

Reliability of comorbidity scores derived from administrative data in the tertiary hospital intensive care setting: a cross-sectional study

Michael Hua-Gen Li,¹ Anastasia Hutchinson,² Mark Tacey,^{2,3} Graeme Duke⁴

To cite: Hua-Gen Li M, Hutchinson A, Tacey M, *et al*. Reliability of comorbidity scores derived from administrative data in the tertiary hospital intensive care setting: a cross-sectional study. *BMJ Health Care Inform* 2019;**26**:e000016. doi:10.1136/bmjhci-2019-000016

Accepted 01 January 2019

ABSTRACT

Background Hospital reporting systems commonly use administrative data to calculate comorbidity scores in order to provide risk-adjustment to outcome indicators.

Objective We aimed to elucidate the level of agreement between administrative coding data and medical chart review for extraction of comorbidities included in the Charlson Comorbidity Index (CCI) and Elixhauser Index (EI) for patients admitted to the intensive care unit of a university-affiliated hospital.

Method We conducted an examination of a random cross-section of 100 patient episodes over 12 months (July 2012 to June 2013) for the 19 CCI and 30 EI comorbidities reported in administrative data and the manual medical record system. CCI and EI comorbidities were collected in order to ascertain the difference in mean indices, detect any systematic bias, and ascertain inter-rater agreement.

Results We found reasonable inter-rater agreement (kappa (κ) coefficient ≥ 0.4) for cardiorespiratory and oncological comorbidities, but little agreement ($\kappa < 0.4$) for other comorbidities. Comorbidity indices derived from administrative data were significantly lower than from chart review: -0.81 (95% CI -1.29 to -0.33 ; $p=0.001$) for CCI, and -2.57 (95% CI -4.46 to -0.68 ; $p=0.008$) for EI.

Conclusion While cardiorespiratory and oncological comorbidities were reliably coded in administrative data, most other comorbidities were under-reported and an unreliable source for estimation of CCI or EI in intensive care patients. Further examination of a large multicentre population is required to confirm our findings.

INTRODUCTION

Administrative data have traditionally been collected to assist funding, policy formulation, and epidemiological research.¹ One example is the derivation of comorbidity scores as a measure of pre-morbid status. Such scores are useful for describing casemix and complexity, and are frequently included in risk-adjustment scores.

In Australia, the National Core Hospital-based Outcome Indicators (CHBOI)² and the Health Round Table reports are derived from coding data, include

Summary

- ▶ The reliability of intensive care coded comorbidity data has not been previously studied.
- ▶ Administrative coding of comorbidities is less reliable when more comorbidities are present.
- ▶ When monitoring quality and safety of care in the intensive care unit, the level of comorbidity present are likely to be underestimated.

comorbidity indices, and provide risk-adjustment for outcome indicators such as hospital mortality. Administrative data, while cost-effective, may have limitations in outcomes-based research.^{3–5}

Two commonly used comorbidity scores are the Charlson Comorbidity Index (CCI)⁶ and the Elixhauser Index (EI).⁷ Both comorbidity indices can be derived from hospital administrative data.⁷ The CCI comprises 19 comorbidities and was developed 45 years ago from 559 cancer patients to predict long-term survival. It can be calculated from administrative data and has been validated for use as a predictor of mortality and morbidity since its inception.^{6 8–10} The EI comprises 30 comorbidities derived from the coding data of 1.8 million hospital patients in 1992. It has been found to be superior to CCI at predicting outcomes.^{11–15} Importantly, the EI attempts to avoid contamination by comorbidities derived on or after admission, such as complications and primary diagnosis, respectively. Summary comorbidities are able to condense a large number of comorbidities to aid in mortality risk prediction. However, a major limitation between research and use of these scores is the way in which such data are collected and presented.¹⁶

Evidence examining the reliability of coding data is variable, with a systematic review by Prins *et al* suggesting that just over



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Northern Clinical Research Centre, The Northern Hospital, Epping, Victoria, Australia

²Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

³Department of Intensive Care, Box Hill Hospital, Box Hill, Victoria, Australia

⁴Department of Intensive Care, The Northern Hospital, Epping, Victoria, Australia

Correspondence to

Dr Michael Hua-Gen Li;
Michael.Huagen.Li@gmail.com

50% of comorbidities are coded in hospital discharge data.¹⁷ This is supported by one local¹⁸ and a number of international reports.^{19–25} Some studies of specific patient populations, however, suggest that coding data are reasonably reliable when compared with chart review by health professionals.^{26–29} Nevertheless, the use and interpretation of routinely collected hospital administrative data to assess patient complexity and performance indicators remain contentious.^{30–32}

There are limited data for intensive care settings where complex patients may have a higher number of comorbidities.³³ Chong *et al* suggest that the reliability of coding data may be inversely proportional to the number of comorbidities per episode.³⁴ To our knowledge, the reliability of comorbidity scores in the Australian intensive care population has not yet been examined.

Medical language processing, which automatically extracts keywords from medical charts, has been shown to be similar to manual chart review of medical records,³⁵ which remains the gold standard when assessing the reliability of coding data.³⁶ Thus, we aimed to compare the reliability of routinely collected hospital data for deriving the CCI and EI scores compared with manual chart review.

METHODS

Design

A cross-sectional study design was used. One hundred independent patient episodes were randomly selected from a total of 828 admissions to a university-affiliated, tertiary, adult, 10-bed intensive care unit (ICU) over a 12-month period from 1 July 2012 to 30 June 2013. The episodes were stratified into two equal subgroups: those requiring mechanical ventilation and those not requiring ventilation, in order to ensure that the population with higher comorbidities was captured as we theorised that patients requiring mechanical ventilation would be more likely to have a higher CCI or EI. Episodes belonging to each group were randomly sampled individually one at a time, alternatively, until each group was populated with 50 episodes. Both groups were examined for repeat episodes, which were removed, and the alternating process of random selection was continued until a total of 100 patients was reached.

All manual and electronic medical records (EMR) relating to the episode of care were routinely scanned for storage. Administrative coding data were derived from these charts and stored electronically. The researchers were not involved in any of the following duties: medical record keeping, scanning of records, and administrative data coding.

Scanned medical records for these episodes were audited by two medically-trained investigators blinded to the results of the administrative coding data. To ensure consistency, investigators cross-checked five

episodes not included in the sample to minimise investigator bias. Data collected from medical records were specific to only those included in the EI and CCI. Administrative coding diagnoses with a ‘c’ prefix, indicating a complication or diagnosis not present on admission, were excluded.

An independent analyst, blinded to the medical review and the data coding process, then extracted International Statistical Classification of Diseases and Related Health Problems, 10th revision, Australian Modification (ICD-10-AM) codes relating to all CCI and EI comorbidity diagnoses using previously validated coding algorithms³⁷ before comparing agreement with retrospective chart analysis.

Hospital ethics approval was provided by the institutional ethics committee before commencement of the study. Data for the investigation were de-identified and patient consent was deemed unnecessary.

Analysis

Frequency of each specific CCI and CI comorbidity was recorded, and CCI and EI scores were obtained.^{6,7} Paired *t*-tests were used to compare frequency of comorbidities from chart review audit and administrative data coding, with 95% confidence intervals (95% CI) of the mean reported. A Bland-Altman plot was prepared for the CCI and EI scores to determine systematic bias. The reliability between the Health Information Systems administrative coding staff and medical-trained coders was assessed by calculating kappa (κ) statistics for multiple raters.^{38,39} A $\kappa \geq 0.4$ was considered to have at least moderate association.⁴⁰ Analysis was conducted using Stata version 14 (Stata Corporation, College Station, Texas). A value of $p \leq 0.05$ was considered to be statistically significant.

RESULTS

From 1 July 2012 to 30 June 2013, there were 828 patients admitted to the ICU, and 257 (31.0%) received mechanical ventilation. The study population included 49 (19.1%) episodes requiring ventilation and 51 (8.9%) of the episodes not requiring ventilation. A total of 100 (12.0%) records were audited.

The characteristics of the study population and the two subgroups are presented in [table 1](#). Study patients had a median of 8 (IQR 5.0–12.5) coded general comorbidities.

The number of Charlson comorbidities identified by chart review (mean 2.26±1.82) was significantly greater ($p < 0.001$) than the number of Charlson comorbidities derived from administrative coding data (mean 1.39±1.19).

The mean CCI derived from the administrative data (cCCI) was 2.52 (95% CI 1.95 to 3.09). The mean CCI derived from a chart review audit (aCCI) was 3.33 (95% CI 2.77 to 3.99). There was a significant difference of

Table 1 Demographics

Variable	MV	No MV	Total	P value
Number of patients, <i>n</i>	49	51	100	
Age	66 (57–79)	59 (39–70)	65 (49–75)	0.011
Male, <i>n</i> (%)	29 (59.2)	28 (54.9)	57 (57)	0.665
ICU LOS, hours	31.0 (22.5–46.8)	97.5 (44.4–163.8)	45.8 (24.5–117.5)	<0.001
Hospital LOS, days	6 (4–9)	8 (5–15)	7 (4–12)	0.091
Death (%)	2 (4.1)	10 (19.6)	12 (12.0)	0.028
Number of comorbidities*	6 (5–10)	11 (6–15)	8 (5–12.5)	0.006
Charlson Comorbidity Score	2 (1–4)	1 (0–3)	2 (1–3.5)	0.224
Elixhauser Score	5 (3–10)	8 (3–15)	8 (3–11.5)	0.238

Patient and clinical characteristics, by mechanical ventilation status. Data presented as median (IQR) unless otherwise stated.

*Based on Primary (P) and Associated (A) ICD-10 code prefixes in diagnostic data.

ICD-10, International Statistical Classification of Diseases and Related Health Problems, 10th revision; ICU, intensive care unit; LOS, length of stay; MV, mechanical ventilation.

–0.81 (95% CI –1.29 to –0.33; $p=0.001$) between the two methods.

The Bland-Altman plot (figure 1) did not reveal evidence of any systematic bias as the CCI score

increased (taken as the average between the two methods of Charlson scores extraction).

As expected, the number of EI comorbidities identified was greater than the number of CCI comorbidities

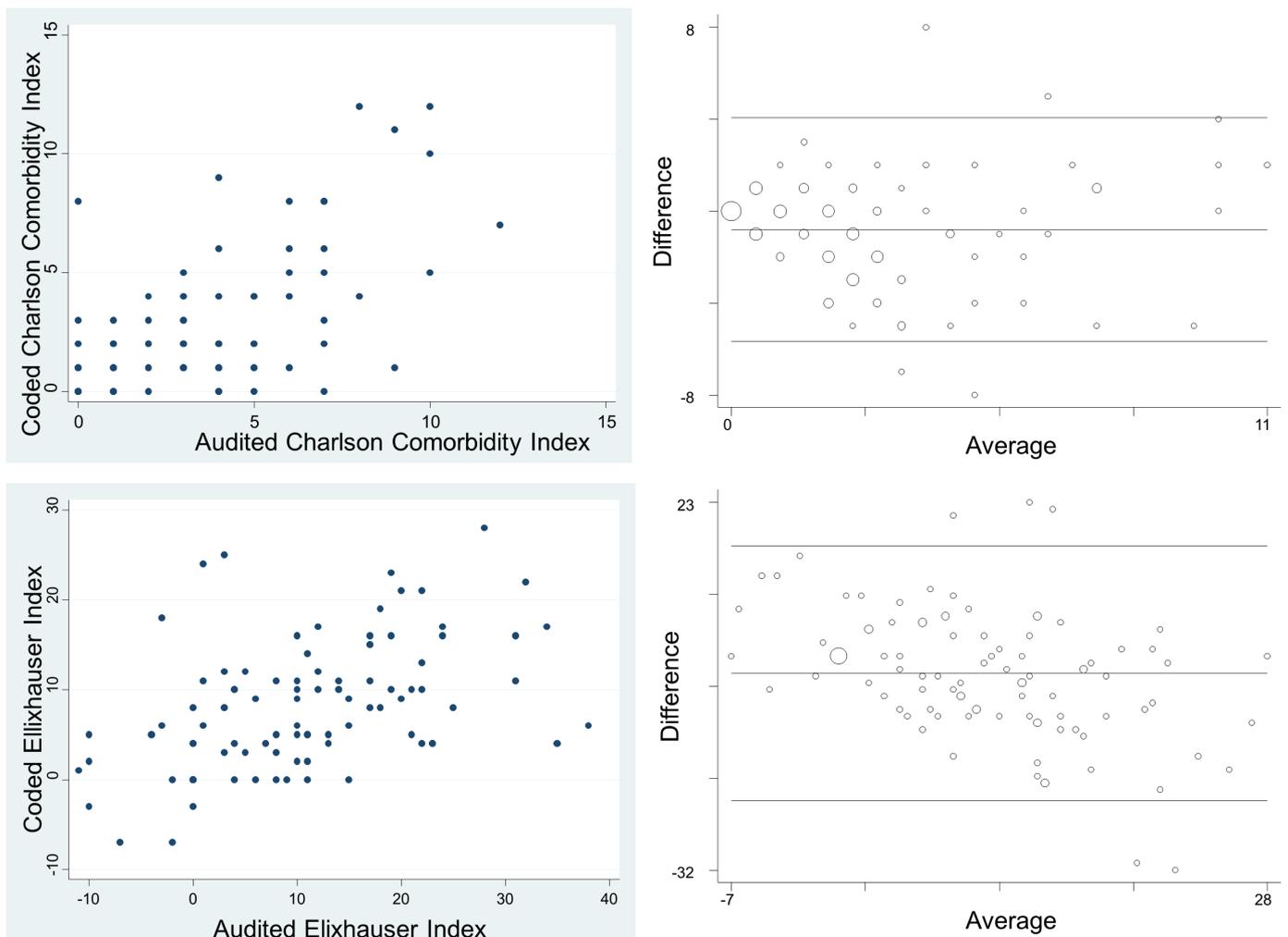


Figure 1 Results. Upper left panel: scatter plot of Charlson Comorbidity Index (CCI). Lower left panel: scatter plot for Elixhauser Index (EI). Upper right panel: Bland-Altman plot for CCI. Lower right panel: Bland-Altman Plot for EI.

in each record. The number of Elixhauser comorbidities identified by chart review (mean 4.15 ± 2.75) was significantly greater ($p < 0.001$) than the number of comorbidities derived from administrative coding data (mean 2.67 ± 1.66).

The mean EI derived from the administrative data (cEI) was 7.96 (95% CI 6.55 to 9.37) and the mean EI derived from a chart review audit (aEI) was 10.53 (95% CI 8.42 to 12.64). Thus, there was a significant difference of -2.57 (95% CI -4.46 to -0.68 ; $p = 0.008$) between the two EI scores.

Unlike the CCI, the Bland-Altman plot (figure 1) for EI did indicate a bias in the difference between coded and audited EI scores. For low range EI scores, the administrative (coding) data produced a greater score than chart review audit scores, whereas the reverse applied for high range EI scores.

The kappa statistic revealed a moderate to high ($\kappa \geq 0.4$) level of inter-rater agreement in only seven (37%) of the CCI comorbidities: congestive heart failure (CHF), myocardial infarction (MI), diabetes

mellitus with complications (DMC), chronic kidney disease (CKD), metastatic cancer, solid-organ cancer, and peripheral vascular disease (PVD) (table 2). The kappa statistic for EI comorbidities showed a moderate to high level of inter-rater agreement for the same group of comorbidities (except MI and PVD), and also for hypertension, chronic pulmonary disease (COPD), anaemia, and drug abuse (table 3).

All remaining comorbidities had a lower level of inter-rater agreement ($\kappa < 0.4$) in 12 (63%) of the CCI and 21 (70%) of the EI comorbidities.

DISCUSSION

We undertook a retrospective cross-sectional review of patient records (chart review) and administrative coding data for comorbidities in 100 patients admitted to an adult general intensive care ward. We found that administrative data significantly under-reported comorbidities present in the patient records in the majority of cases.

Table 2 Charlson Comorbidity Index inter-rater agreement

Charlson Comorbidity Index	Coded records (n)	Audited records (n)	% agreement	κ coefficient	95% CI	Prob >Z
Congestive heart failure	15	28	83.0	0.51	0.32 to 0.70	<0.001
Cerebrovascular disease	2	4	96.0	0.32	-0.18 to 0.81	<0.001
Diabetes mellitus	10	15	90.0	0.32	-0.06 to 0.58	<0.001
Diabetes mellitus with end-organ damage	21	11	72.6	0.64	0.44 to 0.84	<0.001
Dementia	0	5	95.0	0.00	-	-
Connective tissue disease	0	4	96.0	0.00	-	-
AIDS	0	0	/	/	/	/
Mild liver disease	3	16	83.0	0.06	-0.12 to 0.24	0.203
Metastatic solid tumour	10	9	95.0	0.71	0.47 to 0.95	<0.001
Any solid tumour	16	15	85.0	0.43	0.19 to 0.67	<0.001
Myocardial infarction	20	25	79.0	0.40	0.19 to 0.61	<0.001
Paraplegia	1	4	95.0	-0.02	-0.05 to 0.02	0.581
Gastric ulcer disease	1	20	79.0	-0.02	-0.06 to 0.02	0.692
Chronic pulmonary disease	13	32	79.0	0.43	0.24 to 0.61	<0.001
Peripheral vascular disease	3	10	93.0	0.44	0.10 to 0.77	<0.001
Renal disease	22	21	85.0	0.56	0.36 to 0.76	<0.001
Moderate to severe liver disease	1	4	95.0	-0.02	-0.05 to 0.02	0.581
Leukaemia	1	2	97.0	-0.01	-0.04 to 0.02	0.557
Lymphoma	0	1	99.0	0.00	0.00	-

Inter-rater agreement in Charlson Comorbidity Index as specified by kappa (κ) values between audited data and coded data by disease.

Table 3 Elixhauser Index inter-rater agreement

Elixhauser Index	Coded records (n)	Audited records (n)	% agreement	κ value	95% CI	Prob >Z
Congestive heart failure	15	28	85.0	0.57	0.38 to 0.75	<0.001
Cardiac arrhythmia	17	21	76.0	0.22	-0.00 to 0.45	0.013
Valvular disease	2	10	92.0	0.31	-0.02 to 0.64	<0.001
Pulmonary circulation disorder	3	2	97.0	0.39	-0.17 to 0.94	<0.001
Peripheral vascular disease	6	12	88.0	0.28	-0.02 to 0.57	<0.001
Hypertension	33	43	74.0	0.49	0.28 to 0.63	<0.001
Paralysis	1	2	97.0	-0.01	-0.01 to 0.02	0.557
Other neurological disorder	13	21	80.0	0.30	0.07 to 0.53	0.009
Chronic pulmonary disease	13	30	79.0	0.40	0.21 to 0.60	<0.001
Diabetes, uncomplicated	7	15	86.0	0.30	0.03 to 0.56	<0.001
Diabetes with complications	30	13	81.0	0.46	0.27 to 0.65	<0.001
Hypothyroidism	1	7	94.0	0.24	-0.15 to 0.62	<0.001
Renal failure	22	33	81.0	0.53	0.35 to 0.71	<0.001
Liver disease	6	20	80.0	0.15	-0.06 to 0.36	0.029
Peptic ulcer disease	0	19	81.0	0.00	-	-
Lymphoma	0	3	97.0	0.00	-	-
Metastatic cancer	4	9	93.0	0.43	0.09 to 0.77	<0.001
Solid tumour without metastases	16	13	85.0	0.40	0.15 to 0.65	<0.001
Rheumatoid arthritis	0	4	95.0	0.00	-	-
Coagulopathy	7	3	92.0	0.16	-0.02 to 0.41	0.035
Obesity	4	20	82.0	0.20	-0.04 to 0.33	0.003
Weight loss	12	1	89.0	0.14	-0.11 to 0.38	0.003
Fluid and electrolyte disorders	30	29	73.0	0.35	-0.15 to 0.55	<0.001
Blood loss anaemia	0	9	91.0	0.00	-	0.500
Deficiency anaemia	6	10	94.0	0.60	0.30 to 0.89	<0.001
Alcohol abuse	8	10	88.0	0.27	-0.03 to 0.57	0.003
Drug abuse	4	9	93.0	0.43	0.09 to 0.77	<0.001
Psychosis	2	8	94.0	0.38	0.00 to 0.76	<0.001
Depression	5	17	86.0	0.31	0.06 to 0.56	<0.001
AIDS	0	0	/	/	/	/

Inter-rater agreement in Elixhauser Index as specified by kappa (κ) values between audited data and coded data by disease.

Our findings are, in general, consistent with several previous reports.^{17–25}

In contrast to our overall findings, we found a small number of comorbidities that were reliably reported ($\kappa \geq 0.4$) in the administrative (coding) data. These were CHF, CKD, DMC, solid-organ cancer, and metastatic cancer.

In 1999, Kieszak *et al* performed a study examining the CCI of carotid endarterectomy cases at a single health service.²⁵ Coded data obtained from an administrative database were compared with a medical chart review and concluded that medical chart review was superior to audited data. A few years later, Quan *et al* conducted a similar study looking at all inpatients in a large health service and showed that, overall, coded data tended to

under-report comorbidities.⁴¹ Youssef *et al* examined data for general medical inpatients in Saudi Arabia and drew a similar conclusion.²⁹ Recently, this has been confirmed in a Norwegian general intensive care population by Stavem *et al*⁴² (table 4). In addition to those comorbidities in our study that were more reliably reported, Stavem *et al*'s individual comorbidities were also more reliably coded for cerebrovascular disease, dementia, and mild liver disease. As our institution has a similar casemix and size to their study, such differences could be accounted for by differences in coding methodology. Nevertheless, from two studies in separate countries, it is clear that certain comorbidities are more reliably coded than others and may provide guidance regarding data that should be included in risk-prediction models when

Table 4 Audited and coded inter-rater agreement

Condition	This study	Youssef ²⁹	Quan ⁴¹	Kieszak ²⁵	Stavem ⁴²
Myocardial infarction	0.40	0.34	0.59	0.22	0.66
Congestive heart failure	0.51	0.61	0.80	0.38	0.45
Peripheral vascular disease	0.44	0.42	0.34	0.22	0.47
Cerebrovascular disease	0.32	0.83	0.50	–	0.56
Dementia	0	0.66	0.42	0.26	0.42
Chronic pulmonary disease	0.43	0.78	0.72	0.64	0.65
Hemiplegia/paraplegia	–0.02	0.22	0.55	–	0.31
Rheumatologic disease	/	0.58	0.57	0.20	0.29
Peptic ulcer disease	–0.02	0.62	0.63	0.12	0.31
Diabetes	0.32	0.75	0.74	0.68	0.60
Diabetes with chronic complication	0.64	0.34	0.58	0.16	0.42
Mild liver disease	0.06	0.09	0.53	0	0.42
Moderate liver disease	–0.02	0.43	0.47	–	0.26
Renal disease	0.56	0.75	0.49	0.29	0.60
Any malignancy	0.43	0.77	0.78	0.23	0.49
Metastatic solid tumour	0.71	0.87	0.87	0.20	0.56
AIDS/HIV	/	0.91	0.78	–	–

Inter-rater agreement (kappa values) between audited and coded.

comparing health services in different geographical locations.

We selected adult admissions to an intensive care setting at a tertiary hospital with a high proportion of mechanically-ventilated patients because we expected these patients to more likely have comorbidities, and these comorbidities were likely to influence the level of casemix funding and thus be more reliably coded. It is not unexpected for chronic conditions that do not require intervention and do not affect funding to be excluded during the coding process. We did not ascertain the effect of coding on funding of patient episodes since this was not our primary aim; however, it has been suggested that the CCI is an inadequate predictor of resource utilisation.⁴³

While there were no statistically significant differences in CCI and EI scores between the ventilated and non-ventilated patients, non-ventilated patients had a higher number of coded comorbidities, were more likely to stay in intensive care for longer, and had an increased incidence of mortality in the same admission. This may be explained by the possibility that the non-ventilated patient group might have included a sizeable population in which ventilation was either not deemed to be of therapeutic benefit because of a lack of a clear indication, or because of a poor prognosis due to a number of other comorbidities that might not have been captured by the CCI or EI. Overall, our observation of lower inter-rater agreement compared with other hospital settings^{25 29 41} is consistent with the hypothesis that coding reliability may be inversely proportional to the number of comorbidities.³⁴

The Charlson methodology is more commonly used in risk-adjustment than the Elixhauser methodology,⁷ even

though it was derived from a small and specific cancer population using chart review. The EI, which was derived from administrative data from a large population and broad casemix, identifies a higher number of comorbidities. Our results suggest that the under-reporting of EI is comparable with the CCI, and that administrative data may not be reliable in generating either CCI or EI scores for intensive care patients.

There are several practical implications of our findings. The use of administrative data in ICUs to predict mortality through use of the CCI and EI should be viewed with a great degree of caution. The Charlson comorbidities and the derived CCI score are commonly used for risk-adjustment in several mortality prediction models constructed from administrative data. In Victoria, these include the Health Round Table Reports,⁴⁴ the National CHBOI mortality index,² and the Dr Foster methodology.⁴⁵ This is in contrast to models such as the Critical Care Outcome Prediction Equation (COPE) and Hospital Outcome Prediction Equation (HOPE), which do not include comorbidities.^{46 47} Based on our study, predicted morbidity and mortality in ICUs is likely to be under-reported when such models are based on administrative coding data.

If the CCI or EI are included in a mortality prediction model, such as a hospital standardised mortality ratio (HSMR) that is derived from administrative data, then several errors may result. First, a systemic bias due to under-reporting will be incorporated into the model. Reliance on administrative data for CCI may result in under-reporting of comorbidities and incomplete assessment of patient risk.

Second, any variation in reporting of comorbidities between institutions will lead to misleading comparative results. A health service that under-reports comorbidities will have lower CCI and EI scores resulting in these patients appearing to be healthier. This will reduce the size of the mortality denominator and produce a higher than expected HSMR. Such an institution will misleadingly appear to be a poor performer.

Thirdly, chart review of a random selection of patients may aid a 'poor performing' health service in identifying this as a potential source of bias in their report card. A better solution is for prediction models to identify and incorporate only those comorbidities that are reliably coded (CHF, CKD, COPD, cancer), rather than rely on the less accurate index scores (such as CCI and EI) that incorporate comorbidities that are unreliably reported. The optimal source from where not only the most accurate, but also the most efficient, CCI can be obtained also warrants further investigation.⁴⁸ The increasing prevalence of EMR provides a potential for capturing large data more uniformly.³⁶ With this, questions are raised regarding which types of algorithms are more effective and whether medical language processing can be standardised across different practice settings and health services.⁴³ Furthermore, the widespread use of EMR for national safety and quality purposes requires standardisation of data management processes and compliance with regulatory requirements.⁴⁹

Our study had a number of limitations. The sample size was relatively small and limited to a single site, reducing the precision of the estimates and the power to detect differences for some conditions. We conducted our study in the intensive care setting, and our results may not be generalisable to other patient groups, departmental settings, or hospital sites. We found evidence of systematic bias in the EI score that may reflect local coding rules. Our results should be viewed with caution and require validation in a larger cohort.

CONCLUSIONS

Our findings suggest that there is under-reporting of comorbidities that are necessary to calculate the CCI and the EI in administrative data for seriously ill patients, such as those admitted to the intensive care ward. Derived total (CCI and EI) scores may produce misleading results. Consideration should be given to limiting and validating a revised CCI, using an alternative comorbidity model, or negating comorbidities entirely when calculating the HSMR as has been done by the COPE and HOPE models. Further studies are required to establish the reliability of the CCI and EI in other patient groups.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which

permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Dean BB, Lam J, Natoli JL, *et al*. Review: use of electronic medical records for health outcomes research: a literature review. *Med Care Res Rev* 2009;66:611–38.
2. Australian Commission of Safety and Quality in Health Care. *National core, hospital-based outcome indicator specification, consultation draft*. Sydney: ACSQHC, 2012.
3. Green J, Wintfeld N. How accurate are hospital discharge data for evaluating effectiveness of care? *Med Care* 1993;31:719–31.
4. Price J, Estrada CA, Thompson D. Administrative data versus corrected administrative data. *Am J Med Qual* 2003;18:38–45.
5. Bohensky MA, Jolley D, Pilcher DV, *et al*. Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. *J Crit Care* 2012;27:422.e11–422.e21.
6. Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
7. Elixhauser A, Steiner C, Harris DR, *et al*. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
8. Charlson M, Szatrowski TP, Peterson J, *et al*. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994;47:1245–51.
9. Quan H, Li B, Couris CM, *et al*. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol* 2011;173:676–82.
10. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613–9.
11. Baldwin L-M, Klabunde CN, Green P, *et al*. In search of the perfect comorbidity measure for use with administrative claims data: does it exist? *Med Care* 2006;44:745–53.
12. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Med Care* 2004;42:355–60.
13. Stukenborg GJ, Wagner DP, Connors AF. Comparison of the performance of two comorbidity measures, with and without information from prior hospitalizations. *Med Care* 2001;39:727–39.
14. Mehta HB, Dimou F, Adhikari D, *et al*. Comparison of comorbidity scores in predicting surgical outcomes. *Med Care* 2016;54:180–7.
15. Sharabiani MTA, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. *Med Care* 2012;50:1109–18.
16. Austin SR, Wong Y-N, Uzzo RG, *et al*. Why summary comorbidity measures such as the Charlson comorbidity index and Elixhauser score work. *Med Care* 2015;53:e65–72.
17. Prins H, Hasman A. Appropriateness of ICD-coded diagnostic inpatient hospital discharge data for medical practice assessment. A systematic review. *Methods Inf Med* 2013;52:3–17.
18. Powell H, Lim LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease. an Australian perspective. *J Clin Epidemiol* 2001;54:687–93.
19. 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.
20. Seo H-J, Yoon S-J, Lee S-I, *et al*. A comparison of the Charlson comorbidity index derived from medical records and claims data from patients undergoing lung cancer surgery in Korea: a population-based investigation. *BMC Health Serv Res* 2010;10:1–8.
21. Leal JR, Laupland KB. Validity of ascertainment of co-morbid illness using administrative databases: a systematic review. *Clin Microbiol Infect* 2010;16:715–21.
22. Malenka DJ, McLerran D, Roos N, *et al*. Using administrative data to describe casemix: a comparison with the medical record. *J Clin Epidemiol* 1994;47:1027–32.
23. Romano PS, Roos LL, Luft HS, *et al*. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. ischemic heart disease patient outcomes research team. *J Clin Epidemiol* 1994;47:249–60.
24. Hawker GA, Coyte PC, Wright JG, *et al*. Accuracy of administrative data for assessing outcomes after knee replacement surgery. *J Clin Epidemiol* 1997;50:265–73.

25. Kieszak SM, Flanders WD, Kosinski AS, *et al.* A comparison of the Charlson comorbidity index derived from medical record data and administrative billing data. *J Clin Epidemiol* 1999;52:137–42.
26. Humphries KH, Rankin JM, Carere RG, *et al.* Co-morbidity data in outcomes research: are clinical data derived from administrative databases a reliable alternative to chart review? *J Clin Epidemiol* 2000;53:343–9.
27. Newschaffer CJ, Bush TL, Penberthy LT. Comorbidity measurement in elderly female breast cancer patients with administrative and medical records data. *J Clin Epidemiol* 1997;50:725–33.
28. Sarfati D, Hill S, Purdie G, *et al.* How well does routine hospitalisation data capture information on comorbidity in New Zealand? *N Z Med J* 2010;123:50–61.
29. Youssef A, Alharthi H. Accuracy of the Charlson index comorbidities derived from a hospital electronic database in a teaching hospital in Saudi Arabia. *Perspect Health Inf Manag* 2013;10.
30. Scott IA, Ward M. Public reporting of hospital outcomes based on administrative data: risks and opportunities. *Med J Aust* 2006;184:571–5.
31. Narins CR, Dozier AM, Ling FS, *et al.* The influence of public reporting of outcome data on medical decision making by physicians. *Arch Intern Med* 2005;165:83–7.
32. Thomas JW, Hofer TP. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. *Med Care* 1999;37:83–92.
33. Wenner JB, Norena M, Khan N, *et al.* Reliability of intensive care unit admitting and comorbid diagnoses, race, elements of Acute Physiology and Chronic Health Evaluation II score, and predicted probability of mortality in an electronic intensive care unit database. *J Crit Care* 2009;24:401–7.
34. Chong WF, Ding YY, Heng BH. A comparison of comorbidities obtained from hospital administrative data and medical charts in older patients with pneumonia. *BMC Health Serv Res* 2011;11.
35. Chuang J-H, Friedman C, Hripcsak G. A comparison of the Charlson comorbidities derived from medical language processing and administrative data. *Proc AMIA Symp* 2002:160–4.
36. Misset B, Nakache D, Vesin A, *et al.* Reliability of diagnostic coding in intensive care patients. *Crit Care* 2008;12.
37. Quan H, Sundararajan V, Halfon P, *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130–9.
38. Green AM. *Kappa statistics for multiple raters using categorical classifications*. San Diego CA: Proceedings of the 22nd Annual SAS User Group International Conference, 1997: 1110–5.
39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
40. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
41. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Med Care* 2002;40:675–85.
42. Stavem K, Hoel H, Skjaker SA, *et al.* Charlson comorbidity index derived from chart review or administrative data: agreement and prediction of mortality in intensive care patients. *Clin Epidemiol* 2017;9:311–20.
43. Toson B, Harvey LA, Close JCT. The ICD-10 Charlson comorbidity index predicted mortality but not resource utilization following hip fracture. *J Clin Epidemiol* 2015;68:44–51.
44. The Health Roundtable. *Everything you wanted to know about the health roundtable Surry hills*. NSW: The Health Roundtable, 2015.
45. Foster D. About Usus London, United Kingdom Telstra health, 2016. Available: <http://www.drfooster.com/about-us>
46. Duke GJ, Barker A, Rasekaba T, *et al.* Development and validation of the critical care outcome prediction equation, version 4. *Crit Care Resusc* 2013;15:191–7.
47. Duke GJ, Graco M, Santamaria J, *et al.* Validation of the hospital outcome prediction equation (HOPE) model for monitoring clinical performance. *Intern Med J* 2009;39:283–9.
48. Ng X, Low AHL, Thumboo J. Comparison of the Charlson comorbidity index derived from self-report and medical record review in Asian patients with rheumatic diseases. *Rheumatol Int* 2015;35:2005–11.
49. Markoska K, Spasovski G. Clinical data collection and patient phenotyping. In: Vlahou A, Mischak H, Zoidakis J, *et al.*, eds. *Integration of omics approaches and systems biology for clinical applications*. Hoboken NY: Wiley, 2018: p. 3–10.