to be made across NHS institutions, and with overseas health institutions.

There has been little or no critical evaluation of the definition of surgical site infection that is to be used for surveillance in England, namely the nosocomial infection national surveillance scheme (NINSS) version of the definition set out by the Centers for Disease Control (CDC) in 1992.[5] Moreover, the version or interpretation of the definition used varies between hospitals and regions.[6 7] Designers of a national surveillance system must judge the available definitions by their ability to identify infections that matter most to patients and to health services and by the practicability of collecting the required information.

We therefore compared agreement between four common definitions of surgical site infection—namely (*a*) the CDC 1992 definition, (*b*) the NINSS modification of the CDC definition, (*c*) the presence of pus, and (*d*) the ASEPSIS scoring method[8]—applied to the same series of surgical wounds. We also compared the percentage of infection based on the CDC definition and on the NINSS modification to investigate the potential effect of subjective CDC criteria and of variation between hospitals in data collection methods.

Participants and methods

Since May 2000, surgical wound surveillance has been conducted at University College London Hospitals. Cardiac, thoracic, orthopaedic, general, obstetric, gynaecological, urological, maxillofacial, plastic, and vascular surgical specialties have participated, each for at least six months each year. Only patients staying in hospital for at least two nights are included.

ELPS *This is the abridged version of an article that was posted on bmj.com on 14 September 2004: http://bmj.com/cgi/doi/10.1136/bmj.38232.646227.DE*

Comparison of crude rates of surgical site infection reported with Centers for Disease Control (CDC) 1992 definition and with ASEPSIS scoring method. Wounds were considered to be infected if they met the CDC criteria for either superficial or deep infection (top half of table) or if they met the criteria for deep infection only (bottom half of table). Values are numbers (percentages) of wounds, with 95% confidence intervals for percentages, adjusted for multiple wounds in the same patients

	CDC results		
ASEPSIS results	**Uninfected**	**Infected**	**Total**
Wounds with superficial or deep infections according to CDC considered infected			
Uninfected (score ≤20)	4660 (80.3)	750 (12.9)	5410 (93.2, 95% CI 92.5 to 93.9)
Infected (score >20)	28 (0.5)	366 (6.3)	394 (6.8, 95% CI 6.1 to 7.5)
Total	4688 (80.8, 95% CI 79.6 to 81.9)	1116 (19.2, 95% CI 18.1 to 20.4)	5804 (100)
Wounds with deep infections only according to CDC considered infected			
Uninfected (score ≤20)	5218 (89.9)	192 (3.3)	5410 (93.2, 95% CI 92.5 to 93.9)
Infected (score >20)	179 (3.1)	215 (3.7)	394 (6.8, 95% CI 6.1 to 7.5)
Total	5397 (93.0, 95% CI 92.3 to 93.7)	407 (7.0, 95% CI 6.3 to 7.7)	5804 (100)

Agreement statistics: for top half of table, κ=0.43 (95% CI 0.40 to 0.46), Ppos=0.48 (0.45 to 0.52), Pneg=0.92 (0.92 to 0.93); for bottom half, κ=0.50 (0.46 to 0.55), Ppos=0.54 (0.49 to 0.58), Pneg=0.97 (0.96 to 0.97).

Nosocomial Infection Surveillance Unit, HPA Central Public Health Laboratory, London
J Wilson
SSI surveillance programme leader
Health VFM Audit, National Audit Office, London
A Pearson
manager and consultant epidemiologist
Correspondence to: A P R Wilson peter.wilson@uclh.nhs.uk

Information is collected on patients and their surgical wounds, allowing us to apply the different definitions of wound infection.[5 6 8 9]

Definitions of surgical site infection

The 1992 CDC definition requires the observation of 16 wound or patient characteristics in order to classify infection and has two subjective criteria, namely a surgeon's diagnosis of infection and the culture of micro-organisms from the wound.[5] The US national nosocomial infections surveillance system (NNISS) recommends that the latter criterion should be based only on positive cultures of fluid and tissue rather than wound swabs,[5 7] but this interpretation does not seem to be applied generally.[7] The English NINSS method modified the CDC definition to exclude the need for a surgeon's diagnosis and required that pus cells be present to satisfy the criterion of micro-organisms cultured from the wound.[6] Another definition of infection simply requires the presence of pus, even though some infections are missed.[9] ASEPSIS is a quantitative scoring method that provides a numerical score related to the severity of wound infection using objective criteria based on wound appearance and the clinical consequences of the infection.[7 8]

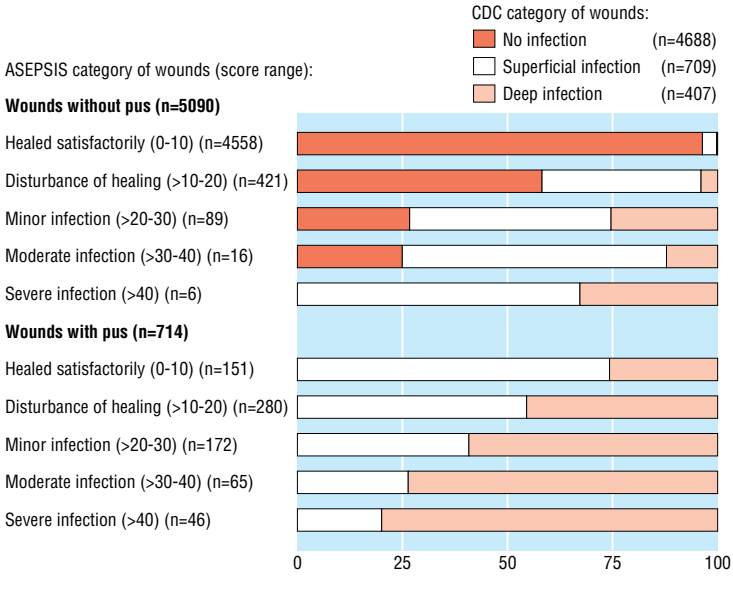
For purposes of comparison, we classified ASEPSIS scores >20 as infected. ASEPSIS scores of 10-20 ("disturbance of healing") are known to describe some infections, but most reflect wound breakdown due to other causes.[10] Moderate to severe infections score >30. The CDC definition also describes the severity of infection, classifying infections as "none," "superficial," or "deep or organ space" (termed "deep" in this article). Both definitions purport to describe the importance of an infection with respect to the patient's morbidity and the likely clinical consequences.

Data collection

Surveillance staff assessed patients every two or three days by direct observation, case note review, and questioning of nurses. We contacted patients by post or telephone one to two months after their operations to complete a questionnaire designed to ascertain late infections. Thus, we followed up patients either until their wounds had healed without infection or until an infection was detected. We therefore classified wounds as infected or not and recorded the proportion of wounds classified as infected at any time during follow up.

Statistical analysis

Counts and percentages presented are of wounds unless otherwise indicated. Confidence intervals for proportions of infection were adjusted for clustering on patient. We summarised agreement between the different definitions of infection using the κ statistic and the proportional agreement of ASEPSIS and CDC respectively for positive (Ppos) and negative (Pneg) diagnoses of infection.[11] Confidence intervals for the agreement statistics were adjusted for clustering on patient and calculated by bootstrap methods. The values shown are "bias-corrected."



Comparison of diagnoses of surgical site infection in 5804 wounds reported with Centers for Disease Control (CDC) 1992 definition and with ASEPSIS scoring method, for wounds with and without pus

Results

A total of 5804 surgical wounds in 4773 patients were assessed during 5028 separate hospital admissions to all surgical specialties in the hospital group between May 2000 and July 2003. The patients' median age was 53.5 years (interquartile range 37.5-69.6), and 2281 (48%) of the patients were female. The median hospital stay was 8 days (6-14), and duration of operation 111 minutes (62-180).

The mean percentage of wound infection differed substantially with the different definitions; 19.2% (95% confidence interval 18.1% to 20.4%) with the CDC definition, 14.6% (13.6% to 15.6%) with the NINSS version, 12.3% (11.4% to 13.2%) with pus alone, and 6.8% (6.1% to 7.5%) with an ASEPSIS score >20. The table shows the level of agreement between the ASEPSIS and CDC systems. When superficial infections (according to CDC category) were included, 13% (778) of all observed wounds received conflicting diagnoses, and 6% were classified as infected by both definitions. When superficial infections were excluded, the two definitions estimated about the same overall percentage infection (6.8% and 7.0% respectively), but there were almost twice as many conflicting infection diagnoses (n=371) as concordant ones (n=215).

Wounds with pus were automatically diagnosed as infected by the CDC, NINSS, and pus alone definitions, but only 39% of these (283/714) had ASEPSIS scores >20 (see figure). For these wounds, the CDC scale also consistently diagnosed greater infection severity than did ASEPSIS.

In wounds without pus the relation of ASEPSIS and CDC scales was less consistent (figure). For example, 42% (177/421) of wounds classified only as "disturbance of healing" by ASEPSIS were classified as infected by the CDC definition. Conversely, four of the six wounds classified as "severe wound infections" by ASEPSIS were classified as superficial by the CDC definition.

Comparison of wound classification with the CDC definition and with the NINSS version also revealed discrepancies. For example, more than 30% of wounds defined as superficially infected with CDC were classified as not infected with NINSS (229/709). In the CDC "superficial infection" category 94% (222/237) of the observed discrepancy was attributable to the NINSS modification of the CDC criterion related to positive bacterial cultures. In the CDC "deep infection" category the discrepancy observed was due to the exclusion of infections based solely on a surgeon's diagnosis.

Discussion

The four different definitions of surgical site infection varied widely in the estimated percentage of wounds infected. Comparing the 1992 CDC definition and the ASEPSIS scoring method, we found more than twice as many wounds were classified as infected by only one definition (n=778) as were classified as infected by both (n=366).

Potential limitations of this study

For the CDC definition, we often assumed the requirement for a surgeon's diagnosis of infection to be satisfied when a decision was made to start specific antibiotic treatment or to provide surgical treatment. In other studies, differences in results between CDC and other surveillance methods have been associated with lack of follow up, use of positive culture results, or clinical criteria.[12] Although our study was conducted in a single group of hospitals, data came from multiple sites, many surgical specialties, and a large number of surgeons, so that most of the relevant sources of variation were represented.

Comparison of the different definitions

Other health outcome measures have been psychometrically evaluated, but similar information is lacking for most definitions of wound infection.[13] ASEPSIS in its original form was reported to be repeatable and related to outcome,[10 14] but it has since been modified and reproducibility is currently being reassessed. Choosing an optimal definition is therefore extremely difficult.

The absence of a clear pattern to the type of wounds classified as infected by CDC but as not infected by NINSS supports the view that the CDC criteria responsible for the discrepancy are difficult to apply consistently. Small changes made to the CDC definition or even to its interpretation, as with the NINSS version, cause substantial variation in the apparent percentage of infected wounds. Although the CDC definition has been adopted in many countries to allow international comparison, this faith seems unwarranted.

Conclusions

Using wound infection rates as a performance indicator to compare centres or countries is premature. Without a means to interpret absolute rates, such comparisons will be compromised by discrepancies in the way that infections are defined. External agencies

What is already known on this topic

Surgical site infections are a major cause of morbidity and increased costs in health care

The percentage of surgical wounds classified as infected is an obvious potential performance indicator, but common definitions have not been validated or compared

What this study adds

To assess the robustness of four common definitions of wound infection, their agreement in wound classification was determined

Classifications with different definitions disagreed for more than twice as many wounds as those for which they agreed, and small changes in the interpretation of a definition caused substantial variation in the percentage of wounds classified as infected

Although feedback of rates of wound infection within an institution using a consistent definition is effective in reducing infection rates, infection rates cannot be used as a performance indicator to compare hospitals without a more robust definition

should not judge the quality of medical care on these measures.[15] Comparative performance tables should be reported only once a scientifically based and agreed definition has been produced.

We thank the members of the wound surveillance team (D Archibald, J Leach, and E O'Donnell).

Contributors: See bmj.com

Funding: Wound surveillance was supported by a start up grant from UCLH trustees for the first two years. Subsequently it has been funded directly by UCLH Trust. The analysis by CG reported in this paper has been funded by a grant from the National NHS R&D Research Methodology Programme. None of these funding sources have contributed to, or influenced the interpretation of, the analyses reported.

Competing interests: None declared.

Ethical approval: This was not deemed necessary as the surveillance was part of the hospital audit programme.

1 Donaldson L, Mullally S. *Surveillance of healthcare associated infections. PLCMO2003/4, PLCNO2003/4.* London: Department of Health, 2003.
2 Gil-Egea MJ, Pi-Sunyer MT, Verdaguer A, Sanz F, Sitges-Serra A, Eleizegui LT. Surgical wound infections: prospective study of 4,468 clean wounds. *Infect Control* 1987;8:277-80.
3 Mead PB, Pories SE, Hall P, Vacek PM, Davis JH Jr, Gamelli RL. Decreasing the incidence of surgical wound infections. Validation of a surveillance-notification program. *Arch Surg* 1986;121:458-61.
4 Department of Health. Indicator 23003. In: *NHS performance indicators: a consultation.* London: DoH, 2001: Annex 1, p57.
5 Horan TC, Gaynes RP, Martone WJ, Jarvis WR, Emori TG. CDC definitions of nosocomial surgical site infections, 1992: a modification of CDC definitions of surgical wound infections. *Infect Control Hosp Epidemiol* 1992;13:606-8.
6 Wilson JA, Ward VP, Coello R, Charlett A, Pearson A. A user evaluation of the nosocomial infection national surveillance system: surgical site infection module. *J Hosp Infect* 2002;52:114-21.
7 Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. *Health Technol Assess* 2001;5(22): 1-194.
8 Wilson AP, Treasure T, Sturridge MF, Gruneberg RN. A scoring method (ASEPSIS) for postoperative wound infections for use in clinical trials of antibiotic prophylaxis. *Lancet* 1986;i:311-3.
9 Cruse PJ, Foord R. The epidemiology of wound infection. A 10-year prospective study of 62,939 wounds. *Surg Clin North Am* 1980;60:27-40.
10 Armitage P, Berry G, Matthews JNS. *Statistical methods in medical research.* 4th ed. Oxford: Blackwell, 2001:703.
11 Wilson AP, Helder N, Theminimulle SK, Scott GM. Comparison of wound scoring methods for use in audit. *J Hosp Infect* 1998;39:119-26.
12 Colson M, Bolsin S. The use of statistical process control methods in monitoring clinical performance. *Int J Qual Health Care* 2003;15:445.
13 Smith BH, Penny KI, Purves AM, Munro C, Wilson B, Grimshaw J, et al. The chronic pain grade questionnaire: validation and reliability in postal research. *Pain* 1997;71:141-7.
14 Byrne DJ, Malek MM, Davey PG, Cuschieri A. Postoperative wound scoring. *Biomed Pharmacother* 1989;43:669-73.
15 Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147-54.

(Accepted 4 August 2004)

A memorable patient

The white coated man's burden

Mrs Jones was a truly delightful woman in her late 60s who struggled with the problems posed by diabetes, hypertension, angina, asthma, peptic ulceration, gout, rheumatoid arthritis, and a modest degree of renal impairment. Unsurprisingly, she was usually markedly symptomatic from one or more of these conditions. As a newly appointed medical registrar, I encountered her in my first outpatient clinic. I saw somebody who was clearly disabled trying to make the best of it but obviously becoming pretty low spirited in consequence of the constant struggle. My role, I knew, was to relieve her plight and restore her to a pleasurable and fulfilling life, so I set about tackling her multiplicity of problems.

Over the weeks, I improved her joint pain and her angina, but the drugs aggravated her ulcer and worsened her asthma. The worrying level of hypertension could, of course, be treated—but then the diabetes, gout, and renal function deteriorated. I grappled with the diabetes, the "improved" control I achieved coinciding with one of her periods of inability to eat because of the abdominal pain caused by her arthritis drugs. She had a nasty hypoglycaemic episode, and the decision to let her blood glucose level run rather higher, after all, led to her becoming dehydrated. The renal function went off like a rocket, and gout laid her low. And so it went on.

As the weeks turned to months, I saw Mrs Jones in the clinic with increasing frequency. I was dismayed by my inability to control her symptoms, and she was dismayed that she was unable to report an improvement to the young doctor who was trying so hard to help her. She began to bring in little gifts, such as eggs from her chickens, to keep my spirits up. These served only to intensify my sense of guilt at having failed her.

As a last resort, I asked my consultant for advice. I'd always done so as a senior house officer, but had vaguely thought there might be something in the rules that said that registrars weren't supposed to need to. Certainly, that was the impression registrars had given me when I was a house officer. How could I render poor, brave Mrs Jones asymptomatic? "Do you know, I really don't think you can," was the consultant's uncompromising, and—to my ears that day—rather unfeeling reply.

I reflected carefully, though, and light slowly dawned. My responsibility, I realised, was to do my very best to help her—including making use of all the advice available to me. If I had done that, and done it conscientiously, it was not my personal fault that medical science had not advanced to the point of being able to cure Mrs Jones and those like her. I should understand, sympathise, and medically do everything I could—but I should not carry a burden of guilt if all had not been solved.

I suspect that most doctors realise this after a while, even if they have never clearly formulated it in their minds. Equally, I suspect that concerns similar to those that I experienced with Mrs Jones trouble a greater number of junior doctors than their seniors might realise or remember. As for Mrs Jones, we both set more realistic goals for symptom control. Or, rather, I did. Her goals had probably always been more realistic than mine, but she had been too polite to mention it.

John Moore-Gillon *consultant physician, Bart's and the London Trust, London (john.moore-gillon@bartsandthelondon.nhs.uk)*

We welcome articles up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake,* or any other piece conveying instruction, pathos, or humour. Please submit the article on http://submit.bmj.com Permission is needed from the patient or a relative if an identifiable patient is referred to. We also welcome contributions for "Endpieces," consisting of quotations of up to 80 words (but most are considerably shorter) from any source, ancient or modern, which have appealed to the reader.