BMJ Open | Accuracy of administrative data for surveillance of healthcare-associated infections: a systematic review

Maaike S M van Mourik,[1] Pleun Joppe van Duijn,[2] Karel G M Moons,[2] Marc J M Bonten,[1,2] Grace M Lee[3,4]

ABSTRACT

Objective: Measuring the incidence of healthcare-associated infections (HAI) is of increasing importance in current healthcare delivery systems. Administrative data algorithms, including (combinations of) diagnosis codes, are commonly used to determine the occurrence of HAI, either to support within-hospital surveillance programmes or as free-standing quality indicators. We conducted a systematic review evaluating the diagnostic accuracy of administrative data for the detection of HAI.

Methods: Systematic search of Medline, Embase, CINAHL and Cochrane for relevant studies (1995–2013). Methodological quality assessment was performed using QUADAS-2 criteria; diagnostic accuracy estimates were stratified by HAI type and key study characteristics.

Results: 57 studies were included, the majority aiming to detect surgical site or bloodstream infections. Study designs were very diverse regarding the specification of their administrative data algorithm (code selections, follow-up) and definitions of HAI presence. One-third of studies had important methodological limitations including differential or incomplete HAI ascertainment or lack of blinding of assessors. Observed sensitivity and positive predictive values of administrative data algorithms for HAI detection were very heterogeneous and generally modest at best, both for within-hospital algorithms and for formal quality indicators; accuracy was particularly poor for the identification of device-associated HAI such as central line associated bloodstream infections. The large heterogeneity in study designs across the included studies precluded formal calculation of summary diagnostic accuracy estimates in most instances.

Conclusions: Administrative data had limited and highly variable accuracy for the detection of HAI, and their judicious use for internal surveillance efforts and external quality assessment is recommended. If hospitals and policymakers choose to rely on administrative data for HAI surveillance, continued improvements to existing algorithms and their robust validation are imperative.

Strengths and limitations of this study

- Administrative data algorithms, based on discharge and procedure codes, are increasingly used to facilitate surveillance efforts and derive quality indicators.
- This comprehensive systematic review explicitly distinguished between administrative data algorithms developed for in-hospital surveillance and those for (external) quality assessment.
- All included primary studies were subjected to a thorough methodological quality assessment; this revealed frequent risk of bias in primary studies.
- The diverse nature of primary studies regarding study methods and algorithms precluded the pooling of results in most instances.

INTRODUCTION

Assessment of quality of care and monitoring of patient complications is a key concept in current healthcare delivery systems.[1] Administrative data, and discharge codes in particular, have been used as a valuable source of information to define patient populations, assess severity of disease, determine patient outcomes and detect adverse events, including healthcare-associated infections (HAI).[2–4] In certain instances, administrative data are employed to measure quality of care and govern payment incentives. Examples include patient-safety indicators (PSIs) developed by the USA Agency for Healthcare Quality Research, reduced payment for Healthcare-Associated Conditions (HACs) considered preventable and the expansion of value-based purchasing (VBP) initiatives, both implemented by US federal payers.[5–8] HAI rates reported to the national surveillance networks such as the US National Healthcare Safety Network (NHSN) are often determined from clinical patient information through chart review. Although these more

clinical rates are increasingly adopted by quality programmes, administrative data are still a key component of HAI detection for payers and some quality measurement programmes.[4 6]

Nonetheless, many cautionary notes have been raised regarding the accuracy of administrative data for the purpose of HAI surveillance.[1 9–11] Their universal use, ease of accessibility and relative standardisation across settings and time make them attractive for large-scale surveillance and research efforts. On the flip side—inherent to their purpose as a means to organise billing and reimbursement of healthcare—administrative data were not designed for the surveillance of HAI. Hence, when assigning primary and secondary discharge diagnosis codes, other interests may have greater priority, for example, maximising reimbursement for care delivered. In addition, the reliability of diagnosis code assignment depends heavily on adequate clinician documentation and the number of diagnoses in relation to the number of fields available.[3 12]

For the purpose of HAI surveillance, different targeted applications of administrative data algorithms define what measures of concordance are most important. First, they may be used as a case-finder to support within-hospital surveillance efforts, either in isolation or combined with other indicators of HAI such as microbiology culture results or antibiotic dispensing. In this case, sufficient sensitivity may be preferred over positive predictive value (PPV) to identify patients who require manual confirmation of HAI. Alternatively, discharge codes may be used in external quality indicator algorithms that directly determine the occurrence of HAI and thus gauge hospital performance.[3 9 13] In this setting, high PPV of observed signals may be of greater importance than detecting all cases of HAI. The primary objective of this systematic review was to assess the overall accuracy of published administrative data algorithms for the surveillance (ie, detection) of a broad range of HAI. We also determined whether the accuracy of algorithms developed for within-hospital surveillance differs from those meant for external quality evaluation. In addition, we rigorously evaluated the methodological quality of included studies using the QUADAS-2 tool developed for systematic reviews of diagnostic accuracy studies and also assessed the impact of a possible risk of bias.

METHODS

This systematic review includes studies assessing the diagnostic accuracy of administrative data algorithms using discharge and/or procedure codes for detecting HAI. Studies assessing infection or colonisation with specific pathogens (eg, methicillin-resistant *Staphylococcus aureus* or *Clostridium difficile*) were not included as laboratory-based surveillance may be considered more appropriate. The results of this analysis are reported in accordance with PRISMA guidelines.[14] This review did not receive protocol registration.

Search

Medline, EMBASE, the Cochrane database and CINAHL were searched for studies published from 1995 onwards with a query combining representations of administrative data and (healthcare-associated) infections (see online supplementary data 1 S1) with limits set to articles published in English, French or Dutch. The search was performed on 8 March 2012 and closed on 1 March 2013.

Study selection

To define suitability for inclusion, the following criteria were applied: (1) the study assessed concordance between administrative data and HAI occurrence, (2) data included were from 1995 or later as earlier data may be of limited generalisability to current practice, (3) the study did not reflect natural language processing and (4) the study presented original research rather than reviews or duplicated results. Selection of studies was done by a single reviewer (MSMvM), with cross-referencing to detect possibly missed studies. Inclusion was not restricted to specific geographical locations or patient populations, and nor was there a requirement for complete data availability.

Definitions

Administrative data algorithms were considered the index test (ie, the test under investigation). These algorithms consist of a selection of diagnosis and/or procedure codes used for billing or other purposes. The selection of codes within each algorithm was either specific for the study or, in some cases, they were predefined metrics used for payment or quality assessment. The latter group includes PSIs, HACs or the code selection defined by the Pennsylvania Healthcare Cost Containment Council (PHC4); most were used and developed in the USA, but the PSIs have also been used in other countries.[6 15] The reference standard was the presence or absence of HAI as determined by a review of patient clinical records, according to national infection surveillance methods (eg, NHSN), definitions from surgical quality monitoring programmes such as the US Surgical Quality Improvement Program (SQIP) or other definitions.

Quality assessment and data extraction

After selection of studies, quality assessment and data extraction was performed independently by two reviewers (MSMvM, PJvD) using modified QUADAS-2 criteria for quality assessment of diagnostic accuracy studies (see online supplementary table S2 for data extraction forms, details and assumptions).[16 17]

In brief, these criteria evaluate risk of bias and applicability to the review question with respect to methods of patient selection, the index test and the reference standard. In addition, the criteria provide a framework to evaluate risk of bias introduced by (in)complete HAI ascertainment, the so-called 'patient flow'. Points of special attention during the quality assessment were whether HAI ascertainment was blinded to the outcome

of the administrative data algorithm and the identification of partial or differential verification patterns. Partial verification occurs when not all patients were assessed for HAI presence (received the reference standard), in a pattern reliant on the result of the index test. In the case of differential verification, not all patients who were evaluated with the index test received the same reference standard. Depending on the pattern of partial and/or differential verification, this may have introduced bias in the observed accuracy estimates of the algorithm under study.[18] Several studies contained multiple types of verification patterns, methods of HAI ascertainment or specifications of administrative data algorithms; quality assessment and data extraction was then applied separately to each so-called comparison. Agreement between observers on methodological quality was reached by discussion.

Analyses

Included studies were stratified by HAI type and by the intended application of the administrative data within the process of HAI surveillance. A distinction was made between algorithms aimed at supporting within-hospital surveillance—either in isolation or in combination with other indicators—and those developed as a means of external quality of care evaluation. In addition, studies were classified by risk of bias based on QUADAS-2 criteria. Forest plots were created depicting the reported sensitivity, specificity, positive and negative predictive values of the administrative data algorithms for HAI detection.

If large enough groups of sufficiently comparable studies with complete two-by-two tables were available, estimates for sensitivity and specificity were pooled using the bivariate method recommended in the Cochrane Handbook for Systematic Reviews of Diagnostic Accuracy.[19 20] This analysis jointly models the distribution of sensitivity and specificity, accounting for correlation between these two outcome measures. There was no formal assessment of publication bias. All analyses were performed using R V.3.0.1 (http://www.r-project.org) and SPSS Statistics 20 (IBM, Armonk, New York, USA).

RESULTS
Study selection

After removal of duplicates, 8478 unique titles were screened for relevance and exclusion criteria were applied to 675 remaining abstracts. Cross-referencing identified four additional articles; in addition, 10 articles were published between the search date and search closure (figure 1). Fifty-seven studies, containing 71 comparisons, were available for the qualitative synthesis and underwent methodological quality assessment.[21–77]

Study characteristics

Study design, selection of the study population, methodology used as reference standard and administrative data specifications varied greatly. This large variability in study characteristics precluded the generation of summary estimates for sensitivity and specificity for most types of HAI. As the reference standard, 35 studies applied NHSN methodology to determine HAI presence, six defined HAI as registered in SQIP, and the remaining studies used clinical or other methods (table 1). Case-definitions were applied by infection preventionists in 24 studies, as well as by trained nurses, physicians or other abstractors. Eighteen studies assessed algorithms for within-hospital surveillance, and a further 15 combined administrative data with other indicators of infection (eg, microbiology culture results or antibiotic use) to detect HAI. Twenty-four studies assessed administrative data algorithms explicitly designed for external quality assessment, such as PSIs or HACs. Only seven studies provided data collected after 2008.[36 45 53 66 69 31 34]

Methodological quality

Figure 2 summarises the risk of bias and applicability concerns for each QUADAS-2 domain (see online supplementary data S3 for details by study; S4 for figures by HAI type). A high risk of bias in the flow component was observed in a considerable fraction of included studies. Ascertainment of HAI status was complete in 37 of 57 studies; in other words, only 65% of studies had the same reference standard applied to all or a random sample of the included patients. Alternative verification patterns were: evaluation of only those patients flagged by administrative data (nine), assessment of patients flagged by either administrative data or another test (eg, microbiological testing) (eight) and reclassification of discrepant cases after a second review. A high risk of bias for the flow component often co-occurred with the inability to extract complete data on diagnostic accuracy, mainly as a result of partial verification. In studies that assessed only the PPV, HAI ascertainment was limited to patients flagged by administrative data; this partial verification in itself was not problematic; however, lack of blinding of assessors may still have introduced an overall risk of bias.

Surgical site infection

Thirty four studies assessed surgical site infection (SSI); most studies identified the population at risk (ie, the denominator) by selecting specific procedure codes from claims data, although a few included all patients admitted to surgical wards. Details on administrative data algorithms are specified in online supplementary table S6. Algorithms in studies applying NHSN methods as a reference standard generally also incorporated diagnosis codes assigned during readmissions to complete the required follow-up duration, and several included follow-up procedures to detect SSI.

Accuracy estimates were highly variable (figure 3A, see online supplementary S5A), also within groups of studies with the same target procedures and intended

Figure 1 Flow chart of study selection and inclusion. HAI, healthcare-associated infections.
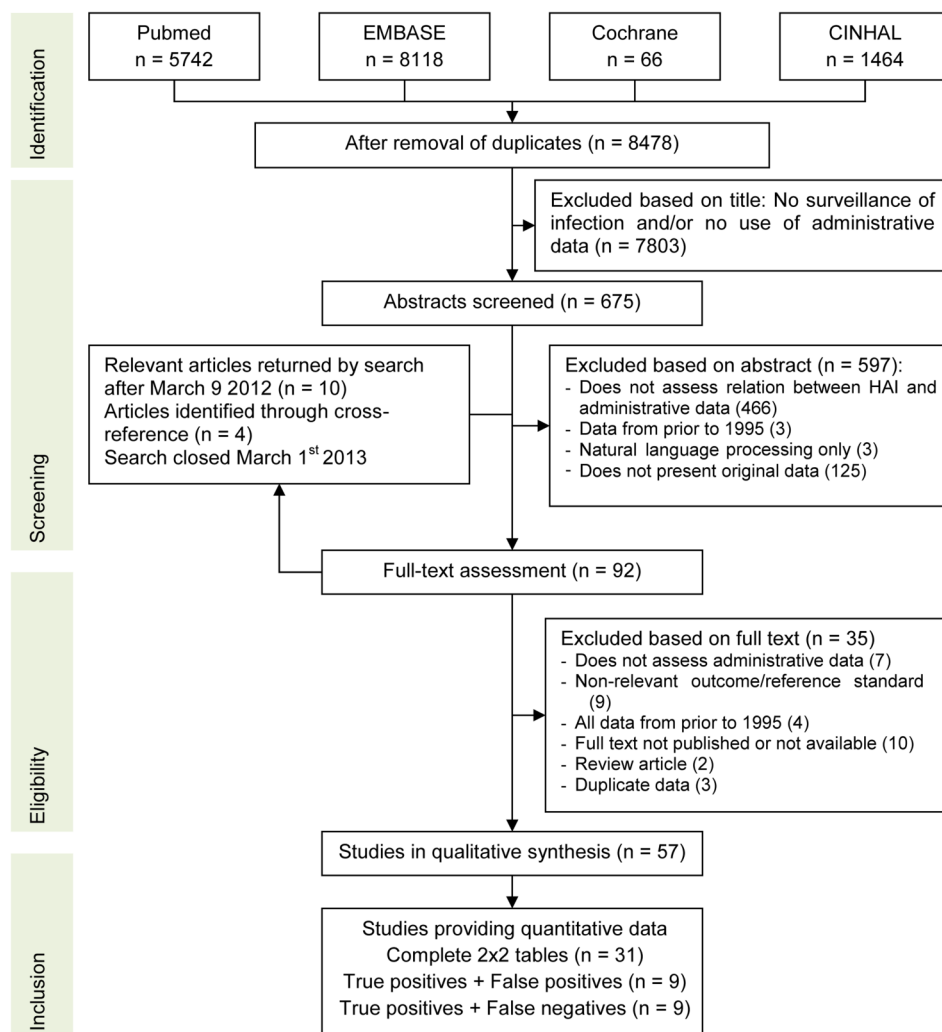
application (range for sensitivity 10–100%, PPV 11–95%). Several studies assessed multiple specifications of administrative data algorithms; as expected, using a broader selection of discharge codes detected more cases of SSI at the cost of lower PPV.[26 47 54] Between studies, there was no apparent relation between the specificity of the codes included and observed accuracy (ICD9 codes 998.5, 996.6 (or equivalent) vs a broader selection, data not shown). Inspection of the forest plots suggests that, in general, studies with a high risk of bias showed a more favourable diagnostic accuracy than those with more robust methodological quality, perhaps with the exception of cardiac procedures.

Bloodstream infections

Of the 24 studies evaluating bloodstream infections (BSI), half focused on central line-associated BSI (CLABSI) and 19 assessed algorithms for external quality assessment. Methods of identifying patients with a central line were very diverse: studies evaluating PSI 7 ('central venous catheter-related BSI') or HAC applied specific discharge codes, whereas other studies only included patients with positive blood cultures[67] or relied on manual surveillance to determine central line

presence (see online supplementary table S6).[69] The sensitivity of CLABSI detection was no higher than 40% in all but one study. Notably, only the studies that did not rely on administrative data to determine central line presence achieved sensitivity over 20% (figure 3B and see online supplementary S5B). The sensitivity of administrative data algorithms for detecting BSI was slightly higher. The pooled sensitivity of PSI 13 ('post-operative sepsis') in studies using SQIP methods as a reference standard was 17.0% (95% CI 6.8% to 36.4%) with a specificity of 99.6% (99.3% to 99.7%). Of the algorithms meant for external quality assessment, the PPVs varied widely and were often <50%, suggesting that these quality indicators detected many events that were not (CLA)BSI. Again, study designs with higher risks of bias tended to show higher accuracy.

Urinary tract infection

Fifteen studies investigated urinary tract infection (UTI), seven focusing specifically on catheter-associated UTI (CAUTI). In algorithms relying on administrative data to identify patients receiving a urinary catheter, the low sensitivity of CAUTI detection was striking (figure 3C,

Table 1   Main characteristics of included studies, stratified by targeted type of HAI

	Total	SSI	BSI	UTI	Pneumonia	Other
N studies	**57**	**34**	**24**	**15**	**14**	**2**
(N comparisons)	**(71)**	**(44)**	**(29)**	**(15)**	**(15)**	**(2)**
Device-associated	20	–	12	7	7	1
ICU only	5	1	3	2	3	0
Type of reference standard						
NHSN	35	26	9	6	7	2
(VA)SQIP	6	2	6	2	3	0
Clinical	4	1	3	1	1	0
Other	12	5	6	6	3	0
Application of administrative data						
External quality assessment	24	9	19*	6	8	0
Within hospital surveillance	18	13	3	7	4	1
Combined with other HAI indicators	15	12	3	2	2	1
Specific quality metric						
PSI	9	1	10	0	2	0
HAC	3	0	2	1	0	0
PHC4	4	4	3	3	4	0
Region of origin						
USA	44 (55)	22 (29)	19 (24)	10 (10)	9 (10)	1 (1)
Europe	8 (10)	8 (9)	4 (4)	4 (4)	4 (4)	1 (1)
Other	4 (6)	4 (6)	1 (1)	1 (1)	1 (1)	0 (0)
High risk of bias on QUADAS domain						
Patient selection	1 (1)	1 (1)	1 (1)	0 (0)	1 (1)	0 (0)
Index test	0 (3)	0 (1)	0 (1)	0 (0)	0 (0)	0 (0)
Reference standard	19 (27)	11 (18)	6 (7)	4 (4)	2 (2)	1 (1)
Flow	19 (29)	10 (18)	8 (11)	4 (4)	3 (4)	1 (1)
Verification pattern						
Complete or random sample	37 (42)	23 (26)	16 (18)	11 (11)	10 (10)	1 (1)
Complete with discrepant analysis	3 (6)	3 (6)	1 (2)	1 (1)	1 (2)	0 (0)
Partial, based on index test only	8 (8)	2 (4)	5 (7)	2 (2)	2 (2)	0 (0)
Partial, based on index and other tests	8 (12)	6 (6)	1 (1)	1 (1)	1 (1)	1 (1)
Other or unclear	1 (3)	0 (2)	1 (1)	0 (0)	0 (0)	0 (0)
Data availability						
Complete 2×2 table, by HAI type	29	20	10	6	6	1
Complete 2×2 table, HAI combined	3	3	2	4	3	0
Positive predictive value only, by HAI	9	3	6	1	2	0
Other	9	2	5	3	3	0
No data extraction possible	7	6	1	1	0	1

Some studies presented multiple comparisons and/or assessed more than 1 type of HAI; the number of comparisons is shown in brackets.
*One study targeting external quality assessment using administrative data combined with other sources of data.
BSI, bloodstream infections; HAC, Healthcare-associated condition as defined by the Centers for Medicare and Medicaid Services; HAI, healthcare-associated infections; ICU, intensive care unit; NHSN, National Healthcare Safety Network; PHC4, Pennsylvania Healthcare Cost Containment Counsel code selection; PSI, Patient Safety Indicator; QUADAS, Quality assessment for diagnostic accuracy studies; SSI, surgical site infection; UTI, urinary tract infection; (VA)SQIP, (Veteran's Administration) Surgical Quality Improvement Project.

see online supplementary S5C, S6).[78 76] Sensitivity was higher for UTI, but PPVs were universally below 25%, except in the study by *Heisler et al*; this study, however, additionally scrutinised flagged records for the presence of UTI.[34]

Pneumonia
Fourteen studies evaluated pneumonia, of which nine specifically targeted ventilator-associated pneumonia (VAP). The presence of mechanical ventilation was either determined within the administrative data algorithm[34 43] or by manual methods.[67] For VAP, sensitivity ranged from 35% to 72% and PPV from 12% to 57%.

For pneumonia, sensitivity and PPV hovered around 40%, although the studies used very diverse methodologies (figure 3D, see online supplementary S5D).

Other HAI and aggregated estimates
One study assessed the value of administrative data for detection of postpartum endometritis (data extraction not possible) and one the occurrence of drain-related meningitis. In addition, six studies presented data aggregated for multiple types of HAI (figure 3E, see online supplementary S5E). Also, for these studies, sensitivity did not exceed 60%, with similar or lower PPVs.
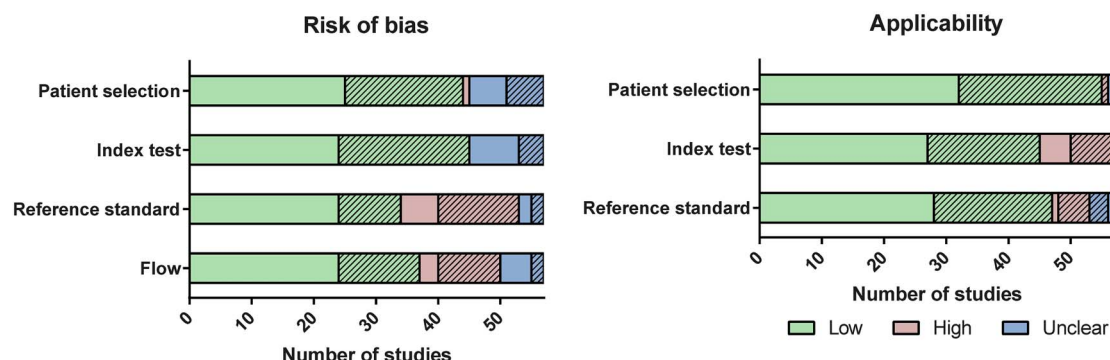
Figure 2 Summary of risk of bias and applicability for all studies (n=57), assessed using the *Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2)* methods. Some studies contain multiple comparisons; in this case, the lowest risk of bias per study is included. Shading denotes studies where extraction of complete two-by-two tables was not possible, including studies only assessing positive predictive values.

Algorithms combining administrative data with clinical data

Fifteen studies in this review evaluated the accuracy of administrative data in an algorithm that also included other (automated) indicators of HAI for within-hospital surveillance. Eight allowed for extraction of accuracy estimates of administrative data alone (labelled as 'Int (C)' in figure 3) and only very few provided the data necessary to fairly assess the incremental benefit of administrative data over clinical data such as antimicrobial dispensing or microbiology results. In these studies, gains in sensitivity obtained by adding administrative data were at most 10 percent points (data not shown).[23 49 50 59 74 75]

DISCUSSION

In the light of the increasing attention for evaluating, improving and rewarding quality of care, efficient and reliable measures to detect HAI are vital. However, as demonstrated by this comprehensive systematic review, administrative data have limited—and very variable— accuracy for the detection of HAI. In addition, algorithms to identify infections related to invasive devices such as central lines and urinary catheters are particularly problematic. All included studies were very heterogeneous in specifications of both the administrative data algorithms and the reference standard. Thorough methodological quality assessment revealed that incomplete ascertainment of HAI status and/or lack of blinding of assessors occurred in one-third of studies, thus introducing a risk of bias and complicating a balanced interpretation of accuracy estimates. Studies employing designs associated with a higher risk of bias appeared to provide a more optimistic picture than those employing more robust methodologies.

The drawbacks of administrative data for the purpose of HAI surveillance have been emphasised previously, especially from the perspective of (external) interfacility comparisons.[3 9 11 79] In comparison with a recent systematic review that assessed the accuracy of administrative data for HAI surveillance,[9] we identified a larger number of primary studies (partly due to broader inclusion criteria) and distinguished between administrative data algorithms developed for different intended applications. This prior review suggests that despite their moderate sensitivity, administrative data may be useful within broader algorithmic (automated) routine surveillance; notably, the studies in our systematic review demonstrated only modest gains in efficiency over other automated methods.[23 25 26 32 63 67 74] Surprisingly, there was no clear difference between administrative data algorithms developed for the purpose of supporting within-hospital surveillance versus those meant for external quality assessment in terms of sensitivity or PPV. Sensitivity was highly variable and PPVs were modest at best, also in algorithms targeting very specific events (CAUTI, CLABSI) for external benchmarking or payment rules. Administrative data may, however, be advantageous when aiming to track HAIs that require postdischarge surveillance across multiple healthcare facilities or levels of care, such as SSI.[41 80] Importantly, a considerable number of studies were performed in the USA, with a specific billing and quality evaluation system; hence, some quality metrics and coding systems may not be applicable to other countries.

A number of previously published studies explored reasons for the inability of administrative data to detect HAI. For specific quality measures, differences in HAI definitions between the quality metrics and NHSN methods may account for a portion of the discordant cases,[81]; other explanations include the erroneous detection of infections present-on-admission (PoA) or infections not related to the targeted device, incorrect coding, insufficient clinician documentation, challenges in identifying invasive devices or the limited number of coding fields available.[53 69 44 51 76 82 83] The precarious balance between the accuracy of administrative data and their use in quality measurement and pay-for-performance programmes has been argued previously, especially as these efforts may encourage coding practices that further undermine the accuracy of
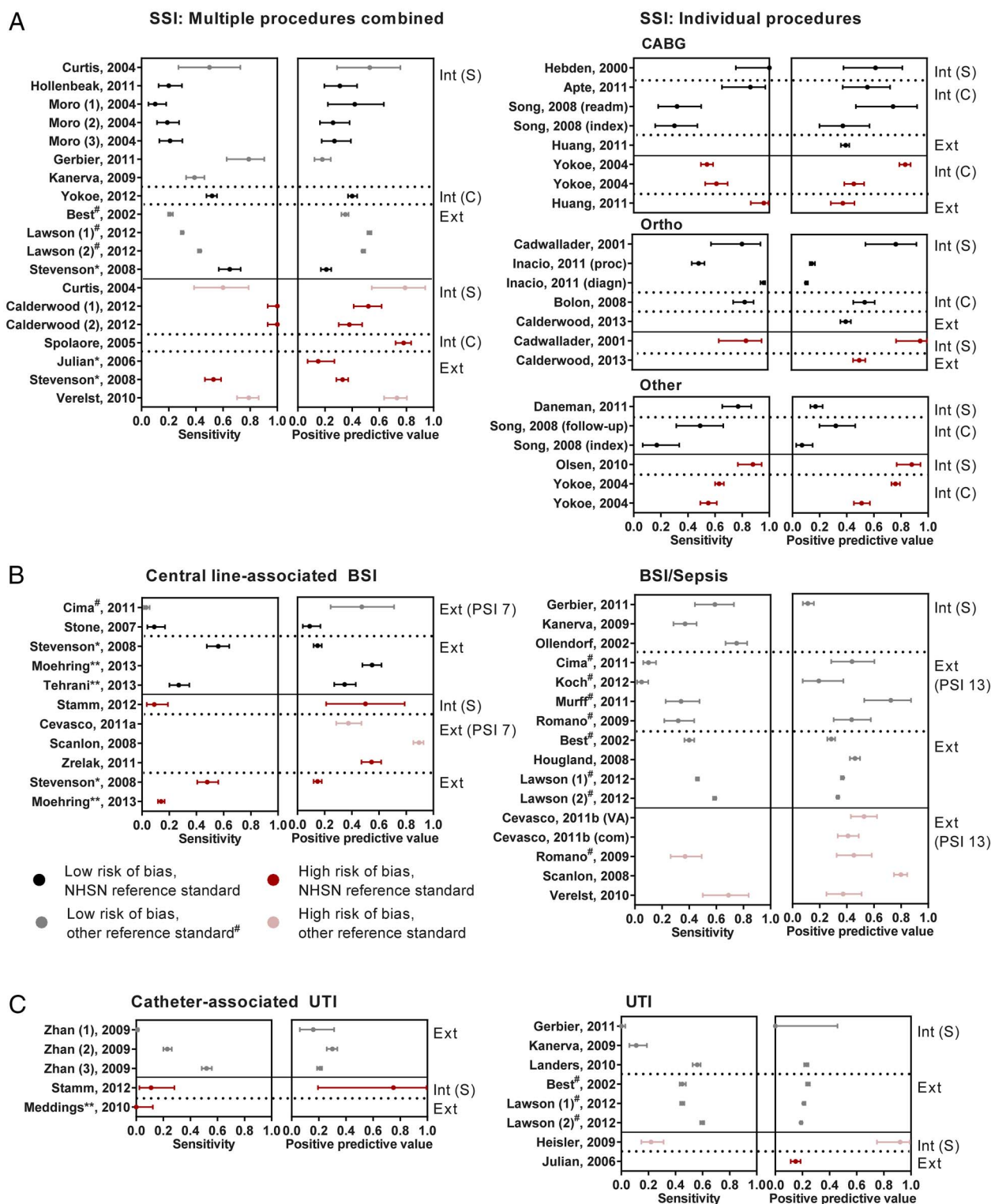
Figure 3 Forest plots for sensitivity and positive predictive value, stratified by HAI type and relevant study characteristics. Studies are grouped by the intended application of administrative data: Int (S)—used in isolation to support within-hospital surveillance efforts, Int (C)—used to support within-hospital surveillance, combined with other indicators of infection, Ext—used for external quality assessment, including public reporting and pay-for-performance. BSI, bloodstream infection; CABG, coronary artery bypass graft; DRM, drain-related meningitis; HAI, healthcare-associated infections; Ortho, orthopedic procedure; PSI, patient safety indicator; Sep, sepsis; SSI, surgical site infection; UTI, urinary tract infection. In studies including multiple specifications of the administrative data algorithm, these are numbered sequentially. 95% CIs are derived using the exact binomial method. If multiple study designs were performed within a single study, they are mentioned separately. #Reference standard from Surgical Quality Improvement Project (NSQIP or VASQIP). *Code selection based on specification from Pennsylvania Health Cost Containment Council. ** HAC specification.
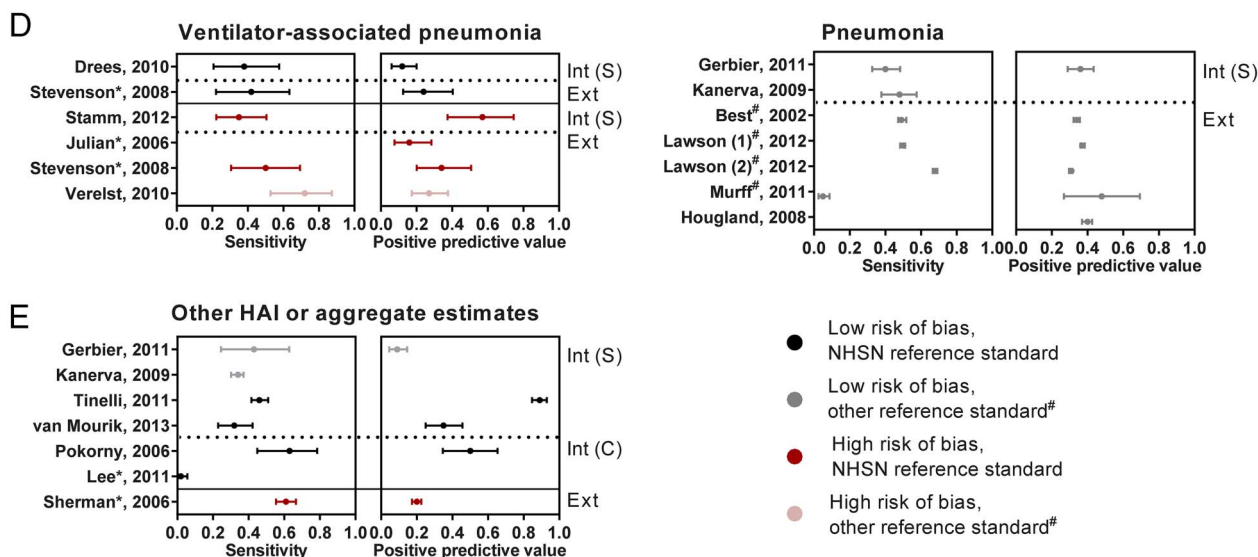
Figure 3 Continued.

administrative data.[11] Recent studies have provided mixed evidence regarding a change in coding practice in response to the introduction of financial disincentives or public reporting programmes.[84–86]

Several refinements in coding systems are currently in progress that may affect the future performance of administrative data. First, the transition to the 10th revision of the International Classification of Disease (ICD-10) may provide increased specificity due to the greater granularity of available codes.[87] Only seven studies in this review used the ICD-10, often in a setting that was not directly comparable to settings using the ICD-9 (mainly the USA), and some studies purposefully mapped the ICD-10 codes to mimic the ICD-9. Second, the number of coding fields available in (standardised) billing records has increased in recent years, allowing for more secondary diagnoses to be recorded; however, it is unclear whether expansion beyond 15 fields will benefit the HAI registration and other complications.[60 88] Third, the adoption and accuracy of PoA indicators in the process of code assignment remains to be validated, and they were incorporated in only a few studies included in this review.[78 89] Finally, this systematic review could not provide sufficient data to evaluate changes in coding accuracy since the US introduction of financial disincentives in 2008 for certain HACs that were not present on admission. Ongoing studies are needed to assess the impact of these changes in coding systems on their accuracy for HAI surveillance.

The frequent use of partial or differential verification patterns may be explained by the well-known limitations with quality of traditional surveillance as the reference standard in conjunction with the workload of applying manual surveillance to large numbers of patients.[23 25 26 32 63 67 74] Although reclassifying missed cases after a second review will result in more accurate detection of HAI, this differential application

of the second review may bias the performance estimates upwards,[18] unless it is applied to (a random sample of) all cases, including concordant HAI-negative and HAI-positive cases.[23 67 90]

Despite efforts to identify all available studies, we cannot exclude the possibility of having missed studies and nor did we assess publication bias. In addition, as the search was closed in March 2013, a number of primary studies within the domain of this systematic review have been published since closure of the search. The findings of these studies were in line with our observations.[80 82 83 90–99] In addition, as a result of our broad inclusion criteria, the included studies were very diverse, complicating the interpretation of the results. Contrary to a previous systematic review,[9] the small number of comparable studies motivated us to refrain from generating pooled summary estimates in most cases. Future evaluations of the accuracy of administrative data should consider applying the same reference standard to all patients, or—if unfeasible—to a random sample in each subgroup of the two-by-two table and ensure blinding of assessors. To facilitate a balanced interpretation of the results, estimates of diagnostic accuracy calculated before and after reclassification should also be reported separately.[100]

CONCLUSION

Administrative data such as diagnosis and procedure codes have limited, and highly variable, accuracy for the surveillance of HAI. Sensitivity of HAI detection was insufficient in most studies and administrative data algorithms that target specific HAI for external quality reporting also had generally poor PPVs, with identification of device-associated infections being the most challenging. The relative paucity of studies with a robust methodology and the diverse nature of the studies,

together with continuous refinements in coding systems, preclude reliable forecasting of the accuracy of administrative data in future applications. If administrative data continue to be used for the purposes of HAI surveillance, benchmarking or payment, improvement to existing algorithms and their robust validation is imperative.

Author affiliations
[1]Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands
[2]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
[3]Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, Massachusetts, USA
[4]Division of Infectious Diseases, Boston Children's Hospital, Boston, Massachusetts, USA

Twitter Follow Maaike van Mourik at @vanmourikmaaike

Contributors MSMvM designed the study, performed the search, critically appraised studies, performed the analysis and drafted the manuscript; PJvD critically appraised studies and helped write the manuscript; MJMB and KGMM assisted in the study design, critical appraisal, data analysis and writing of the manuscript; GML assisted in the study design, data interpretation and writing of the manuscript.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Panzer RJ, Gitomer RS, Greene WH, et al. Increasing demands for quality measurement. *JAMA* 2013;310:1971–80.
2. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
3. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003;12(Suppl 2):ii58–63.
4. Jarman B, Pieter D, van der Veen AA, et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Qual Saf Health Care* 2010;19:9–13.
5. Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. *N Engl J Med* 2007;357:1573–5.
6. Agency for Healthcare Research and Quality. Patient Safety Indicators Overview. http://wwwqualityindicatorsahrqgov/modules/psi_resourcesaspx 2014 (cited 18 Mar 2014); http://www.qualityindicators.ahrq.gov/modules/psi_resources.aspx
7. Department of Health and Human Services, Centers for Medicare and Medicaid Services. FY 2014 IPPS Final Rule Medicare Program: hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and fiscal year 2014 rates; quality reporting requirements for specific providers; hospital conditions of participation; payment policies related to patient status. *Federal Register* 2013;78:50495–1040.
8. Centers for Medicare and Medicaid Services. Press release: CMS issues proposed inpatient hospital payment system regulation, 2014 (cited 27 May 2014). http://wwwcmsgov/Newsroom/MediaReleaseDatabase/Press-releases/2014-Press-releases-items/2014-04-30.html
9. Goto M, Ohl ME, Schweizer ML, et al. Accuracy of administrative code data for the surveillance of healthcare-associated infections: a systematic review and meta-analysis. *Clin Infect Dis* 2014;58:688–96.
10. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. *Clin Infect Dis* 2009;49:949–55.
11. Farmer SA, Black B, Bonow RO. Tension between quality measurement, public quality reporting, and pay for performance. *JAMA* 2013;309:349–50.
12. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997;127(8 Pt 2):666–74.
13. Woeltje KF. Moving into the future: electronic surveillance for healthcare-associated infections. *J Hosp Infect* 2013;84:103–5.
14. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
15. Drosler SE, Romano PS, Tancredi DJ, et al. International comparability of patient safety indicators in 15 OECD member countries: a methodological approach of adjustment by secondary diagnoses. *Health Serv Res* 2012;47(1 Pt 1):275–92.
16. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
17. Benchimol EI, Manuel DG, To T, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–9.
18. Naaktgeboren CA, de Groot JA, van SM, et al. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med* 2013;159:195–202.
19. Macaskill P, Gatsonis C, Deeks JJ, et al. Chapter 10: Analysing and Presenting the Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy Version 1.0.* The Cochrane Collaboration, 2010:20–29.
20. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
21. Apte M, Landers T, Furuya Y, et al. Comparison of two computer algorithms to identify surgical site infections. *Surg Infect (Larchmt)* 2011;12:459–64.
22. 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.
23. Bolon MK, Hooper D, Stevenson KB, et al. Improved surveillance for surgical site infections after orthopedic implantation procedures: extending applications for automated data. *Clin Infect Dis* 2009;48:1223–9.
24. Braun BI, Kritchevsky SB, Kusek L, et al. Comparing bloodstream infection rates: the effect of indicator specifications in the evaluation of processes and indicators in infection control (EPIC) study. *Infect Control Hosp Epidemiol* 2006;27:14–22.
25. Cadwallader HL, Toohey M, Linton S, et al. A comparison of two methods for identifying surgical site infections following orthopaedic surgery. *J Hosp Infect* 2001;48:261–6.
26. Calderwood MS, Ma A, Khan YM, et al. Use of medicare diagnosis and procedure codes to improve detection of surgical site infections following hip arthroplasty, knee arthroplasty, and vascular surgery. *Infect Control Hosp Epidemiol* 2012;33:40–9.
27. Calderwood MS, Kleinman K, Bratzler DW, et al. Use of medicare claims to identify US hospitals with a high rate of surgical site infection after hip arthroplasty. *Infect Control Hosp Epidemiol* 2013;34:31–9.
28. Campbell PG, Malone J, Yadla S, et al. Comparison of ICD-9-based, retrospective, and prospective assessments of perioperative complications: assessment of accuracy in reporting. *J Neurosurg Spine* 2011;14:16–22.
29. Cevasco M, Borzecki AM, O'Brien WJ, et al. Validity of the AHRQ Patient Safety Indicator "central venous catheter-related bloodstream infections". *J Am Coll Surg* 2011;212:984–90.
30. Cevasco M, Borzecki AM, Chen Q, et al. Positive predictive value of the AHRQ Patient Safety Indicator "Postoperative Sepsis": implications for practice and policy. *J Am Coll Surg* 2011;212:954–61.
31. Cima RR, Lackore KA, Nehring SA, et al. How best to measure surgical quality? Comparison of the Agency for Healthcare Research and Quality Patient Safety Indicators (AHRQ-PSI) and the American College of Surgeons National Surgical Quality

Improvement Program (ACS-NSQIP) postoperative adverse events at a single institution. *Surgery* 2011;150:943–9.

32. Curtis M, Graves N, Birrell F, *et al*. A comparison of competing methods for the detection of surgical-site infections in patients undergoing total arthroplasty of the knee, partial and total arthroplasty of hip and femoral or similar vascular bypass. *J Hosp Infect* 2004;57:189–93.

33. Daneman N, Ma X, Eng-Chong M, *et al*. Validation of administrative population-based data sets for the detection of cesarean delivery surgical site infection. *Infect Control Hosp Epidemiol* 2011;32:1213–15.

34. Drees M, Hausman S, Rogers A, *et al*. Underestimating the impact of ventilator-associated pneumonia by use of surveillance data. *Infect Control Hosp Epidemiol* 2010;31:650–2.

35. Gerbier S, Bouzbid S, Pradat E, *et al*. [Use of the French medico-administrative database (PMSI) to detect nosocomial infections in the University hospital of Lyon.]. *Rev Epidemiol Sante Publique* 2011;59:3–14.

36. Haley VB, Van AC, Tserenpuntsag B, *et al*. Use of administrative data in efficient auditing of hospital-acquired surgical site infections, new york state 2009–2010. *Infect Control Hosp Epidemiol* 2012;33:565–71.

37. Hebden J. Use of ICD-9-CM coding as a case-finding method for sternal wound infections after CABG procedures. *Am J Infect Control* 2000;28:202–3.

38. Heisler CA, Melton LJ 3rd, Weaver AL, *et al*. Determining perioperative complications associated with vaginal hysterectomy: code classification versus chart review. *J Am Coll Surg* 2009;209:119–22.

39. Hollenbeak CS, Boltz MM, Nikkel LE, *et al*. Electronic measures of surgical site infection: implications for estimating risks and costs. *Infect Control Hosp Epidemiol* 2011;32:784–90.

40. Hougland P, Nebeker J, Pickard S, *et al*. Using ICD-9-CM codes in Hospital claims data to detect adverse events in patient safety surveillance. In: Hendrikson K, Battles JB, Keyes A, *et al*, eds. *Advances in patient safety: new directions and alternative approaches*. Rockville, MD: Agency for Healthcare Research and Quality, 2008:1–18.

41. Huang SS, Placzek H, Livingston J, *et al*. Use of Medicare claims to rank Hospitals by surgical site infection risk following coronary artery bypass graft surgery. *Infect Control Hosp Epidemiol* 2011;32:775–83.

42. Inacio MC, Paxton EW, Chen Y, *et al*. Leveraging electronic medical records for surveillance of surgical site infection in a total joint replacement population. *Infect Control Hosp Epidemiol* 2011;32:351–9.

43. Julian KG, Brumbach AM, Chicora MK, *et al*. First year of mandatory reporting of healthcare-associated infections, Pennsylvania: an infection control-chart abstractor collaboration. *Infect Control Hosp Epidemiol* 2006;27:926–30.

44. Kanerva M, Ollgren J, Virtanen MJ, *et al*. Estimating the annual burden of health care-associated infections in Finnish adult acute care hospitals. *Am J Infect Control* 2009;37:227–30.

45. Koch CG, Li L, Hixson E, *et al*. What are the real rates of postoperative complications: elucidating inconsistencies between administrative and clinical data sources. *J Am Coll Surg* 2012;214:798–805.

46. Landers T, Apte M, Hyman S, *et al*. A comparison of methods to detect urinary tract infections using electronic data. *Jt Comm J Qual Patient Saf* 2010;36:411–17.

47. Lawson EH, Louie R, Zingmond DS, *et al*. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg* 2012;256:973–81.

48. Lee J, Imanaka Y, Sekimoto M, *et al*. Validation of a novel method to identify healthcare-associated infections. *J Hosp Infect* 2011;77:316–20.

49. Leth RA, Moller JK. Surveillance of hospital-acquired infections based on electronic hospital registries. *J Hosp Infect* 2006;62:71–9.

50. Leth RA, Norgaard M, Uldbjerg N, *et al*. Surveillance of selected post-caesarean infections based on electronic registries: validation study including post-discharge infections. *J Hosp Infect* 2010;75:200–4.

51. Meddings J, Saint S, McMahon LF Jr. Hospital-acquired catheter-associated urinary tract infection: documentation and coding issues may reduce financial impact of Medicare's new payment policy. *Infect Control Hosp Epidemiol* 2010;31:627–33.

52. Miner AL, Sands KE, Yokoe DS, *et al*. Enhanced identification of postoperative infections among outpatients. *Emerg Infect Dis* 2004;10:1931–7.

53. Moehring RW, Staheli R, Miller BA, *et al*. Central line-associated infections as defined by the Centers for Medicare and Medicaid Services' Hospital-Acquired Condition versus Standard Infection Control Surveillance: why Hospital compare seems conflicted. *Infect Control Hosp Epidemiol* 2013;34:238–44.

54. Moro ML, Morsillo F. Can hospital discharge diagnoses be used for surveillance of surgical-site infections? *J Hosp Infect* 2004;56:239–41.

55. Murff HJ, FitzHenry F, Matheny ME, *et al*. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011;306:848–55.

56. Ollendorf DA, Fendrick AM, Massey K, *et al*. Is sepsis accurately coded on hospital bills? *Value Health* 2002;5:79–81.

57. Olsen MA, Fraser VJ. Use of diagnosis codes and/or wound culture results for surveillance of surgical site infection after mastectomy and breast reconstruction. *Infect Control Hosp Epidemiol* 2010;31:544–7.

58. Platt R, Kleinman K, Thompson K, *et al*. Using automated health plan data to assess infection risk from coronary artery bypass surgery. *Emerg Infect Dis* 2002;8:1433–41.

59. Pokorny L, Rovira A, Martin-Baranera M, *et al*. Automatic detection of patients with nosocomial infection by a computer-based surveillance system: a validation study in a general hospital. *Infect Control Hosp Epidemiol* 2006;27:500–3.

60. Romano PS, Mull HJ, Rivard PE, *et al*. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. *Health Serv Res* 2009;44:182–204.

61. Sands KE, Yokoe DS, Hooper DC, *et al*. Detection of postoperative surgical-site infections: comparison of health plan-based surveillance with hospital-based programs. *Infect Control Hosp Epidemiol* 2003;24:741–3.

62. Scanlon MC, Harris JM, Levy F, *et al*. Evaluation of the agency for healthcare research and quality pediatric quality indicators. *Pediatrics* 2008;121:e1723–31.

63. Sherman ER, Heydon KH, St John KH, *et al*. Administrative data fail to accurately identify cases of healthcare-associated infection. *Infect Control Hosp Epidemiol* 2006;27:332–7.

64. Song X, Cosgrove SE, Pass MA, *et al*. Using hospital claim data to monitor surgical site infections for inpatient procedures. *Am J Infect Control* 2008;36(3 Suppl):S32–6.

65. Spolaore P, Pellizzer G, Fedeli U, *et al*. Linkage of microbiology reports and hospital discharge diagnoses for surveillance of surgical site infections. *J Hosp Infect* 2005;60:317–20.

66. Stamm AM, Bettacchi CJ. A comparison of 3 metrics to identify health care-associated infections. *Am J Infect Control* 2012;40:688–91.

67. Stevenson KB, Khan Y, Dickman J, *et al*. Administrative coding data, compared with CDC/NHSN criteria, are poor indicators of health care-associated infections. *Am J Infect Control* 2008;36:155–64.

68. Stone PW, Horan TC, Shih HC, *et al*. Comparisons of health care-associated infections identification using two mechanisms for public reporting. *Am J Infect Control* 2007;35:145–9.

69. Tehrani DM, Russell D, Brown J, *et al*. Discord among performance measures for central line-associated bloodstream infection. *Infect Control Hosp Epidemiol* 2013;34:176–83.

70. Tinelli M, Mannino S, Lucchi S, *et al*. Healthcare-acquired infections in rehabilitation units of the Lombardy Region, Italy. *Infection* 2011;39:353–8.

71. van Mourik MS, Troelstra A, Moons KG, *et al*. Accuracy of hospital discharge coding data for the surveillance of drain-related meningitis. *Infect Control Hosp Epidemiol* 2013;34:433–6.

72. Verelst S, Jacques J, Van den HK, *et al*. Validation of Hospital Administrative Dataset for adverse event screening. *Qual Saf Health Care* 2010;19:e25.

73. Yokoe DS, Christiansen CL, Johnson R, *et al*. Epidemiology of and surveillance for postpartum infections. *Emerg Infect Dis* 2001;7:837–41.

74. Yokoe DS, Noskin GA, Cunnigham SM, *et al*. Enhanced identification of postoperative infections among inpatients. *Emerg Infect Dis* 2004;10:1924–30.

75. Yokoe DS, Khan Y, Olsen MA, *et al*. Enhanced surgical site infection surveillance following hysterectomy, vascular, and colorectal surgery. *Infect Control Hosp Epidemiol* 2012;33:768–73.

76. Zhan C, Elixhauser A, Richards CL Jr, *et al*. Identification of hospital-acquired catheter-associated urinary tract infections from Medicare claims: sensitivity and positive predictive value. *Med Care* 2009;47:364–9.

77. Zrelak PA, Sadeghi B, Utter GH, *et al*. Positive predictive value of the Agency for Healthcare Research and Quality Patient Safety Indicator for central line-related bloodstream infection ("selected infections due to medical care"). *J Healthc Qual* 2011;33:29–36.

78. Meddings JA, Reichert H, Rogers MA, *et al*. Effect of nonpayment for hospital-acquired, catheter-associated urinary tract infection: a statewide analysis. *Ann Intern Med* 2012;157:305–12.

79. Safdar N, Anderson DJ, Braun BI, *et al*. The evolving landscape of healthcare-associated infections: recent advances in prevention and a road map for research. *Infect Control Hosp Epidemiol* 2014;35:480–93.

80. Calderwood MS, Kleinman K, Bratzler DW, *et al*. Medicare claims can be used to identify US hospitals with higher rates of surgical site infection following vascular surgery. *Med Care* 2014;52:918–25.

81. Mull HJ, Borzecki AM, Loveland S, *et al*. Detecting adverse events in surgery: comparing events detected by the Veterans Health Administration Surgical Quality Improvement Program and the Patient Safety Indicators. *Am J Surg* 2014;207:584–95.

82. Cass AL, Kelly JW, Probst JC, *et al*. Identification of device-associated infections utilizing administrative data. *Am J Infect Control* 2013;41:1195–9.

83. Atchley KD, Pappas JM, Kennedy AT, *et al*. Use of administrative data for surgical site infection surveillance after congenital cardiac surgery results in inaccurate reporting of surgical site infection rates. *Ann Thorac Surg* 2014;97:651–7.

84. Thompson ND, Yeh LL, Magill SS, *et al*. Investigating Systematic Misclassification of Central Line-Associated Bloodstream Infection (CLABSI) to secondary bloodstream infection during health care-associated infection reporting. *Am J Med Qual* 2013;28:56–9.

85. Lindenauer PK, Lagu T, Shieh MS, *et al*. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003–2009. *JAMA* 2012;307:1405–13.

86. Calderwood MS, Kleinman K, Soumerai SB, *et al*. Impact of Medicare's payment policy on mediastinitis following coronary artery bypass graft surgery in US hospitals. *Infect Control Hosp Epidemiol* 2014;35:144–51.

87. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD -10) Version for 2010. 2010. 2–5–2014. Ref Type: Report.

88. Drosler SE, Romano PS, Sundararajan V, *et al*. How many diagnosis fields are needed to capture safety events in administrative data? Findings and recommendations from the WHO ICD-11 Topic Advisory Group on Quality and Safety. *Int J Qual Health Care* 2014;26:16–25.

89. Cram P, Bozic KJ, Lu X, *et al*. Use of present-on-admission indicators for complications after total knee arthroplasty: an analysis of Medicare administrative data. *J Arthroplasty* 2014;29:923–8.

90. Letourneau AR, Calderwood MS, Huang SS, *et al*. Harnessing claims to improve detection of surgical site infections following hysterectomy and colorectal surgery. *Infect Control Hosp Epidemiol* 2013;34:1321–3.

91. Patrick SW, Davis MM, Sedman AB, *et al*. Accuracy of hospital administrative data in reporting central line-associated bloodstream infections in newborns. *Pediatrics* 2013;131(Suppl 1):S75–80.

92. Knepper BC, Young H, Jenkins TC, *et al*. Time-saving impact of an algorithm to identify potential surgical site infections. *Infect Control Hosp Epidemiol* 2013;34:1094–8.

93. Quan H, Eastwood C, Cunningham CT, *et al*. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ Open* 2013;3:e003716.

94. Leclere B, Lasserre C, Bourigault C, *et al*. Computer-enhanced surveillance of surgical site infections: early assessment of a generalizable method for French hospitals. *European Conference of Clinical Microbiology and Infectious Diseases*; 2014.

95. Murphy MV, Du DT, Hua W, *et al*. The utility of claims data for infection surveillance following anterior cruciate ligament reconstruction. *Infect Control Hosp Epidemiol* 2014;35:652–9.

96. Grammatico-Guillon L, Baron S, Gaborit C, *et al*. Quality assessment of hospital discharge database for routine surveillance of hip and knee arthroplasty-related infections. *Infect Control Hosp Epidemiol* 2014;35:646–51.

97. Knepper BC, Young H, Reese SM, *et al*. Identifying colon and open reduction of fracture surgical site infections using a partially automated electronic algorithm. *Am J Infect Control* 2014;42(10 Suppl):S291–5.

98. Warren DK, Nickel KB, Wallace AE, *et al*. Can additional information be obtained from claims data to support surgical site infection diagnosis codes? *Infect Control Hosp Epidemiol* 2014;35 (Suppl 3):S124–32.

99. Leclere B, Lasserre C, Bourigault C, *et al*. Matching bacteriological and medico-administrative databases is efficient for a computer-enhanced surveillance of surgical site infections: retrospective analysis of 4,400 surgical procedures in a French university hospital. *Infect Control Hosp Epidemiol* 2014;35:1330–5.

100. de Groot JA, Bossuyt PM, Reitsma JB, *et al*. Verification problems in diagnostic accuracy studies: consequences and solutions. *BMJ* 2011;343:d4770.