Epidemiol. Infect. (2016), **144**, 1999–2005. © Cambridge University Press 2016
doi:10.1017/S0950268815003271

Administrative data is as good as medical chart review for comorbidity ascertainment in patients with infections in Singapore

J. HWANG[1]*, A. CHOW[2,3], D. C. LYE[1,4] AND C. S. WONG[2,3]

[1] *Department of Infectious Diseases, Institute of Infectious Diseases and Epidemiology, Tan Tock Seng Hospital, Singapore*
[2] *Department of Clinical Epidemiology, Institute of Infectious Diseases and Epidemiology, Tan Tock Seng Hospital, Singapore*
[3] *Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore*
[4] *Yong Loo Lin School of Medicine, National University of Singapore, Singapore*

Received 17 August 2015; Final revision 6 November 2015; Accepted 16 December 2015; first published online 13 January 2016

SUMMARY

The Charlson comorbidity index (CCI) is widely used for control of confounding from comorbidities in epidemiological studies. International Classification of Diseases (ICD)-coded diagnoses from administrative hospital databases is potentially an efficient way of deriving CCI. However, no studies have evaluated its validity in infectious disease research. We aim to compare CCI derived from administrative data and medical record review in predicting mortality in patients with infections. We conducted a cross-sectional study on 199 inpatients. Correlation analyses were used to compare comorbidity scores from ICD-coded administrative databases and medical record review. Multivariable regression models were constructed and compared for discriminatory power for 30-day in-hospital mortality. Overall agreement was fair [weighted kappa 0·33, 95% confidence interval (CI) 0·23–0·43]. Kappa coefficient ranged from 0·17 (95% CI 0·01–0·36) for myocardial infarction to 0·85 (95% CI 0·59–1·00) for connective tissue disease. Administrative data-derived CCI was predictive of CCI $\geqslant 5$ from medical record review, controlling for age, gender, resident status, ward class, clinical speciality, illness severity, and infection source ($C = 0{\cdot}773$). Using the multivariable model comprising age, gender, resident status, ward class, clinical speciality, illness severity, and infection source to predict 30-day in-hospital mortality, administrative data-derived CCI ($C = 0{\cdot}729$) provided a similar C statistic as medical record review ($C = 0{\cdot}717$, $P = 0{\cdot}8548$). In conclusion, administrative data-derived CCI can be used for assessing comorbidities and confounding control in infectious disease research.

Key words: Infectious disease, medical informatics (veterinary and medical), public health.

INTRODUCTION

Early and accurate risk stratification is often key to the clinical management and choice of treatment for patients with infections. There are numerous scoring systems available, such as Acute Physiology and Chronic Health Evaluation (APACHE) [1], Sepsis-related Organ Failure Assessment (SOFA) [2] and Multiple Organ Failure Score [3]. These systems rely on laboratory data, which may not be easily available for initial patient assessment.

A scoring system based on a patient's comorbid disease can potentially offer an easier way to quantify the mortality risk for these patients. Charlson *et al*. developed a comorbidity index which quantifies the overall burden of comorbid disease based on the relative risk

* Author for correspondence: Dr J. Hwang, 1E Kent Ridge Road, Singapore 119228.
(Email: jeff_yf_hwang@nuhs.edu.sg)

https://doi.org/10.1017/S0950268815003271 Published online by Cambridge University Press

of mortality for each comorbidity, and subsequently validated the score on a cohort of patients with primary breast carcinoma [4]. The Charlson comorbidity index (CCI) has been shown to predict mortality accurately in different settings, such as in patients with infections, in which Oltean *et al.* [5] demonstrated that a higher index score was associated with increased mortality. Oltean *et al.* also showed that the index correlated positively with another score, the SOFA. Besides CCI, Elixhauser *et al.* [6] also derived another comorbidity index, and both Charlson and Elixhauser comorbidity indices are comparable in predicting mortality in critically ill patients [7].

However, deriving such a score is not without its challenges. Manual record review may be a laborious process, and time-intensive as well, especially in situations where clinical decisions need to be made promptly. Information technology (IT) may thus play a useful role in obtaining information from administrative databases. Administrative data is an increasingly utilized source of information, as it covers a wide population and contains in-depth clinical data [8].

In Singapore, electronic administrative databases such as the Cluster Patients Records System and National Electronic Health Record maintain hospital discharge records of every patient admitted to a public hospital. The discharge records comprise a detailed recording of the patient's inpatient stay, including details of the patient's pre-existing medical problems, reason for admission, as well as the International Classification of Diseases (ICD) codes for discharge diagnosis and comorbid conditions. The administrative databases also obtain information about a patient's co-existing medical problems from other sources such as investigation orders and emergency room clinical records.

Using ICD codes from administrative data to search for Charlson comorbidities is potentially an efficient way of deriving the CCI. Deyo *et al.* [9] developed a coding algorithm for the Charlson comorbidities based on International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) in 1992, and showed that an increasing comorbidity index is associated with more adverse outcomes such as in-hospital complications and post-operative deaths in patients who underwent lumbar surgery. In 2005, Quan *et al.* [10] derived an algorithm based on the newer ICD-10, which was shown to have a better model performance compared with Deyo *et al.*'s ICD-9-CM coding algorithm. However, these two studies were not compared with a criterion standard such as chart review data.

A recent study [11] assessed the positive predictive value of ICD-10 diagnostic codes by comparing them with discharge summaries or medical records, and found an overall positive predictive value of 98·0%. However, another study [12] found that the CCI derived from record data was superior to the same index derived from administrative billing data as a predictor of inpatient mortality. In Singapore, a study by Chong *et al.* [13] noted that the prevalence of almost all Elixhauser comorbidities obtained from administrative data was lower than that obtained from manual medical record review.

Several studies have evaluated the performance of administratively derived CCI in patients with stroke [14], cardiovascular conditions [12] and renal disease [15]. However, none thus far has evaluated the performance of CCI derived from administrative databases in patients with infections. As there are many studies that use the CCI for studies on patients with infections, it is important to evaluate the CCI from administrative databases.

Our study aims to compare the comorbidity scores derived from ICD-coded administrative databases with those from medical record review of patients with infectious conditions in Singapore, and to evaluate the discriminatory power of the administrative data-derived CCI and medical record-derived CCI for 30-day in-hospital mortality.

METHODS

Study population

The study population comprised patients admitted to a 1500-bed tertiary-care hospital in Singapore, from 1 October 2011 to 30 September 2012, who were prescribed piperacillin-tazobactam or a carbapenem for empirical therapy of an infectious condition and automatically triggered to receive antibiotic recommendations by the institutional antibiotic computerized decision support system. A random sample of 199 (10·6%) patients from the cohort was included in this study. These patients represented the hospital patient population admitted for an acute infectious condition. With a sample of 199 patients, our study was adequately powered at 80% with a type I error rate of no more than 5%, to detect most meaningful odds ratios ($\geqslant$1·5) for the primary outcome of interest.

https://doi.org/10.1017/S0950268815003271 Published online by Cambridge University Press

Record review

We used each patient's unique personal identification number, and date of admission to obtain their discharge medical record. A manual review of the discharge record was performed by a single physician to derive the list of comorbid conditions on admission. The comorbid conditions were tabulated and given a score in accordance with the Charlson Weighted Index of Comorbidity score [4].

Separately, comorbidity scores for each participant were derived from the ICD-coded administrative databases, based on comorbidities present at start of hospitalization, using the ICD-9-CM coding algorithm used by Quan *et al.* [10].

Data analysis

Correlation analyses were subsequently performed to compare the scores derived from discharge record review, and from administrative databases. Agreement was quantified based on Cohen's kappa coefficient.

We also evaluated the CCI, based on both medical record review as well as administrative data, as a predictor for 30-day in-hospital mortality. Multivariate logistic regression models were constructed and compared for discriminatory power for 30-day in-hospital mortality. Statistical analyses were performed using SPSS for Windows v. 16.0 (SPSS Inc., USA)

Ethical approval was obtained from the Domain Specific Research Board, National Healthcare Group (Singapore).

RESULTS

Table 1 summarizes the characteristics of the study population. The mean age was 72 years (S.D. = 15·1 years). The most common source of infection was pneumonia, followed by urinary tract infection.

Table 2 shows the agreement between both data sources for each comorbidity. The overall agreement between both data sources was fair [weighted kappa 0·33, 95% confidence interval (CI) 0·23–0·43]. Spearman correlation was 0·38 ($P < 0·0001$). Kappa coefficient ranged from 0·17 (95% CI –0·01 to 0·36) for myocardial infarction, to 0·85 (95% CI 0·57–1·00) for connective tissue disease. Three comorbid conditions (peripheral vascular disease, leukaemia, lymphoma) had good agreement of both sources. Four conditions (chronic pulmonary disease, diabetes, diabetes with complications, metastatic solid tumour) were in moderate agreement.

CCI derived from administrative data was predictive of CCI $\geqslant 5$ from medical record review, after controlling for age, gender, resident status, ward class, clinical speciality, illness severity, and infection source ($C = 0·773$) (Fig. 1).

Using the multivariable logistic regression model comprising age, gender, resident status, ward class, clinical speciality, illness severity, and infection source to predict 30-day in-hospital mortality, CCI derived from administrative data provided a similar C statistic as CCI computed from medical record review ($C = 0·729$ and $C = 0·717$, $P = 0·8548$) (Fig. 2).

DISCUSSION

Our study shows that administrative data-derived CCI performed as well as CCI computed from medical record review in predicting 30-day in-hospital mortality in patients being treated for infections in Singapore. Luthi *et al.* [16] previously evaluated the performance of administrative data-derived CCI in Switzerland, and found that it provided a higher C statistic compared to single-day chart review for hospital mortality ($C = 0·863$ and $C = 0·795$, respectively). That study included all adult patients hospitalized in acute wards and did not specifically assess patients with infections.

Although the agreement between administrative data and medical record review differed for different conditions, eight conditions had moderate to good agreement, including conditions with higher scores in the CCI computation, such as AIDS and metastatic tumour. A previous work reported similar findings in a study on adult in-patients in Switzerland [16], with six out of 17 comorbidities having moderate agreement between CCI derived from administrative data and CCI from single-day chart review, among which chronic pulmonary disease, diabetes, diabetes with complications, and metastatic solid tumour also had moderate to good agreement in our study. Kieszak *et al.* [12] found that in patients undergoing carotid endarterectomy in the United States, chart index and ICD-9-derived CCI had moderate to good agreement for three conditions (chronic pulmonary disease, diabetes, leukaemia or lymphoma). We found similarly moderate to good agreement for these three conditions in our study. In the study by Chong *et al.* on patients with pneumonia in Singapore, which investigated Elixhauser comorbidities instead of Charlson comorbidities, diabetes (both complicated and uncomplicated), metastatic cancer, chronic pulmonary

https://doi.org/10.1017/S0950268815003271 Published online by Cambridge University Press

Table 1. *Characteristics of the study population*

Characteristic	Patients ($N = 199$) *n* (%)
Demographic information	
Age, years, mean (S.D.)	72·3 (15·1)
Males	113 (56·8)
Ethnicity	
Chinese	148 (74·4)
Malay	20 (10·1)
Indian	16 (8·0)
Others	15 (7·5)
Resident status	
Resident	191 (96·0)
Foreign	8 (4·0)
Ward type	
General ward	183 (92·0)
Intensive care unit/high-dependency ward	16 (8·0)
Ward class	
Subsidized	180 (90·5)
Private	19 (9·6)
Source of infection	
Pneumonia	134 (67·3)
Urinary tract infection	18 (9·1)
Sepsis	15 (7·5)
Intra-abdominal	10 (5·0)
Skin and soft tissue	6 (3·0)
Hepatobiliary	6 (3·0)
Neurological	5 (2·5)
Bone and joint	4 (2·0)
Cardiac	1 (0·5)
Clinical speciality	
Medical	162 (81·4)
Surgery (includes neurosurgery, general surgery, orthopaedics, urology, otorhinolaryngology)	37 (18·6)
30-day mortality	32 (16·1)

disease, lymphoma and peripheral vascular disease had moderate or good agreement between medical chart review and administrative data [13], as seen in our study. Certain conditions such as connective tissue disease and renal disease had varying degrees of agreement between chart data and administrative data, and possible reasons include the wide spectrum of these conditions in terms of severity and/or diagnoses, which could affect the quality of the coding of these conditions. Any difference in coding was more likely to be related to the disease condition itself than the speciality to which the patients were admitted, as the team of clinical coders in our hospital would code for the disease conditions in the same way for all clinical specialities in the hospital. All clinical coders had undergone prior training to code for disease conditions in a standardized and consistent manner.

CCI is currently not widely utilized in the clinical setting, with possible reasons being that it requires additional time and effort to compute the score from manual record review, and clinicians may not have access to all the patient's records to compute the index. Using IT to derive the CCI from administrative databases is a quick and efficient method of deriving the CCI and delivering the score to clinicians, which may thus increase the adoption of use of CCI in routine clinical care. The readily available administrative data-derived CCI can be made available to physicians for the prognostication of 30-day in-hospital mortality on admission.

Besides clinical work, infectious disease research and other epidemiological studies often encounter the issue of whether subjects with comorbidities are eligible for their study. Many studies have used restrictive eligibility criteria based on specific diagnoses to exclude patients with certain comorbid diseases, reducing the statistical power of the study. CCI provides a validated score to adjust for potential confounding due to comorbidities in clinical and epidemiological studies. Our study demonstrated that administrative data-derived CCI performed as well as medical record review in predicting in-patient mortality. Administrative data-derived CCI can offer an efficient means to quantify comorbidity risk and to adjust for potential confounding in the study population.

Our study could be limited by the potential measurement error from chart reviews performed by a single physician. Nonetheless, prior to the study, the physician had received training on how to obtain the CCI in a systematic manner, and had derived the CCI systematically from medical records. As such, any potential misclassification was likely to be non-differential. Having a single physician also prevents inter-rater reliability issues due to interpersonal differences, which may be large when reviewing medical records.

In addition, as our patient population comprised patients who were prescribed piperacillin-tazobactam or a carbapenem for an infectious condition, the results of the study may possibly be more representative of patients with more severe infectious conditions, although the severity of a patient's condition was less likely to affect the use of administrative data or chart review to derive the CCI.

Our study may be underpowered to evaluate certain conditions, as the study population had a low number

https://doi.org/10.1017/S0950268815003271 Published online by Cambridge University Press

Table 2. *Agreement between medical record review and administrative data*

Comorbidity	Patients* with comorbidity from medical record, *n* (%)	Patients* with comorbidity from administrative data, *n* (%)	Kappa coefficient (95% CI)
Connective tissue disease	4 (2·0)	3 (1·5)	0·85 (0·59 to 1·00)
Peripheral vascular disease	12 (6·0)	6 (3·0)	0·65 (0·40 to 0·91)
Leukaemia	1 (0·5)	2 (1·0)	0·66 (0·05 to 1·00)
Lymphoma	2 (1·0)	4 (2·0)	0·66 (0·22 to 1·00)
Chronic pulmonary disease	40 (20·1)	24 (12·1)	0·60 (0·45 to 0·74)
Diabetes with complications	30 (15·1)	47 (23·6)	0·57 (0·43 to 0·71)
Metastatic solid tumour	9 (4·5)	12 (6·0)	0·55 (0·29 to 0·81)
Diabetes	55 (27·6)	24 (12·1)	0·53 (0·39 to 0·66)
AIDS	4 (2·0)	1 (0·5)	0·40 (–0·15 to 0·94)
Moderate or severe liver disease	3 (1·5)	2 (1·0)	0·39 (–0·15 to 0·94)
Dementia	20 (10·1)	16 (8·0)	0·33 (0·12 to 0·54)
Mild liver disease	6 (3·0)	11 (5·5)	0·33 (0·03 to 0·62)
Moderate or severe renal disease	17 (8·5)	46 (23·1)	0·33 (0·17 to 0·48)
Cerebrovascular disease	51 (25·6)	36 (18·1)	0·29 (0·13 to 0·44)
Congestive heart failure	17 (8·5)	23 (11·6)	0·28 (0·08 to 0·48)
Hemiplegia	3 (1·5)	14 (7·0)	0·22 (–0·05 to 0·48)
Any tumour	24 (12·1)	21 (10·6)	0·22 (0·03 to 0·41)
Myocardial infarction	25 (12·6)	20 (10·1)	0·17 (–0·01 to 0·36)
Ulcer disease	13 (6·5)	0 (0)	n.a.

* Patients ($N = 199$).
CI, Confidence interval; n.a., not applicable.

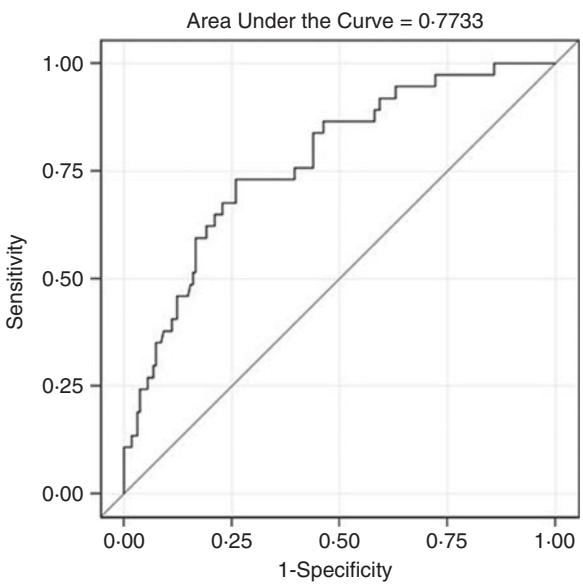

Fig. 1. Receiver operating characteristic curve showing prediction of Charlson comorbidity index (CCI) $\geqslant 5$ from medical record review by CCI derived from administrative data, after adjusting for age, gender, resident status, ward class, clinical speciality, illness severity, and infection source.

of several comorbidities such as connective tissue disease, leukaemia, lymphoma and AIDS, which makes interpretation of the data more difficult. Moreover, our study showed that less than half (8/19) of the comorbidity conditions had moderate to good agreement between administrative data and medical record review. Nonetheless, our study did show that administrative data was comparable with medical record review in predicting patients with CCI $\geqslant 5$, which represents patients with more severe pre-existing conditions or a higher number of comorbidities. While it can be argued that the results of the study may possibly be limited to this group of patients, these patients may be the very ones for which knowledge of the comorbidity score has more significant clinical implications, as a higher comorbidity score could lead to more cautious clinical management.

Nonetheless, our study had several strengths. The administrative-based CCI was derived from the ICD-coded administrative databases, based on comorbidities documented in hospital discharge summaries. All summaries were reviewed by senior physicians in charge of the patients and diagnoses assigned ICD codes by trained clinical coders, who were either non-practising physicians or professionally trained coders with a clinical background [13]. The coding practice adheres to the Singapore Coding Directive, a national coding standard, and the coding accuracy is monitored through regular audits by the

https://doi.org/10.1017/S0950268815003271 Published online by Cambridge University Press

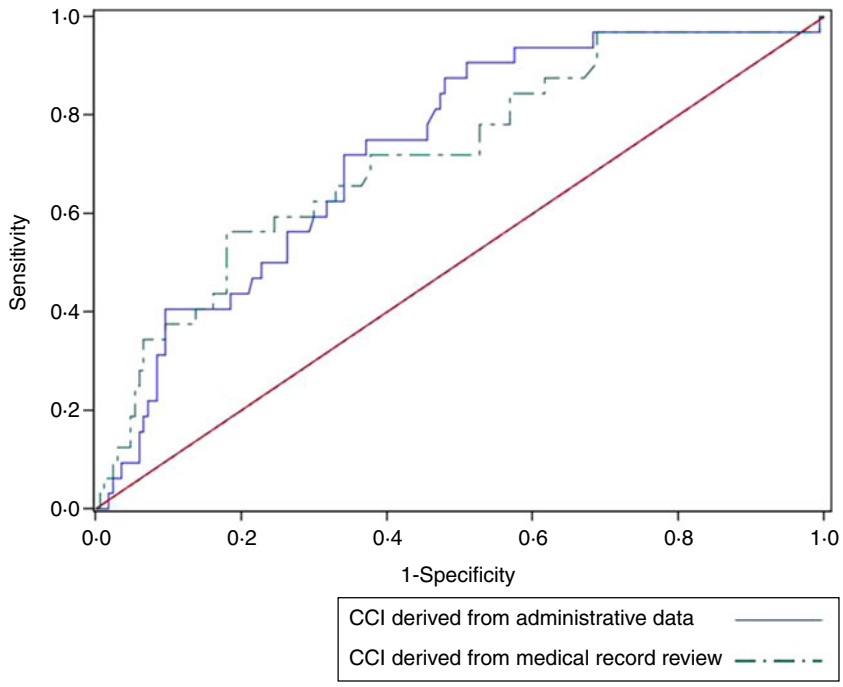

Fig. 2. Receiver operating characteristic curves comparing the area under the curve for Charlson comorbidity index (CCI) derived from administrative data and CCI derived from medical record review. The area under the curve for 30-day in-hospital mortality is shown for each CCI. CCI derived from administrative data ($C = 0{\cdot}729$) and CCI computed from medical record review ($C = 0{\cdot}717$), $P = 0{\cdot}8548$, after adjusting for age, gender, resident status, ward class, clinical speciality, illness severity, and infection source.

Ministry of Health of Singapore. As such, measurement error (if any) in the administrative-based CCI is likely to be minimal. That said, the usability of administrative data is dependent on the quality of clinical coding, which may vary from country to country, therefore the validity of the results for other countries is less certain. It has been shown that administrative databases were subjected to coding error, lack of standardized guidelines for coding of secondary diagnoses, differences in coding both across different institutions and individual coders, and difficulty in distinguishing between diseases that were present prior to admission, and those that arose during the hospitalization [17]. In some countries such as the United States, coding is also affected by insurance claims, although this is not an issue in Singapore.

Furthermore, we evaluated both administrative-based CCI and CCI derived from medical record review as predictors of inpatient mortality in patients admitted for acute infections. To the best of our knowledge, this is the first study to evaluate the two different methods of deriving CCI as predictor of inpatient mortality in patients with infections. In addition, the study assessed patients with a wide range of infections, including pneumonia, urinary tract infection, and sepsis – the commonest infections managed in hospitals. Hence, our findings may be generalized to other adult acute hospital populations.

In conclusion, administrative data-derived CCI performed as well as CCI computed from medical record review in Singapore. Administrative data offers a quicker and easier means for assessing comorbidities and confounding control in infectious disease research.

ACKNOWLEDGEMENTS

The authors thank the staff at the Institute of Infectious Diseases and Epidemiology, Tan Tock Seng Hospital for their contribution to the study. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Wong DT, Knaus WA.** Predicting outcome in critical care: the current status of the APACHE prognostic

https://doi.org/10.1017/S0950268815003271 Published online by Cambridge University Press

scoring system. *Canadian Journal of Anaesthesia* 1991; **38**: 374–383.

2. **Vincent JL, *et al.*** Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on 'sepsis-related problems' of the European Society of Intensive Care Medicine. *Critical Care Medicine* 1998; **26**: 1793–1800.
3. **Marshall JC, *et al.*** Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Critical Care Medicine* 1995; **23**: 1638–1652.
4. **Charlson ME, *et al.*** A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases* 1987; **40**: 373–383.
5. **Oltean S, *et al.*** Charlson's weighted index of comorbidities is useful in assessing the risk of death in septic patients. *Journal of Critical Care* 2012; **27**: 370–375.
6. **Elixhauser A, *et al.*** Comorbidity measures for use with administrative data. *Medical Care* 1998; **36**: 8–27.
7. **Ladha KS, *et al.*** The Deyo-Charlson and Elixhauser-van Walraven Comorbidity Indices as predictors of mortality in critically ill patients. *BMJ Open* 2015; **5**: e008990.
8. **Januel JM, *et al.*** Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. *BMC Health Services Research* 2011; **11**: 194.
9. **Deyo RA, *et al.*** Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology* 1992; **45**: 613–619.
10. **Quan H, *et al.*** Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care* 2005; **43**: 1130–1139.
11. **Thygesen SK, *et al.*** The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Medical Research Methodology* 2011; **11**: 83.
12. **Kieszak SM, *et al.*** A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *Journal of Clinical Epidemiology* 1999; **52**: 137–142.
13. **Chong WF, *et al.*** A comparison of comorbidities obtained from hospital administrative data and medical charts in older patients with pneumonia. *BMC Health Services Research* 2011; **11**: 105.
14. **Goldstein LB, *et al.*** Charlson Index comorbidity adjustment for ischemic stroke outcome studies. *Stroke* 2004; **35**: 1941–1945.
15. **Johnston MC, *et al.*** Charlson index scores from administrative data and case-note review compared favourably in a renal disease cohort. *European Journal of Public Health* 2015; **25**: 391–396.
16. **Luthi JC, *et al.*** Administrative data outperformed single-day chart review for comorbidity measure. *International Journal for Quality in Health Care* 2007; **19**: 225–231.
17. **Leal JR, Laupland KB.** Validity of ascertainment of comorbid illness using administrative databases: a systematic review. *Clinical Microbiology and Infection* 2010; **16**: 715–721.

https://doi.org/10.1017/S0950268815003271 Published online by Cambridge University Press