

Validation of Hospital Administrative Dataset for adverse event screening

Sandra Verelst,¹ Jessica Jacques,² Koen Van den Heede,¹ Pierre Gillet,² Philippe Kolh,² Arthur Vleugels,¹ Walter Sermeus¹

► Supplementary appendices are published online only. To view these files please visit the journal online (<http://qshc.bmj.com>)

¹Center for Health Services and Nursing Research, University Hospital of Liège, Liège, Belgium

²Department of Medico-Economic Information, University Hospital of Liège, Liège, Belgium

SV and JJ both contributed equally to this study.

Accepted 19 December 2009
Published Online First
27 April 2010

ABSTRACT

Objective To assess whether the Belgian Hospital Discharge Dataset (B-HDDS) is a valid source for the detection of adverse events in acute hospitals.

Design, setting and participants Retrospective review of 1515 patient records in eight acute Belgian hospitals for the year 2005.

Main outcome measures Predictive value of the B-HDDS and medical record reviews and degree of correspondence between the B-HDDS and medical record reviews for five indicators: pressure ulcer, postoperative pulmonary embolism or deep vein thrombosis, postoperative sepsis, ventilator-associated pneumonia and postoperative wound infection.

Results Postoperative wound infection received the highest positive predictive value (62.3%), whereas postoperative sepsis and ventilator-associated pneumonia were rated as only 44.2% and 29.9% respectively. Excluding present on admission from the screening substantially decreased the positive predictive value of pressure ulcer from 74.5% to 54.3%, as pressure ulcers present on admission were responsible for more B-HDDS-medical record mismatches than any other indicator. Over half (56.8%) of false-positive cases for postoperative sepsis were due to a lack of specificity of the ICD-9-CM code, whereas in 58.6% of false-positive cases for ventilator-associated pneumonia, clinical criteria appeared to be too stringent.

Conclusions The B-HDDS has the potential to accurately detect some but not all adverse events. Adding a code 'present on admission' and improving the ICD-9-CM codes might already partially improve the correspondence between the B-HDDS and the medical record review.

INTRODUCTION

The report 'To Err is Human' created awareness among the general public that adverse events are common and one of the leading causes of mortality within the USA.¹

Adverse events are defined as 'an unintended injury or complication which results in disability, death or prolongation of hospital stay and is caused by healthcare management rather than the patient's disease.'² Methods for identifying events are voluntary reporting, direct observation of healthcare personnel, computerised screening algorithms, screening administrative data and retrospective chart review.³ Medical records are generally considered as the gold standard.⁴ They contain rich clinical details to identify various medical injuries and allow the analysis of causes of errors. A significant limitation, however, is that medical

records are mostly in paper or electronic format that is not readily accessible for research. Moreover, the quality of medical records is highly variable, and some information, such as the effect of an adverse event, is not generally recorded.^{5 6} Moreover, transforming medical records into research data is expensive and resource-intensive.⁴

Sources that are increasingly being used are administrative data, containing demographic characteristics, length of stay, and diagnoses and procedures based on International Classification of Diseases.^{4 7} They are readily available and inexpensive, and provide insight into the characteristics of large populations of patients.⁷

However, secondary diagnoses codes were originally never intended to measure adverse events, being originally created to assist in describing the prevalence of major causes of morbidity and mortality, and later adapted for use in hospital reimbursement with the advent of prospective payment. Lacking in detailed standard clinical definitions universally applied by medical record coders, the coding system is open to clinical and coding interpretation. Coders working with hospital discharge records also depend on what is dictated in physicians' discharge summaries to guide them in coding.⁷ As a result, the accuracy and reliability of administrative data in describing adverse events have been repeatedly questioned.^{7 8}

The main objective of this study was to assess whether the Belgian Hospital Discharge Dataset (B-HDDS) is a valid source for detecting adverse events at the individual patient level in acute Belgian hospitals.

METHODS

Definition and selection of adverse events

Of two well-recognised sets of patient safety indicators—the Agency for Healthcare Research and Quality (AHRQ) patient safety indicators⁹ and the Organization for Economic Co-operation and Development (OECD) set of patient safety indicators—we selected five adverse events based on prevalence, availability of a clear clinical definition and validity: pressure ulcers,¹⁰ postoperative pulmonary embolism or deep vein thrombosis (PE/DVT),^{11–13} postoperative sepsis,^{14 15} ventilator-associated pneumonia (VAP)^{16–20} and postoperative wound infection.¹¹ The algorithms were based on the technical manual provided by the AHRQ (version 3.1)⁹ or the OECD study²¹ (appendix A²²).

Study population

All 116 acute hospitals in Belgium were invited to participate. Long-term care and rehabilitation

facilities as well as psychiatric hospitals, transfers from another acute care facility and one-day clinics were excluded. The study targeted adult non-obstetric patients. Twenty-one acute hospitals volunteered to participate. Of these, eight hospitals were selected according to region, hospital size and response time. Among the eight selected hospitals, there were two teaching hospitals. Of the remaining six hospitals, one had fewer than 300 beds, two had between 300 and 450 beds, and three had more than 450 beds.

Table 1 shows the characteristics of the eight selected hospitals and their population compared with all acute Belgian hospitals. It highlights the similarities between the eight selected hospitals compared with all acute hospitals in Belgium in terms of number of beds, patient characteristics, length of stay and number of secondary diagnoses coded in the administrative data.

Data source and sampling strategy

Selection of patients was based on the B-HDDS—similar to international administrative data—which has been compulsory since 1990 for all in-patients in acute hospitals in Belgium. The B-HDDS for the registration year 2005 (Y2005) was provided by the selected hospitals. Whenever necessary, additional registration years (Y2004 and Y2006) were used to obtain a sufficient number of cases.

The B-HDDS was used to screen for the five selected adverse events. A hospital discharge case was ‘flagged’ positive if at least

one of the selected indicators was scored positive based on the algorithm (appendix A). For each flagged case, a matched control negative case was selected. The matching was performed per hospital based on the All-Patient Refined Diagnosis-Related Groups (APR-DRG); the severity of illness (SOI, 1–4); age (<30, (30–49), (50–64), (65–79) and ≥80 years); gender, year and semester of registration. We randomly selected 20 flagged cases and 20 control cases per adverse event for each hospital to obtain a total of 200 patients per hospital. Another 50 reserve patients were randomly drawn from the pool of unselected cases to serve as alternate cases. However, the latter were not balanced for the five indicators due to the unequal prevalence of adverse events. Fifty reserve patients could not be obtained in one Flemish hospital due to its small size.

Data collection and recruitment

The B-HDDS obtained from the eight hospitals entailed 285 617 hospital stays (figure 1). There were 4490 (1.6%) cases flagged positive for one or more of the five selected adverse events, of which 2407 could be linked to a control case. After stratifying the cases according to the predefined adverse events, 1950 hospital stays were randomly selected (ie, 975 flagged and 975 control cases) for medical record review. A total of 378 files were excluded since: no informed consent given (141 cases), file unavailable (173 cases), medical record incomplete (47 cases), patient changed address (13 cases) and patients transferred from another acute hospital (four cases). Thus, a total of 1572 medical files were selected for medical record review, of which 1515 (94.7%) out of the 1600 targeted cases were reviewed. Of these, 741 files were positively flagged in the hospital discharge dataset for at least one of the selected indicators, and 774 cases were controls.

Data abstraction

Medical records were screened using a data abstraction tool (see appendix B) to standardise the data-collection process. It consisted of an anonymous patient number, admission and discharge dates, and an indication of the completeness of the medical file. The patient records were considered complete if two of three data sources were present: nursing notes, procedures notes and discharge notes. The patient information was acquired from administrative data and was verified by the review team: age and sex, admission type (elective or emergency), residence before admission and destination after discharge, length of stay and comorbidities. If an indicator was judged to be present on admission, no further analysis on the indicator was made.

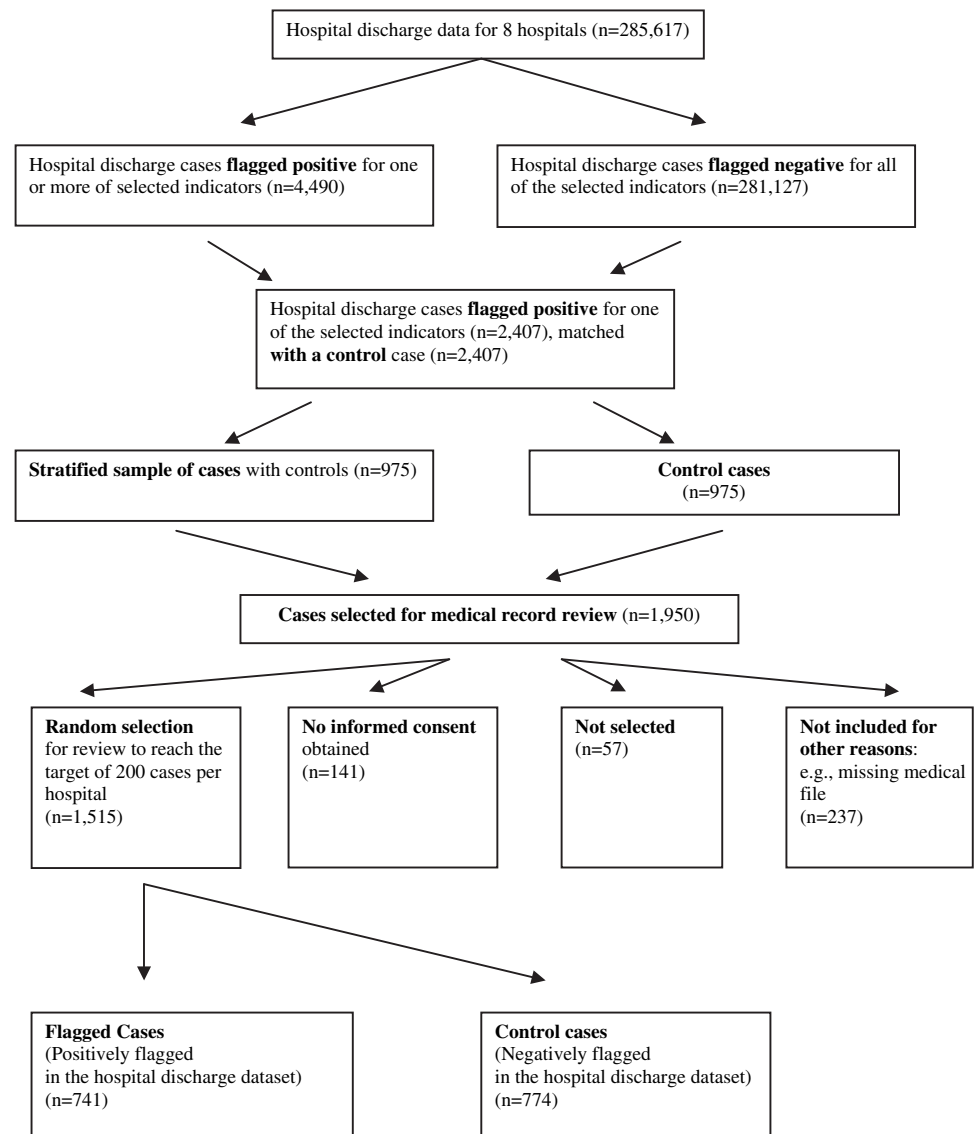
Medical record review

Two research teams (two team members per team) independently reviewed the medical records from four hospitals each. One team comprised one internal medicine specialist and one clinical pharmacist. The other team comprised one surgeon and one nurse. Reviewers were unaware whether a case was positively flagged for one of the five indicators. To familiarise themselves with the tool, each team conducted a pretest on 20 medical records. Whenever the two reviewers disagreed on the occurrence of an adverse event, they discussed the case until a consensus was reached. For 22 records (1.5%), they had no consensus for which the medical record was evaluated by an external panel of experts. Nineteen of these cases involved a question regarding PE/DVT, two regarding postoperative sepsis and one regarding postoperative wound infection.

Table 1 Characteristics of the selected hospitals (n=8) compared with all Belgian acute hospitals (n=109) for the registration year 2003 (Y2003)

	Selected hospitals	Belgian hospitals (Y2003)
University hospitals	2	7
Hospitals (n)		
<300 beds	1	37
300–450 beds	2	31
>450 beds	3	36
Inhospital stays (n)	193774	1271136
Length of stay		
Mean	8	8
P25	2	2
P50	4	4
P75	9	10
Age		
Mean	60	60
P25	48	46
P50	62	63
P75	75	75
Sex (female ratio)	48.51	51.14
Percentage of emergency admissions	44.54	44.25
Secondary diagnosis (mean)	7	7
Most frequent major diagnostic categories (%)*		
05 Circulatory System and related condition	18.41	16.36
08 Musculoskeletal System and Connective Tissue and related condition	15.99	17.64
06 Digestive System and related condition	11.67	12.00
01 Nervous System and related condition	9.35	8.87
04 Respiratory System and related condition	8.51	8.46

*Exclusion of major diagnostic categories 23, 14, 19, patients aged 18 years or older

Figure 1 Selection of cases.

Statistical analysis

Adverse events rates were calculated as a percentage of hospitalisations during which they were detected. For the B-HDDS and the medical record screening results, positive predictive value (PPV) and negative predictive values (NPV) were calculated for all five indicators together and for each indicator separately. Because of the case-control design, PPV and NPV were calculated directly. The B-HDDS was considered to be the test value, while the medical record screening was considered to be the true value. In four hospitals, an in-depth qualitative analysis was performed on all false-negative and positive cases in order to explain the observed mismatches between the B-HDDS and the medical record screening. This was achieved by re-evaluating the medical records. All analyses were performed using SAS v9.1 (SAS Institute, Cary, North Carolina).

RESULTS

Characteristics of the study sample

The characteristics of selected hospitals versus all Belgian acute hospitals were well matched (table 1). Selected study cases (positive vs negative flagged) were also well matched except for mean length of stay (33.82 vs 21.11 days) (table 2). The sample consists of 152 APR-DRG with a matching percentage of 95.7%. The mean (SD) age of the reviewed patients was 68.2 (15.9)

years. Given the specific choice of selected indicators, most cases (79%) were surgical.

Results of predictive values

Table 3 presents the results of predictive values for each indicator depending on whether we included or excluded events present on admission during the medical record screening. Postoperative

Table 2 Characteristics of positive flagged cases compared with control or negative flagged cases

	Positive cases (n = 741)	Controls (n = 774)
Age		
Mean	68.20	68.16
P25	59	58
P50	72	71
P75	80	80
Sex (female ratio)	49.53	47.93
Severity of illness (mean)	3	3
Length of stay (days)		
Mean	33.82	21.11
P25	12	8
P50	23	14
P75	42	24

Table 3 Predictive values for each indicator included or excluded from medical record screening when present on admission

	Pressure ulcer		Postoperative pulmonary embolism or deep vein thrombosis		Postoperative sepsis		Ventilator-associated pneumonia		Postoperative wound infection	
	POA+ (%)	POA- (%)	POA+ (%)	POA- (%)	POA+ (%)	POA- (%)	POA+ (%)	POA- (%)	POA+ (%)	POA- (%)
Positive predictive value	74.5	54.3	58.5	49.6	45.0	44.2	29.9	29.9	69.1	62.3
Negative predictive value	93.3	93.8	99.7	99.9	98.2	98.2	98.9	98.9	95.5	96.4

POA+, indicator from medical record screening included when present on admission; POA-, indicator from medical record screening excluded when present on admission.

wound infection achieved the highest PPV (62.3%) when the present on admission was excluded from the medical record screening. The PPV for postoperative PE/DVT was low (58.5%) and became even lower when the present on admission was excluded from the medical record screening (49.6%). Furthermore, there was a substantial loss in PPV after excluding pressure ulcers present on admission from the medical record screening (from 74.5% to 54.3%). Regardless of in- or excluding present on admission, PPVs were very similar (respectively 45% and 44.2%) for postoperative sepsis and the same (29.9%) for VAP. Except for pressure ulcer, all indicators had a high NPV (>96%).

In depth analysis

Of the 763 medical records from four hospitals, 269 (35.3%) showed a mismatch between the two datasets (table 4). Administrative data were false-negative or under-reported in 49 cases of pressure ulcer and false-positive in 20 cases.

Thirty-one cases of postoperative PE/DVT were considered false positive, mainly because no indication of PE/DVT was found (14 cases or 45.2%) or PE/DVT appeared to be present on admission (11 cases or 35.5%).

Thirty-seven false-positive cases for postoperative sepsis were identified. Twenty-one (56.8%) of these were due to the lack of specificity of the ICD-9-CM code (785.50; 998.0) to distinguish postoperative sepsis from postoperative haemorrhagic shock (15 cases) and postoperative cardiogenic shock (six cases).

In 58 cases of VAP, a false-positive result was withheld. In the majority of cases (34 or 58.6%), the clinical criteria for VAP appeared to be too strict.

For postoperative wound infection, in 15 (45.4%) of the 33 false-positive cases no evidence was found for the indicator, whereas in nine cases (27.3%) the indicator could not be scored due to the strictness of clinical criteria.

DISCUSSION

Using retrospective medical record review as the gold standard, we searched for evidence supporting the presence of any of the five selected adverse events reported in the B-HDDS.

Except for pressure ulcer, all indicators had high NPVs (>96%) but very low PPVs (29.9% to 62.3%).

These results were generally consistent with prior research. The preliminary validation study of screens by Iezzoni *et al* found a sensitivity of 93% and a specificity of 64%.²³ Lawthers and colleagues found that particularly medical screens were sensitive for the present-on-admission coding.²⁴ The large drop in PPV for pressure ulcers in the present study supports this finding.

To understand the mismatches between medical record reviews and administrative data, the current study also provided a detailed analysis on all false-negative and false-positive cases by re-evaluating the medical records involved. The lack of specificity of the ICD-9-CM codes accounted for nearly 57% of false-positive cases for postoperative sepsis, whereas the main problem for VAP was the strictness of clinical criteria used, resulting in 58.6% false-positive cases. The latter finding is consistent with recent work by Klompas and colleagues, in which the clinical diagnosis of VAP appeared to be notoriously inaccurate.²⁵

Our study had several limitations. First of all, for methodological reasons, we focused on five adverse events which represent only a small part of all possible complications that could occur during hospitalisation. For individual indicators, the number of cases examined was relatively small, although it was sufficient to make reasonable assessments about validity.

The retrospective medical record review of the selected cases may not represent a true gold standard. First, no inter-rater reliability test was performed by random abstraction. Second, a medical record was judged to be accurate whenever two of three data sources were present. However, missing nursing progress notes could have significantly affected identification of an indicator. Concerning these issues, Weingart *et al* already suggested that a physician review is at best a 'bronze standard' for evaluating quality.⁵ Finally, the eight selected hospitals accounted for just 6.9% of all acute care hospitals in Belgium. From all acute hospitals invited, only 17 (14.7%) agreed which potentially introduced a bias towards hospitals that are more active in the field of quality of care.

Table 4 Detailed results analysis of false-negative and false-positive cases from four hospitals

	Pressure ulcer	Postoperative pulmonary embolism or deep vein thrombosis	Postoperative sepsis	Ventilator-associated pneumonia	Postoperative wound infection
False negative	49/117 (41.9%)	0/28 (0.0%)	10/32 (31.3%)	8/29 (27.6%)	23/110 (20.9%)
False positive	20/646 (3.1%)	31/735 (4.2%)	37/731 (5.1%)	58/734 (7.9%)	33/653 (5.1%)
Over-reporting	20/20 (100%)	14/31 (45.2%)	9/37 (24.3%)	5/58 (8.6%)	15/33 (45.4%)
Present on admission	0/20 (0.0%)	11/31 (35.5%)	0/37 (0.0%)	8/58 (13.8%)	7/33 (21.2%)
Stringency criteria	0/20 (0.0%)	2/31 (6.4%)	6/37 (16.2%)	34/58 (58.6%)	9/33 (27.3%)
International Classification of Disease, 9th revision, Clinical Modification specificity	0/20 (0.0%)	4/31 (12.9%)	21/37 (56.8%)	11/58 (19.0%)	2/33 (6.1%)
Other*	0/20 (0.0%)	0/31 (0.0%)	1/37 (2.7%)	0/58 (0.0%)	0/33 (0.0%)

*Sepsis without a previous recent surgical procedure.

CONCLUSION

Our results provide new insight on the usefulness of identifying adverse events through administrative data. First, adverse events are indeed subject to under-reporting (and over-reporting to a lesser extent) in the B-HDDS. Whether this was in view of maximising reimbursements or, alternatively, coding errors could not be concluded. Except for a pressure ulcer which has a higher prevalence in hospitals, the NPVs we observed cannot be extrapolated to the general population. Furthermore, ICD-9-CM codes sometimes lack sufficient specificity for describing a certain adverse event which leads to false-positive results. Finally, some false-positive results are also related to the inability of the B-HDDS to correct for an event that was present on admission.

Although there are some problems with under- or over-reporting, administrative data are shown to be of good quality. Most under- or over-reporting problems are related to the lack of specificity of the ICD-9-CM codes or to adverse events being present on admission. Therefore, since administrative data provide a very inexpensive and readily accessible source of clinical information, we advocate that efforts should be made to refine administrative data by adding a code describing 'present on admission.' Indeed, this item was recently added to the new version of B-HDDS. Furthermore, providing more specific ICD-9-CM codes and examining the entire patient record, instead of the medical record only, will probably also increase the accuracy of administrative data. Finally, transparency in results on adverse events based on administrative data is essential. Hospitals should use them as feedback on their practices to allow for an improvement of their quality of care and of the codification.

In conclusion, the B-HDDS can probably accurately detect a select group of adverse events. However, further study is recommended since a code for present on admission was only recently added to our administrative data; this addition will likely correct for a substantial amount of false-positive results. Since no prevalence data on adverse events in Belgian hospitals based on medical record review are available, we recommend urgent study on this matter.

Acknowledgement We thank S von Winckelmann for her help in data collection. We thank S Devriese (KCE) and F Vrijens (KCE) for their support and feedback during the project. The full KCE report 93A can be retrieved from the KCE website (<http://www.kce.fgov.be>).

Funding The study was funded and contracted by the Belgian Health Care Knowledge Centre (KCE), Brussels, Belgium under grant agreement 2006-21.

Competing interests None.

Ethical approval The study was approved by the ethics committee and the medical board of each hospital. Informed consent was obtained from all patients.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Kohn LT, Corrigan JM, Donaldson M, et al. *To err is human: building a safer health system*. Washington, DC: National Academy Press, 1999.
2. Wilson RM, Runciman WB, Gibberd RW. The quality in Australian health care study. *Med J Aust* 1995;**163**:458–71.
3. Karson AS, Bates DW. Screening for adverse events. *J Eval Clin Pract* 1999;**5**:23–32.
4. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003;**12**:ii58–63.
5. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. *Med Care* 2000;**38**:796–806.
6. Woloshynowych M, Neale G, Vincent C. Case record review of adverse events: a new approach. *Qual Saf Health Care* 2003;**12**:411–15.
7. Best WR, Khuri SF, Phelan M, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg* 2002;**194**:257–66.
8. Fry DE, Pine MB, Jordan HS, et al. The Hazards of using administrative data to measure surgical quality. *Am Surg* 2006;**72**:1031–7.
9. AHRQ Quality Indicators. *Guide to patient safety indicators*. 2003. Rockville, MD: Agency for Healthcare Research and Quality, Version 3.1 March 12, 2007. AHRQ Pub.03–R203.
10. European Pressure Ulcer Advisory Panel. Guidelines on treatment of pressure ulcers. *EPUAP Review* 1999;**1**:31–3.
11. McCarthy EP, Iezzoni LI, Davis RB, et al. Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Med Care* 2000;**38**:868–76.
12. Stein PD, Woodard PK, Weg JG, et al. Diagnostic pathways in acute pulmonary embolism: recommendations of the PIOPE II Investigators. *Am J Med* 2006;**119**:1048–55.
13. Palla A, Bardi G, Ribas C. Diagnosis of pulmonary embolism. *Semin Thromb Hemost* 2006;**32**:822–30.
14. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992;**101**:1644–55.
15. Levy MM, Fink MP, Marchall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. *Crit Care Med* 2003;**31**:1250.
16. Calandra T, Cohen J. The international sepsis forum consensus conference on definitions of infection in the intensive care unit. *Crit Care Med* 2005;**33**:1538–48.
17. Morehead RS, Pinto SJ. Ventilator-associated pneumonia. *Arch Intern Med* 2000;**160**:1926–36.
18. Pugin J, Auckenthaler R, Mili N, et al. Diagnosis of ventilator-associated pneumonia by bacteriologic analysis of bronchoscopic and non-bronchoscopic 'blind' bronchoalveolar lavage fluid. *Am Rev Respir Dis* 1991;**143**:1121–9.
19. Farthoukh M, Maître B, Honoré S, et al. Diagnosing pneumonia during mechanical ventilation: the clinical pulmonary infection score revisited. *Am J Respir Crit Care Med* 2003;**168**:173–9.
20. Wunderink RG. Clinical criteria in the diagnosis of ventilator-associated pneumonia. *Chest* 2000;**117**:191S–4S.
21. Millar J, Matke S. *Selecting indicators for patient safety at the Health Systems Level in OECD Countries*. (Technical Paper). OECD 2004 (**18**).
22. Gillet P, Kolh P, Sermeus W, et al. Detection of adverse events in administrative databases—supplement. *KCE reports vol S* 2008.
23. Iezzoni LI, Foley SM, Heeren T, et al. A method for screening the quality of hospital care using administrative data: preliminary validation results. *QRB Qual Rev Bull* 1992;**18**:361–71.
24. Lawthers AG, McCarthy EP, Davis RB, et al. Identification of in-hospital complications from claims data: is it valid? *Med Care* 2000;**38**:785–95.
25. Klompas M. Does this patient have ventilator-associated pneumonia? *JAMA* 2007;**297**:1583–93.