K. Taxis*, V. Wirtz, N. Barber
*Department of Practice and Policy, The School of Pharmacy, University of London, London, UK*
*E-mail address:* katja.taxis@uni-tuebingen.de

*Corresponding author. Address: Pharmazeutische Biologie, Pharmazeutisches Institut, Auf der Morgenstelle 8, 72076 Tubingen, Germany. Tel.: +49-70712975483; fax: +49-7071295250
doi:10.1016/j.jhin.2003.10.006

## Missing data and participation or non-participation in a surgical-site infection surveillance network

Sir,

Missing data reflect the quality of information and may bias epidemiological studies, including the surveillance of hospital-acquired infections (HAI). The method of data collection may contribute to the frequency of missing data. Data on HAI surveillance can be obtained through a network or from isolated surveillance performed by one unit. The objective of our study was to compare the proportion of missing data according to participation or non-participation in a surgical-site infection (SSI) surveillance network.

A voluntary, patient-based surveillance network was started in 1993 with the objective of monitoring SSI in surgical units of south-eastern France.[1] Data on infected and non-infected patients were collected in a standardized way and sent to the co-ordination centre for validation and analysis. For the purpose of this study, data from three units in Lyon University Hospital were available. Two of them were part of the network:[1] an orthopaedic surgery unit (unit A) and a gynaecology–obstetrics unit (unit B). The third, a digestive surgery unit (unit C), implemented its own surveillance programme. We identified 20 common items between the two surveillance methods [network (units A and B), and extra-network (unit C)] and missing values were counted for these items during common surveillance periods: January 1999, April to June 1999, and January to April 2000. Comparisons of proportions of missing values were based on the Chi-square test or Fisher's exact test. $P < 0.05$ values were considered to be statistically significant.

The total sample size was 1781. The study population was composed of 719 patients from unit A, 322 from unit B, and 740 from unit C. The number and percentages of missing values are shown in Table I.

For unit C, the surveillance database contained several missing items. Only five items out of 20 were complete, and known or suspected risk factors

**Table I** Number and proportion of missing values according to participation in a surgical-site infection surveillance (SSI) network

| Variables | Unit A ($N = 719$) | Unit B ($N = 322$) | Unit C ($N = 740$) | $P$[a] |
|---|---|---|---|---|
| Age | 0 | 0 | 0 | - |
| Sex | 0 | 0 | 0 | |
| Date of entry | 0 | 0 | 0 | |
| Date of discharge | 0 | 0 | 9 (1.2%) | <0.001 |
| Date of surgery | 0 | 0 | 0 | - |
| Surgery code | 0 | 0 | 0 | - |
| Multiple surgery | 3 (0.4%) | 0 | 30 (4.0%) | <0.000001 |
| Emergency | 3 (0.4%) | 0 | 22 (3.0%) | <0.00001 |
| Implant/prosthesis | 5 (0.7%) | 0 | 32 (4.3%) | <0.000001 |
| Keyhole surgery | 1 (0.1%) | 0 | 28 (3.8%) | <0.000001 |
| Contamination class | 3 (0.4%) | 0 | 41 (5.5%) | <0.000001 |
| American Society of Anesthesiologist (ASA) score | 7 (1.0%) | 6 (5.0%) | 37 (5.0%) | <0.01 |
| Duration of surgery | 39 (5.4%) | 3 (0.9%) | 88 (11.9%) | <0.000001 |
| Antibiotic prescription for operation | 23 (3.2%) | 4 (1.2%) | 33 (4.4%) | <0.05 |
| Starting date of antibiotic prophylaxis[b] | 0 | 0 | 15 (2.7%) | <0.00001 |
| Duration of antibiotic prophylaxis[b] | 0 | 0 | 117 (21.2%) | <0.000001 |
| Surgical site infection (SSI): yes/no | 0 | 0 | 5 (0.7%) | <0.05 |
| Starting date of SSI[c] | 0 | 0 | 5 (21.7%) | 0.15 |
| Site of infection[c] | 0 | 0 | 3 (13.0%) | 0.55 |
| NNIS index | 44 (6.1%) | 19 (5.9%) | 103 (13.9%) | <0.000001 |

[a] Chi-square test or Fisher's exact test between units participating in the network analysed together (unit A + unit B) *versus* the unit not in the network (unit C).
[b] Denominators were 666 in unit A, 129 in unit B, and 552 in unit C.
[c] Denominators were 5 in unit A, 6 in unit B, and 23 in unit C.

for SSI were missing for a high percentage of cases (i.e. 11.9% of missing data on the duration of surgery) in unit C. Likewise, date of SSI occurrence and site of infection were often missing (21.7% and 13.0%, respectively). Statistical differences for frequency of missing values were found whether the data were collected in a surveillance network (units A and B) or by independent surveillance (unit C) for 13 items out of 20 (*P*-values ranged from 0.05 to $<10^{-6}$; Table I). In addition, missing data relating to the National Nosocomial Infection Surveillance (NNIS) index were more frequent with non-participation in the network ($P < 10^{-6}$). However, differences could occur because of various surgical specialties. We compared the percentages of missing data between units A and B, which participated in the network. We noted only one statistically significant difference between units A and B for the proportion of missing items ('duration of surgery', $P < 0.001$). Thus, we concluded that the proportion of missing values was related more to network participation than to surgical specialty.

Our findings show that participation in a SSI surveillance network protects, at least partially, against a high frequency of missing data. Data quality is crucial in obtaining relevant results and for any subsequent control measures.[2]

Strict adherence to the basic conditions for data collection is, obviously, vital. Participation in a network will encourage participants to comply with a methodology and to complete data collection sheets accurately.

To the best of our knowledge, such descriptive analyses of missing data are not reported frequently, even in the field of hospital epidemiology. In most epidemiological studies, the proportion of missing values is reported, but often not discussed, except for controlled clinical trials.[3] Indeed, when data are missing, epidemiologists know that analyses based on participants with complete data[4] can be biased unless it can be assumed that the probability of missing items is unrelated to the values of the missing binary outcomes (i.e. SSI, yes/no).

Missing items have a dual impact, because data could be analysed separately and could be included in a composite predictive score such as the NNIS index.[5] Then, in the case of missing data included in the NNIS index, the predictive index will not be used in that subgroup of patients who might have specific characteristics, including being at higher risk of infection.

In multivariate analysis, deleting from the analysis subjects on whom information is missing for any of the relevant characteristics could be inappropriate. As stated by Miettinen,[6] with 20 characteristics involved, and with 50% of the subjects having information missing, up to 97.5% of the expected information may be available. Deletion of 50% of subjects from the analysis, when only 2.5% of information is missing, is disproportionate.

One limitation of our study is its small size, which makes the generalization of these results difficult. Similar analysis of larger databases would be interesting.

In conclusion, these results highlight the possible benefits of network participation for improving data quality in SSI surveillance. Unit C agreed to participate in the network after discussion of the study results.

## References

1. Hajjar J, Monnet D, Startor C, Savey A, Derrieu R, Fabry J. Infection du site opératoire: naissance d'un réseau de surveillance dans le Sud-Est de la France. *Hygiènes* 1993;**3**:12—16.
2. Haley RW, Culver DH, White JW, *et al*. The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals. *Am J Epidemiol* 1985;**121**:182—205.
3. Magder L. Simple approaches to assess the possible impact of missing outcome information on estimates of risk ratios, odds ratios, and risk differences. *Control Clin Trial* 2003;**24**:411—421.
4. Rothman K, Greenland S. Fundamentals of epidemiologic analysis. In: Rothman K, Greenland S, editors. *Modern epidemiology*, 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1998. p. 207—208.
5. Smyth ETM, Emmerson AM. Surgical site infection surveillance. *J Hosp Infect* 2000;**45**:173—184.
6. Miettinen O. Regression analysis. In: Miettinen O, editor. *Theoretical epidemiology*. New York: Wiley; 1985. p. 231—233.

G. Scélo[a], L. Gambotti[a], J.-C. Boulez[b], J. Bejui-Hugues[c], J.-P. Carret[c], C. Chémorin[d], M. Berland[e], A.-M. Dumas[e], G. Bégou[c], B. Duchemin-Pelletier[b], R. Girard[f], J. Fabry[a,d], P. Vanhems[a,d,*]

[a]*Epidemiology and Public Health Laboratory, INSERM U271, Claude Bernard-Lyon 1-University, Lyon, France*

[b]*Digestive Surgery Unit, Edouard Herriot Hospital, Lyon, France*

[c]*Orthopaedic Surgery Unit, Edouard Herriot Hospital, Lyon, France*

[d]*Laboratory of Hygiene, Edouard Herriot Hospital, Lyon, France*

[e]*Gynaecology-Obstetrics Unit, Lyon Sud Hospital, Lyon, France*

[f]*Laboratory of Hygiene, Lyon Sud Hospital, Lyon, France*

*E-mail address:* philipva@lyon-sud.univ-lyon1.fr

*Corresponding author

doi:10.1016/j.jhin.2003.09.012