



Identifying New-Onset Conditions And Pre-Existing Conditions Using Lookback Periods In Australian Health Administrative Datasets

Author

Palamuthusingam, Dharmenaan, Ratnayake, Gishan, Kuenstner, Kym, Hawley, Carmel M, Pascoe, Elaine M, Jose, Matthew D, Johnson, David W, Fahim, Magid

Published

2020

Journal Title

International Journal for Quality in Health Care

Version

Accepted Manuscript (AM)

DOI

[10.1093/intqhc/mzaa154](https://doi.org/10.1093/intqhc/mzaa154)

Downloaded from

<http://hdl.handle.net/10072/399865>

Griffith Research Online

<https://research-repository.griffith.edu.au>

Original Research Article

Identifying new-onset conditions and pre-existing conditions using lookback periods in Australian health administrative datasets

DHARMENAAN PALAMUTHUSINGAM^{1,2,3}, GISHAN RATNAYAKE⁴, KYM KUENSTNER^{5,6}, CARMEL M. HAWLEY^{7,2}, ELAINE M. PASCOE⁸, MATTHEW D. JOSE^{9,10}, DAVID W. JOHNSON^{7,2,11} and MAGID FAHIM^{7,2,12}

¹Metro South Integrated Nephrology and Transplant Services, Logan Hospital, Armstrong Road & Loganlea Road, Meadowbrook, QLD 4131, Australia, ²Department of Medicine, University of Queensland, 20 Weightman St, St Lucia, QLD 4072, Australia, ³School of Medicine, Griffith University, 68 University Dr, Mount Gravatt, QLD 4122, Australia, ⁴Radiation Oncology, Princess Alexandra, 31 Raymond Terrace, South Brisbane, QLD 4101, Australia, ⁵Health Information Surgery and Critical Care, The Prince Charles Hospital, 627 Rode Rd, Chermside, QLD 4032, Australia, ⁶Health Information Management Association of Australia, 51 Wicks Rd, North Ryde, NSW 2113, Australia, ⁷Metro South Integrated Nephrology and Transplant Services, Princess Alexandra Hospital, 199 Ipswich Road, Woolloongabba, QLD 4102, Australia, ⁸Centre for Health Services Research, University of Queensland, 48 Liverpool street, QLD 4102, Australia, ⁹Department of Nephrology, Royal Hobart Hospital, Hobart, 34 Cornwall St, Woolloongabba, TAS 7000, Australia, ¹⁰Faculty of Medicine, University of Tasmania, 17 Liverpool St, Hobart TAS 7000, Australia, ¹¹Translational Research Institute, 37 Kent street, Woolloongabba, QLD 4102, Australia, and ¹²Metro North Hospital and Health Service, Royal Brisbane and Women's Hospital, Herston, Butterfield St, QLD 4029, Australia

Address reprint request to: Dharmenaan Palamuthusingam, Metro South Integrated Nephrology and Transplant Services, Logan Hospital, Armstrong Road and Loganlea Road, Meadowbrook, QLD 4131, Australia. Phone: +61409891385; Fax: +61731765747; E-mail: Dharmenaan.palamuthusingam@health.qld.gov.au

Received 24 August 2020; Editorial Decision 7 November 2020; Revised 29 September 2020; Accepted 17 November 2020

Abstract

Background: The condition onset flag (COF) variable was introduced into the hospitalization coding practice in 2008 to help distinguish between the new and pre-existing conditions. However, Australian datasets collected prior to 2008 lack the COF, potentially leading to data waste. The aim of this study was to determine if an algorithm to lookback across the previous admissions could make this distinction.

Methods: All patients requiring kidney replacement therapy (KRT) identified in the Australia and New Zealand Dialysis and Transplant Registry in New South Wales, South Australia and Tasmania between July 2008 and December 2015 were linked with hospital admission datasets using probabilistic linkage. Three different lookback periods entailing either one, two or three admissions prior to the index admission were investigated. Conditions identified in an index admission but not in the lookback periods were classified as a new-onset condition. Conditions identified in both the index admission and the lookback period were deemed to be pre-existing. The degrees of agreement were determined using the kappa statistic. Conditions examined for new onset were myocardial infarction, pulmonary embolism and pneumonia. Conditions examined for prior existence were diabetes mellitus, hypertension and kidney failure. Secondary analyses evaluated whether the conditions identified as pre-existing using COF were captured consistently in the subsequent admissions.

Results: 11 140 patients on KRT with 69 403 admissions were analysed. Lookback over a single admission interval (Period 1) provided the highest rates of true positives with COF for all three

new-onset conditions, ranging from 89% to 100%. The levels of agreement were almost perfect for all conditions ($k = 0.94\text{--}1.00$). This was consistent across the different time eras. All lookback periods identified additional new-onset conditions that were not classified by COF: Lookback Period 1 picked up a further 474 myocardial infarction, 84 pulmonary embolism and 1092 pneumonia episodes. Lookback Period 1 had the highest percentage of true positives when identifying the pre-existing conditions (64–80%). The level of agreement was moderate to strong and was similar across the time eras. Secondary analysis showed that not all pre-existing conditions identified using COF carried forward to the subsequent admission (61–82%) but increased when looking forward across >1 admission (87–95%).

Conclusion: The described algorithm using a lookback period is a pragmatic, reliable and robust means of identifying the new-onset and pre-existing patient conditions, thereby enriching the existing datasets predating the availability of the COF. The findings also highlight the value of concatenating a series of hospital patient admissions to more comprehensively adjudicate the pre-existing conditions, rather than assessing the index admission alone.

Key words: International Classification of Disease, hospital complications, comorbidity, admissions, administrative datasets

Background

Hospital admission datasets are statutory collections containing administrative, demographic and clinical data pertaining to each patient admission for both private and public hospitals. In Australia, clinical information is coded by trained professionals using the International Classification of Diseases (ICD), an international standard diagnostic classification, and the Australian Classification of Health Interventions (ACHI), an Australian standard classification system for procedures and interventions [1, 2]. For each hospital admission, ICD/ACHI codes are assigned in a single linear entry, with the first diagnosis indicating the principal admission diagnosis followed by either new associated diagnoses arising during admission or pre-existing diagnoses, in no particular order. As such, these datasets contain a wealth of information that can be used for quality assurance, service planning and research.

To maximize the potential usefulness of hospital admission datasets for research purposes, researchers need a nuanced understanding of both data content and quality. The process of care measures, such as length of stay, readmission rates, critical care utilization and inter-hospital transfer, is relatively easy to calculate retrospectively. In contrast, distinguishing between the new onset and pre-existing conditions and ascertaining a complete list of comorbid conditions is far more challenging.

Prior to 2008, differentiating new-onset conditions/diagnoses arising during the course of a person's admission from pre-existing conditions/diagnoses based on a single admission's coded diagnoses was not possible without a manual review of medical charts as the onset of condition was not flagged systematically in the codeset. Validating the recency of ICD diagnosis codes with medical chart reviews is impractical in large administrative datasets, and as such the longitudinal value of the datasets predating 2008 is severely limited. Since 1 July 2008, reporting a variable called condition onset flag (COF) was mandated on a national basis in Australia [3, 4]. This variable is used to flag each diagnosis ICD code as pre-existing, new or status unknown. The principal diagnosis is always flagged as not arising during the admission [3, 5]. As such, the Australian Commission on Safety and Quality in Health Care in collaboration with the Independent Hospital Pricing Authority utilizes this variable as one of the many measures when assessing the quality of healthcare delivery [3, 4]. However, this means that important historical data prior to

the availability of the COF variable (pre-2008) are potentially of less value.

Furthermore, previous studies have reported a poor agreement between medical chart reviews and administrative datasets in adjudicating the status of conditions and the comprehensiveness of comorbidities, but these studies were limited by small record numbers and the fact that only the index admission was evaluated rather than several consecutive admissions from patient data in adjudicating comorbidities and complications [6, 7].

The aim of this study was to report an algorithm for the secondary analysis of administrative data. The algorithm uses different lookback durations from the indexed admission to identify both the new-onset and pre-existing conditions accurately compared with those identified using the COF variable, thereby potentially allowing data predating the COF variable introduction to be used for longitudinal cohort investigations to evaluate the trends of hospital-acquired complications. A secondary aim of this study was to evaluate the consistency of pre-existing conditions using the COF variable being carried forward to the subsequent admissions, as would be expected.

Methods

Data sources: hospital admission datasets and Australia and New Zealand Dialysis and Transplant Registry

The study population was derived from three health jurisdictions in Australia. All patients in New South Wales, South Australian and Tasmanian hospital admission datasets were identified as receiving chronic kidney replacement therapy (KRT) by Australia and New Zealand Dialysis and Transplant (ANZDATA) between July 2008 and December 2015. Chronic KRT was defined as chronic in-centre dialysis, home haemodialysis, peritoneal dialysis or a functioning kidney transplant, as identified in ANZDATA. All incident KRT patients aged <18 years of age were excluded. ANZDATA is a clinical quality registry that collects patient demographic and robust dialysis treatment data prospectively on all incident and prevalent dialysis patients in Australia and New Zealand and has a near 100% participation rate [8]. Hospital admission datasets are state-based administrative datasets that record all separations, including discharges, transfers and episode of care changes from both public and

private hospitals, with the exception of Tasmania where only public hospital admissions are recorded. Patient demographics, referral origin, discharge destination on separation, diagnosis according to ICD and procedures are all reported in these datasets.

Data linkage

All incident and prevalent patients on KRT were identified in ANZ-DATA during the specified time periods and matched to individual patient records in hospital admission datasets using probabilistic record linkage. Probabilistic linkage uses a combination of personal identifier variables, such as date of birth, first name and last name, sex and address, to form probability weights for the likelihood of a particular hospital admission being associated with one person [9]. This was undertaken by each jurisdiction's authorized data linkage unit, with linkage rates >95% for each jurisdiction. New South Wales Admitted Patient Data Collection linkage was undertaken by The Centre for Health Record Linkage, South Australian Inpatient Hospital Separations by SANT DataLink and Tasmania Public Hospital Admitted Patient Collection by the Tasmanian Data Linkage Unit.

Selection of outcomes

The selection of outcomes was based on the clinical judgement and detailed examination of the ICD-10 AM codebook. Prior to selecting the outcomes, the previously validated Multipurpose Australian Comorbidity Scoring System for hospital admission datasets was reviewed to ensure that none of the codes selected overlapped [10]. Myocardial infarction, pulmonary embolism and pneumonia were selected as new-onset conditions. Diabetes mellitus, hypertension and kidney failure were chosen as pre-existing conditions as these conditions are common, have significant prognostic effects on both fatal and non-fatal outcomes and therefore are of potentially greater interest to researchers when modelling important patient outcomes. All datasets were mapped to the ICD-10AM version at the time of separation. All possible iterations of each condition across the different versions of ICD-10AM were included to account for the variations in coding (Supplementary Table 1). Further, only admissions ≥ 3 days were included, as the recording of comorbidities in day-only dialysis and day-procedure admissions are often not comprehensive.

Data extraction

New-onset condition

The index record was identified by the presence of the relevant condition (*vide supra*). Of note, a single condition may be coded by >1 ICD code and therefore all variations of an ICD code representing the condition of interest were considered (Supplementary Table 1). Conditions identified in an index admission but not in a set number of previous admission/admissions (also referred to as lookback period) were classified as a new-onset condition. Three different lookback periods were evaluated. Lookback Period 1 examined the previous admission only, Lookback Period 2 extended to the previous 2 admissions and Lookback Period 3 evaluated the previous 3 admissions (see Figure 1).

Pre-existing condition

Lookback. Similarly, the pre-existing conditions were identified by the presence of the condition in the indexed admission, as well as in each of the lookback periods.

Look forward. To investigate if concatenation of multiple coded admissions could more comprehensively identify the pre-existing conditions, the look forward algorithm examined if a condition identified as pre-existing using the COF variable in the index admission was also coded in a subsequent admission (Look Forward 1). For Look Forward 3, the algorithm examined whether a condition identified as pre-existing using the COF variable in the index admission was coded in any of the following 3 admissions.

Statistical methods

Three different lookback periods preceding the indexed admission were evaluated to best identify the optimum lookback period for ascertaining the percentage of true positives for each outcome. True-positive was present if an event was determined to be a new-onset condition arising in-hospital as adjudicated using the various lookback periods and was also deemed to be of new-onset using the COF variable as a reference standard. Agreement was also assessed using Landis and Koch's qualitative assessment of the kappa statistic (k) [11]. The possibility that events were missed by the three different lookback periods due to the absence of previous hospital separations to compare with were mitigated by restricting analyses

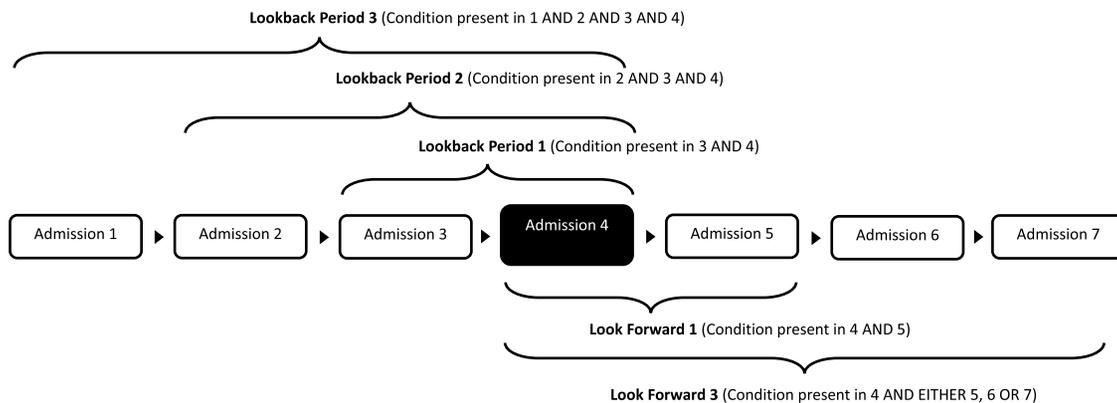


Figure 1 Lookback and look forward periods.

to records where a lookback was possible (effective sample size). This also minimized bias towards patients with a limited history of comorbidity, as it can happen in first-time admissions. The median time and interquartile range (IQR) between hospital admissions was calculated. Subgroup analysis of true-positive percentages and agreement (kappa statistic) between the lookback periods and COF was also assessed over three different time periods to assess for variation over time: July 2008 to July 2011 (Era 1), August 2011 to July 2013 (Era 2) and August 2013 to December 2015 (Era 3). Principal diagnoses were excluded as these conditions are always classified as pre-existing on admission. Admissions with no admission or separation date were also excluded.

Secondary analysis was also undertaken to investigate if a concatenation of multiple coded admissions could more comprehensively identify the pre-existing conditions. Thus, for conditions identified as pre-existing using COF = the admission/admissions following the indexed admission will be reviewed and assessed if the presence of the condition was coded again in two different look forward periods. Look Forward 1 evaluated if the condition was recorded in the very next admission; Look Forward 3 evaluated if the condition was recorded in any 1 of the 3 admissions following the indexed admission (See Figure 1). Analyses were restricted to records where a look forward was possible. This analysis was also conducted over the different time eras (Era 1–3) to examine the variation in coding practices. A one-way analysis of variance was used to determine if the percentages of pre-existing conditions carried forward to the subsequent admission were different between jurisdictions. Statistical analysis was undertaken with Stata version 14.2.

Ethics approval

Ethics approvals were received from New South Wales (HREC/17/CIPHS/41), South Australia (HREC/17/SAH/115), South Australia Aboriginal Health and Research Council (HREC 04-17-746) and Tasmania (H0017537).

Results

There were 12 490 patients identified as receiving KRT in the three jurisdictions between July 2008 and December 2015. These patients were matched with 2 846 173 hospital separations; 7085 records had missing dates. After excluding admissions <3 days, 11 140 patients with 69 403 hospital admissions remained for evaluation. The median time between admissions for each patient was 86 (IQR 23–244) days. The median age of the patients on admission was 62.6 (50.1–73.4) years with 58% being identified as males and mostly of Caucasian ethnicity (84%) (Figure 2).

Outcomes

Primary analyses

New-onset conditions. From July 2008 to December 2015, ICD codes identifying a myocardial infarction were recorded in 1052 hospital separations, of which 309 (29%) were classified as being of new-onset by the COF variable. Pulmonary embolism was recorded in 133 occasions, with 15 (11%) cases being categorized as a new condition, and pneumonia was recorded in 2087 episodes of care with 442 (21%) deemed to have arisen newly during a hospital admission.

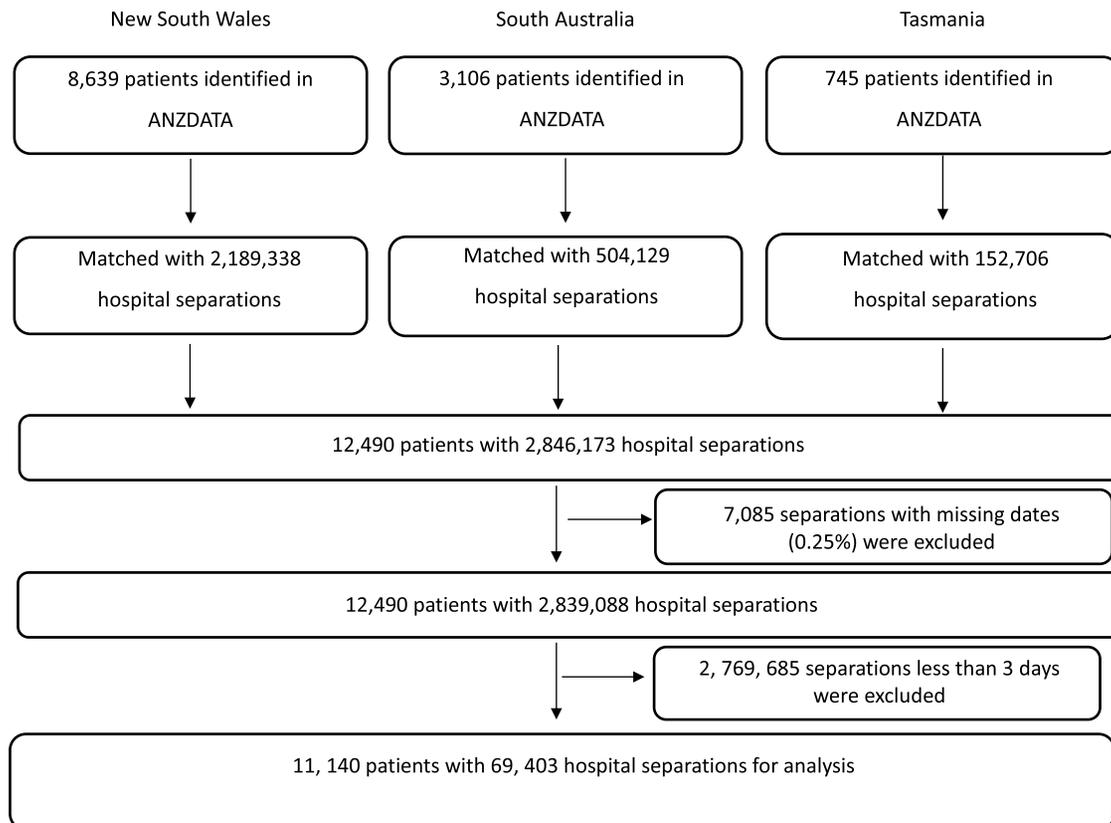


Figure 2 Patient flow chart of linkage process and inclusion into analyses.

Agreement of new-onset conditions identified using the different lookback periods with events identified using the COF variable as the reference standard

The percentage of true positives was highest for all three outcomes in Lookback Period 1: 100% for pulmonary embolism ($k = 1.00$), 93% for myocardial infarction ($k = 0.96$) and 89% for pneumonia ($k = 0.94$). True positives were lowest in Lookback Period 3: 79% for pneumonia ($k = 0.96$), 86% for myocardial infarction ($k = 0.92$) and 100% for pulmonary embolism ($k = 1.00$) (Table 1 and Table 2). Regardless of the lookback period, the level of agreement was almost perfect ($k > 0.90$) for all three conditions. Importantly, the Lookback Period 1 identified potentially additional new-onset conditions including 474 episodes of myocardial infarction, 84 pulmonary embolism occurrences and 1092 cases of pneumonia that were not identified using the COF variable.

Further analyses demonstrated that the level of agreement remained almost perfect for all three new-onset conditions across the three different eras in Lookback Period 1. This was a similar finding in Lookback Period 2 with an exception of pneumonia where the level of agreement was lower in Era 1 but remained strong. Table 2 presents the number of true positives and the agreement for each outcome over the different time eras.

Pre-existing conditions. As expected, kidney failure was the most frequently identified comorbidity using COF, followed by hypertension and diabetes mellitus. The analyses demonstrated that pre-existing conditions were not consistently coded across admissions (Table 3). Thus, while a pre-existing diagnosis of kidney failure was identified in 18 075 index admissions using the COF variable, it was not coded at all in 1015 admissions immediately preceding the index admission (Lookback Period 1).

Agreement between pre-existing conditions identified using the different lookback periods and those identified using the COF variable as the reference standard

The percentage of true positives was highest in Lookback Period 1 and lowest in Lookback Period 3 (80% vs. 56% for diabetes mellitus, 64% vs. 31% for hypertension and 75% vs. 49% for kidney failure, respectively). The level for agreement was strong for diabetes mellitus ($k = 0.86$) and moderate for hypertension ($k = 0.71$) and kidney failure ($k = 0.79$) for Lookback Period 1, whereas agreement was moderate to weak in Lookback Period 3. Table 3 presents the percentage of agreement for each pre-existing condition. Further analysis demonstrated that the level of agreement remained strong for diabetes mellitus and moderate for both hypertension and kidney failure in Lookback Period 1 across the different eras Table 4.

Secondary analyses

Secondary analysis also revealed that the conditions identified as pre-existing using the COF in an index admission were variably carried forward to subsequent admissions as shown in Table 5. Furthermore, concatenation of successive admissions improved the identification of pre-existing (chronic) conditions. Thus, diabetes mellitus was the condition that was recorded most consistently in both Look Forward 1 and Look Forward 3 (82% and 95%, respectively), followed by kidney failure (75% and 94%, respectively) and hypertension (61% and 87%, respectively). Compared to Look Forward 1, all three conditions had a higher rate of being coded again in Look Forward 3 across the different eras. There was a statistically significant difference in the consistency of coding the pre-existing conditions between jurisdictions in Look Forward 1, but not in Look Forward 3 (Table 5).

Table 1 Comparison of true-positive rates of new-onset conditions using three different lookback periods (overall)

	Total events identified using COF in dataset	Lookback Period 1		Lookback Period 2		Lookback Period 3	
		Events identified by COF (effective sample size)	Events identified using Lookback Period 1 True positives (%)	Events identified by COF (effective sample size)	Events identified using Lookback Period 2 True positives (%)	Events identified by COF (effective sample size)	Events identified using Lookback Period 3 True positives (%)
Myocardial infarction	309	252	726 235/252 (93)	195	532 180/195 (92)	150	367 129/150 (86)
Pulmonary embolism	15	11	95 11/11 (100)	6	66 6/6 (100)	5	58 5/5 (100)
Pneumonia	442	367	1459 327/367 (89)	279	1053 237/1053 (85)	214	717 168/717 (79)

Table 2 True positives (%) and agreement of new-onset conditions across eras

	Era 1	Era 2	Era 3	Overall
Total admissions	July 2008–July 2011 25 502	August 2011–July 2013 19 407	August 2013–December 2015 24 494	69 403
Median days between admissions [IQR]	74 [21–216]	96 [26–294]	89 [24–277]	86 [23–244]
Lookback Period 1 (% true positives with 95% CI)				
Myocardial infarction	92% (85–98)	89% (82–97)	97% (93–100)	93% (90–96)
Agreement (<i>k</i>)	0.96 Almost perfect	0.94 Almost perfect	0.98 Almost perfect	0.96 Almost perfect
Pulmonary embolism	0 ^a	100%	100%	100%
Agreement (<i>k</i>)	^a	1.00 Almost perfect	1.00 Almost perfect	1.00 Almost perfect
Pneumonia	84% (75–93)	90% (85–96)	90% (86–94)	89% (86–92)
Agreement (<i>k</i>)	0.91 Almost perfect	0.95 Almost perfect	0.95 Almost perfect	0.94 Almost perfect
Lookback Period 2 (% true positives with 95% CI)				
Myocardial infarction	87% (77–97)	90% (82–98)	96% (92–100)	92% (87–96)
Agreement (<i>k</i>)	0.93 Almost perfect	0.95 Almost perfect	0.98 Almost perfect	0.96 Almost perfect
Pulmonary embolism	0 ^a	0 ^a	100%	100%
Agreement (<i>k</i>)	^a	^a	1.00 Almost perfect	1.00 Almost perfect
Pneumonia	73% (60–87)	87% (79–95)	87% (82–92)	85% (81–89)
Agreement (<i>k</i>)	0.84 Strong	0.93 Almost perfect	0.93 Almost perfect	0.92 Almost perfect
Lookback Period 3 (% true positives with 95% CI)				
Myocardial infarction	72% (53–88)	86% (76–96)	91% (85–97)	86% (80–92)
Agreement (<i>k</i>)	0.83 Strong	0.92 Almost perfect	0.95 Almost perfect	0.92 Almost perfect
Pulmonary embolism	0 ^a	0 ^a	100%	100%
Agreement (<i>k</i>)	^a	^a	1.00 Almost perfect	1.00 Almost perfect
Pneumonia	72% (53–88)	86% (76–96)	80% (73–87)	79% (73–84)
Agreement (<i>k</i>)	0.86 Strong	0.87 Strong	0.89 Strong	0.88 Strong

^aNo events identified using the COF variable.

Table 3 Comparison of true-positive rates of the pre-existing conditions using three different lookback periods (Overall)

	Lookback Period 1			Lookback Period 2			Lookback Period 3		
	Events identified by COF (Effective sample size)	Events identified using Lookback Period 1	True positives (%)	Events identified by COF (Effective sample size)	Events identified using Lookback Period 2	True positives (%)	Events identified by COF (Effective sample size)	Events identified using Lookback Period 3	True positives (%)
Diabetes mellitus	12 248	11 775	9822/12 248 (80)	9915	7709	6589/9915 (67)	7913	5106	4462/7913 (56)
Hypertension	13 834	10 747	8784/13 834 (64)	10 843	5611	4699/10 843 (43)	8424	3009	2571/8424 (31)
Kidney failure	18 075	17 060	13 554/18 075 (75)	14 417	10 591	8592/14 417 (60)	11 318	6715	5530/11 318 (49)

Discussion

Statement of principal findings

The accurate measurement of in-hospital complications and patient comorbidities is necessary to assess both the quality of care delivered and the health burden on society. Administrative datasets routinely collect clinical data, which if categorized and classified judiciously can be a powerful means of providing valuable information for researchers, clinicians and health administrators. This report highlights two key findings. Firstly, the almost perfect agreement demonstrated in the analysis suggests that using an algorithm to lookback across successive coded admissions is a pragmatic, reliable and robust means of identifying both new-onset and pre-existing conditions, thereby enhancing the value of datasets predating the availability of COF (pre-2008). Secondly, pre-existing conditions are not con-

sistently reported in subsequent admissions, introducing a risk of misclassification if a single coded admission is chosen at random. The latter finding underscores the importance of concatenating a series of patient hospital admissions to reliably determine the pre-existing conditions, rather than evaluating the index admission alone.

Lookback Period 1 had the highest percentage of true positives as well as almost perfect agreement with new-onset conditions identified using the COF variable. These findings were also observed across the three different eras. The lookback periods also identified additional new-onset conditions that may have been correctly coded and misclassified by the COF.

The percentage of true-positives and level of agreement in identifying the pre-existing conditions decreased with increasing lookback periods, with the highest agreement observed in Lookback Period 1.

Table 4 True positives (%) and agreement of the pre-existing conditions across eras

	Era 1	Era 2	Era 3	
	July 2008–July 2011	August 2011–July 2013	August 2013–December 2015	Overall
Total admissions	25 502	19 407	24 494	69 403
Median days between admissions [IQR]	74 [21–216]	96 [26–294]	89 [24–277]	86 [23–244]
Lookback Period 1 true positives (95% CI)				
Diabetes mellitus	75% (73–77)	73% (71–74)	86% (85–87)	80% (79–81)
Agreement (<i>k</i>)	0.84 Strong	0.80 Strong	0.88 Strong	0.86 Strong
Hypertension	60% (58–61)	61% (59–62)	67% (66–68)	64% (63–64)
Agreement (<i>k</i>)	0.71 Moderate	0.67 Moderate	0.76 Moderate	0.71 Moderate
End-stage kidney failure	74% (73–76)	74% (73–75)	76% (75–77)	75% (74–76)
Agreement (<i>k</i>)	0.82 Strong	0.77 Moderate	0.75 Moderate	0.79 Moderate
Lookback Period 2 true positives (95% CI)				
Diabetes mellitus	59% (56–61)	56% (54–58)	74% (73–75)	67% (66–68)
Agreement (<i>k</i>)	0.72 Moderate	0.67 Moderate	0.79 Moderate	0.76 Moderate
Hypertension	40% (36–44)	39% (37–41)	48% (47–49)	43% (42–44)
Agreement (<i>k</i>)	0.52 Weak	0.48 Weak	0.55 Weak	0.54 Weak
Kidney failure	59% (58–61)	58% (57–60)	61% (60–62)	60% (59–60)
Agreement (<i>k</i>)	0.71 Moderate	0.65 Moderate	0.63 Weak	0.67 Weak
Lookback Period 3 true positives (95% CI)				
Diabetes mellitus	48% (44–51)	43% (45–47)	64% (63–65)	56% (55–58)
Agreement (<i>k</i>)	0.63 Moderate	0.57 Weak	0.72 Moderate	0.68 Moderate
Hypertension	25% (23–27)	26% (24–27)	35% (34–37)	31% (30–32)
Agreement (<i>k</i>)	0.38 Minimal	0.35 Minimal	0.44 Weak	0.42 Weak
Kidney failure	48% (46–50)	47% (45–49)	50% (49–52)	49% (48–50)
Agreement (<i>k</i>)	0.62 Moderate	0.56 Weak	0.56 Weak	0.59 Weak

Table 5 Percentage of the pre-existing conditions identified using COF that are carried forward to the subsequent admissions

	Look Forward 1 admission (95% CI)			
	July 2008–July 2011	August 2011–July 2013	August 2013–December 2015	Overall (95% CI) <i>two-tailed P value</i> *
Diabetes mellitus	1294/1887 69% (66–71)	1884/2295 82% (81–84)	4443/5004 89% (88–90)	9343/11 364 82% (81–83) <i>P</i> = 0.039
Hypertension	1276/2289 56% (54–58)	1944/3009 65% (63–66)	3272/5325 61% (60–63)	8230/13 556 61% (60–62) <i>P</i> = 0.002
Kidney failure	2263/3092 73% (72–75)	2814/3769 75% (73–76)	5013/6394 78% (77–79)	12 760/16 913 75% (74–76) <i>P</i> = 0.197
Look Forward 3 admissions (95% CI)				
Diabetes mellitus	718/786 91% (89–93)	786/819 96% (95–97)	2044/2090 98% (97–98)	6183/6477 95% (95–96) <i>P</i> = 0.623
Hypertension	722/854 85% (82–87)	958/1065 90% (88–92)	1977/2267 87% (86–89)	6702/7705 87% (86–88) <i>P</i> = 0.988
Kidney failure	1072/1153 93% (91–94)	1256/1344 93% (92–95)	2413/2529 95% (95–96)	8898/9511 94% (93–94) <i>P</i> = 0.506

*Two-tailed *P* value for one-way analysis of variance between the three different jurisdictions.

The level of agreement remained moderate to strong in Lookback Period 1, and this was a consistent finding across the three eras.

Secondary analysis showed that the pre-existing conditions identified using COF were not consistently coded in subsequent admissions even when there was no possibility of the condition having resolved. For example, hypertension was only recorded again in 61% of cases. This in part may explain why pre-existing conditions/comorbidities are underreported in research using administrative datasets (*vide infra*). However, a much higher percentage of conditions were coded again in any of the following 3 admissions (87% for hypertension and 95% for diabetes mellitus). This suggests that the evaluation of a single coded admission has a substantial likelihood of missing chronic pre-existing conditions and that, when available, multiple consecutive admissions should be examined to ensure comprehensive and reliable ascertainment of pre-existing conditions. Sole reliance on the COF variable does not mitigate this risk as the condition must be coded prior to assignment of the COF. Furthermore, the differences observed in coding between jurisdictions is likely to be due to the inherent biases in local coding practices and

the potential impact this may have on diagnosis-related group codes for each admission, which in turn affect hospital funding as some health jurisdictions are funded by the number and complexity of the patients treated.

Interpretation within the context of wider literature

Administrative datasets can be imperfect and variable when assessing the outcomes from an indexed admission. This study showed that concatenating a series of hospital admissions for each person is necessary to determine the conditions that may have varied in their recency and duration, whereas previous studies frequently evaluated coding in the index admission only. This may have been due to various reasons, including restricted data access, lack of understanding of administrative datasets or convenience. Using longitudinal data to evaluate the outcomes has the added advantage of more accurately determining incidence rates by minimizing the prevalent pool effect caused by the erroneous inclusion of prevalent cases. The fact that 100% agreement was not achieved with the various lookback periods

may be explained by the limitations of using the COF variable itself. Firstly, COF has been inconsistent in its uptake across jurisdictions and institutions. For instance, reporting of COFs was 74% in public hospitals in 2009–2010 and had only risen to 81% in 2011–2012, whilst the rates in private hospitals were even lower [3]. Similarly, higher rates of reporting were observed in larger hospitals compared to smaller sub-acute health services [3–5]. Secondly, auditing of COF accuracy is not routinely reported across jurisdictions. In fact, some datasets do not undertake quality assurance audits to validate the accuracy of COF reporting at all. In those that do, reported error rates are up to 21% in public hospitals and up to 49% in the private sector [5, 12]. Thirdly, complications arising after surgery between admissions or conditions that are managed in the outpatient setting may be omitted when estimating surgical complications, as subsequent admissions would not identify them as being pre-existing. This would also explain why more events were identified using the various lookback periods. Fourthly, clinical information is coded by trained professionals, and not clinical experts. The Australian Coding Standards Condition Onset Flag guidelines stipulate that in situations where there is ambiguity about the recency of a condition, coders are advised to default that condition as being pre-existing [3]. Further, coders are time limited and therefore it is impractical for them to review multiple previous admissions to verify recency of the conditions. Nevertheless, not infrequently, clinical coders clarify ambiguity in in-hospital discharge data with clinicians. Finally, besides clerical errors, inadequate medical documentation can make identifying the new and pre-existing conditions difficult. Notwithstanding these limitations, the lack of subgroup specification detail in ICD diagnosis and procedure codes does not allow for more granular coding of clinical information. Subsequent revisions are more comprehensive. For instance, ICD-9 had approximately 17 000 codes, whereas ICD-10 has now incorporated more than 1 55 000 codes, permitting increased specificity for describing a diagnosis. Importantly, researchers need to acknowledge that transitions between different versions of the ICD coding schema may also result in disruption in the observed trends [13].

Strengths and limitations

A major strength of our study was both the large numbers of patients ($n = 11\,140$) and matched records ($n = 69\,403$) for analysis. The patients in this study also spanned across multiple jurisdictions, including both large metropolitan and regional hospitals. A limitation of this study was that it only involved patients with kidney failure receiving chronic KRT. In addition, due to the large numbers of admissions involved, it was not possible to evaluate the observed discrepancies in reporting by reviewing medical records. Previous studies have demonstrated suboptimal accuracy of coding complications in administrative datasets when compared to clinical chart reviews, and in fact, have concluded that there is a systematic under-reporting of new and pre-existing conditions [6, 7]. Interestingly, Reilly *et al.* also demonstrated poor inter-rater agreement between anaesthetists in identifying postoperative complications in their retrospective chart review of 482 patients undergoing surgery, suggesting that this approach itself has shortcomings. Notably, these studies are limited by small numbers.

Conclusion

The findings of this study highlight the value of concatenating longitudinal patient data to improve the reliability of coding of pre-existing conditions, rather than using coding information from the

index admission only. The pragmatic algorithm described herein provides a reliable and robust approach to identifying the recency of conditions in health datasets predating the availability of the COF variable, thereby improving patient-orientated research, enriching existing datasets and minimizing data waste. These findings have important implications for future researchers using administrative databases to quantify new-onset and pre-existing conditions when modelling important patient outcomes.

Acknowledgements

None declared.

Conflicts of interests

D.P. has received speaking honoraria from the Australian Medical Forum. D.J. is a current recipient of an Australian National Health and Medical Research Council Practitioner Fellowship. D.J. has previously received consultancy fees, research grants, speaker's honoraria and travel sponsorships from Baxter Healthcare and Fresenius Medical Care. CH has received funding from Janssen and GlaxoSmithKline to her institution for trial steering committee roles and research grant support to her institution from Shire, Baxter, Fresenius, and Otsuka and travel sponsorship from Otsuka.

Funding

This work was supported by the Royal Australasian College of Physicians [D.P. is the recipient of the Royal Australasian College of Physicians Jacquot Research Entry Scholarship, 2019–2020] and the Metro South Health Research Support Scheme Small Project Grant, 2018.

Contributorship

D.P., G.R., K.K., C.H., E.P., M.J., D.J. and M.F. contributed to the study concept and design. D.P., E.P., C.H., D.J., M.J. and M.F. were responsible for the acquisition of data. D.P., G.R., E.P. and M.F. conducted the analysis of the data. D.P., G.R., K.K., C.H., E.P., M.J., D.J. and M.F. contributed to the interpretation of the results. D.P. drafted the manuscript with critical revision from all authors. M.F. was the study supervisor.

Ethics and other permissions

Not applicable.

Data availability statement

The data underlying this article cannot be shared publicly as stipulated by the respective data custodians and data-linkage units of each jurisdiction. Data will be shared on reasonable request to the corresponding author with permission from the respective data custodians and data-linkage units.

Supplementary material

[Supplementary material](#) is available at *International Journal for Quality in Health Care* online.

References

1. Independent Hospital Pricing Authority. *Chronicle of the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM)*. 1st to 10th edn. Darlinghurst, New South Wales: Independent Hospital Pricing Authority, 2017, 1–2736.
2. Independent Hospital Pricing Authority. *Chronicle for Australian Classification of Health Interventions (ACHI)*. 1st to 10th edn. Darlinghurst, New South Wales: Independent Hospital Pricing Authority, 2017, 1–1425.
3. Australian Commission on Safety and Quality in Health Care. *Analysis of Hospital-acquired Diagnoses and Their Effect on Case Complexity and Resource Use – Final Report, in Health Policy Analysis 2013*. Sydney, Australia: Health Policy Analysis Pty Ltd, 2013.
4. Independent Hospital Pricing Authority. *Risk Adjustment Model for Hospital Acquired Complications - Technical Specifications*. Darlinghurst, New South Wales: Independent Hospital Pricing Authority, 2017.
5. Utz M, Johnson T, Halech R. *A Review of the Classification of Hospital-acquired Diagnoses (Chadx)*. Sydney, New South Wales, Citeseer, 2012.
6. Reilly JR, *et al*. Towards a national perioperative clinical quality registry: The diagnostic accuracy of administrative data in identifying major postoperative complications. *Anaesth Intensive Care* 2020; 48(3): p. 203–212.
7. McIsaac DI, *et al*. Validation of new ICD-10-based patient safety indicators for identification of in-hospital complications in surgical patients: a study of diagnostic accuracy. *BMJ Quality & Safety* 2020; 29(3): p. 209–216.
8. McDonald SP, Russ GR. Australian registries—ANZDATA and ANZOD. *Transplant Rev* 2013; 27: 46–49.
9. Palamuthsingam D *et al*. Health data linkage research in Australia remains challenging. *Intern Med J* 2019; 49: 539–544.
10. Holman CD *et al*. A multipurpose comorbidity scoring system performed better than the Charlson index. *J Clin Epidemiol* 2005; 58:1006–14.
11. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977; 33: 363–374.
12. Burgoyne J. *Audit of Identified Hospital Acquired Cardiac Complications, in Activity Based Funding*. Darwin, Northern Territory: Department of Health, 2017.
13. Utter GH, Atolagbe OO, Cooke DT. The use of the International Classification of Diseases, Tenth Revision, Clinical Modification and Procedure Classification System in clinical and health services research: the devil is in the details. *JAMA Surg* 2019; 154: 1089–1090.