BMJ Open

Validation of administrative hospital data for identifying incident pancreatic and periampullary cancer cases: a population-based study using linked cancer registry and administrative hospital data in New South Wales, Australia

Nicola Creighton,[1] Richard Walton,[1] David Roder,[1,2] Sanchia Aranda,[3] David Currow[1]

For numbered affiliations see end of article.

Correspondence to
Nicola Creighton; Nicola. Creighton@cancerinstitute. org.au

ABSTRACT

Objectives: Informing cancer service delivery with timely and accurate data is essential to cancer control activities and health system monitoring. This study aimed to assess the validity of ascertaining incident cases and resection use for pancreatic and periampullary cancers from linked administrative hospital data, compared with data from a cancer registry (the 'gold standard').

Design, setting and participants: Analysis of linked statutory population-based cancer registry data and administrative hospital data for adults (aged ≥18 years) with a pancreatic or periampullary cancer case diagnosed during 2005–2009 or a hospital admission for these cancers between 2005 and 2013 in New South Wales, Australia.

Methods: The sensitivity and positive predictive value (PPV) of pancreatic and periampullary cancer case ascertainment from hospital admission data were calculated for the 2005–2009 period through comparison with registry data. We examined the effect of the look-back period to distinguish incident cancer cases from prevalent cancer cases from hospital admission data using 2009 and 2013 as index years.

Results: Sensitivity of case ascertainment from the hospital data was 87.5% (4322/4939), with higher sensitivity when the cancer was resected (97.9%, 715/730) and for pancreatic cancers (88.6%, 3733/4211). Sensitivity was lower in regional (83.3%) and remote (85.7%) areas, particularly in areas with interstate outflow of patients for treatment, and for cases notified to the registry by death certificate only (9.6%). The PPV for the identification of incident cases was 82.0% (4322/5272). A 2-year look-back period distinguished the majority (98%) of incident cases from prevalent cases in linked hospital data.

Conclusions: Pancreatic and periampullary cancer cases and resection use can be ascertained from linked

Strengths and limitations of this study

- This study uses statutory population-based cancer registry data as a 'gold standard' to assess the ascertainment of cancer cases from administrative hospital data.
- Sensitivity was examined by patient demographic and tumour characteristics to identify potential biases in case ascertainment from hospital data.
- A limitation is that we could only identify false positives in the hospital data with a (non-pancreatic) cancer case recorded on the cancer registry, which may lead to an underestimate of the number of false positives since false positives without any invasive cancer case recorded on the registry were not identified.
- Look-back periods of up to 5 years were examined to distinguish incident cases from prevalent cases in the hospital data.

hospital admission data with sufficient validity for informing aspects of health service delivery and system-level monitoring. Limited tumour clinical information and variation in case ascertainment across population subgroups are limitations of hospital-derived cancer incidence data when compared with population cancer registries.

INTRODUCTION

System-level monitoring of appropriateness and quality of cancer care is an essential part of cancer control.[1] Population-based cancer registries have a key role to play in system performance reporting since they generally

have a high level of completeness and accuracy of cancer case information that is obtained from multiple sources, including hospital admission and outpatient data, pathology reports and death certificates.[2] Increasingly, population-based cancer registries are expanding their collection of clinical and treatment information, or are linking to other clinical or treatment databases, to gain more comprehensive data to produce performance indicators for informing health service delivery.[3] A critical limitation is that the processes of receiving notifications and compiling data in a population-based cancer registry can be time-consuming, with lag times typically of 18 months or more before complete incidence data are available.[4] This can reduce the utility of cancer registry data for timely feedback on health system performance.

Population-level hospital admission data can be a more timely data source for obtaining incident cancer case data and evaluating treatment outcomes. However, hospital data are usually collected for general administrative purposes and may lack the accuracy and completeness required for measuring patterns of cancer care and outcomes. Validation of the data as being 'fit for purpose' is therefore necessary.[5 6] Another challenge is that the diagnosis, treatment and management of cancer can require multiple hospital admissions. As a result, there is a need to distinguish incident cases from prevalent cases. One approach is to only extract hospital admissions with both a cancer diagnosis code and procedure or treatment code for the initial diagnosis or treatment of the cancer.[7] This approach may be needed where an individual cannot be identified across multiple hospital admission records. When multiple hospital admission records for an individual can be identified, the first admission with the cancer diagnosis code recorded can be used to indicate an incident case.

The ascertainment of incident cancers from hospital admission data has good validity for some cancers, such as breast cancer, with sensitivity ranging from 77% to 86% and positive predictive values (PPVs) from 86% to 93%.[8–10] Lower validity has been found for other cancers, such as prostate and colorectal cancer, where treatment is more likely to be in an ambulatory setting or readmissions for prevalent cancers have not been distinguished from newly diagnosed cancer cases.[8 11 12] The accuracy of case ascertainment from hospital data depends on the quality of coding and completeness of the hospital data, as well as on the patterns of hospitalisation for the particular cancer.

Pancreatic cancer has poor survival and is among the fifth most common cause of cancer death in Australia, the USA and Europe.[13–15] Pancreatic cancer surgery has been the subject of system performance improvement programmes in multiple jurisdictions with the aim of improving outcomes for this cancer.[16–19] Timely data were required for the development of a service delivery programme for pancreatic surgery for cancer in New South Wales (NSW), Australia. Hospital admission data

have the potential to inform aspects of this programme, but there are few studies examining the accuracy of identifying pancreatic cancer cases from hospital admission data. One NSW study found good accuracy for the recording of pancreatic cancer diagnoses in hospital admission data,[20] but the study was performed in only one administrative health district and it is not known whether these results are generalisable. The aim of the present study was to determine the validity of administrative hospital admission data for the ascertainment of incident cases and resection use for pancreatic and periampullary cancers by measuring the sensitivity and PPV using NSW cancer registry data as the 'gold standard' and to determine the look-back period required to distinguish incident cancer cases from prevalent cancer cases in linked hospital data.

METHODS
Study population, design and data sources
The study population comprised NSW residents (aged ≥18 years) diagnosed with pancreatic or periampullary cancer (International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification (ICD-10-AM) C17.0, C24-25). Periampullary cancers were included because they have similar clinical presentations and surgery to pancreatic cancer.[21] NSW has 7.5 million residents with the majority (4.8 million) living in the greater Sydney metropolitan area. We used linked de-identified data from the NSW Central Cancer Registry (CCR), the Admitted Patient Data Collection (APDC) and the NSW Admitted Patient, Emergency Department Attendance and Deaths Register (APEDDR).

The CCR is a statutory population-based cancer registry of all incident primary invasive cancer cases (excluding non-melanoma skin cancer) and in situ breast and melanoma neoplasms diagnosed in NSW residents (referred to here as 'registry data').[22] Pathology laboratories, radiotherapy and medical oncology departments, hospitals, residential aged-care facilities and day procedure centres are required to notify the registry with clinical information about new cases of cancer and demographic details about the person. The CCR is extending data collection to include additional clinical and treatment information and is increasing automation to improve efficiency and, ultimately, timeliness. During this transformation process, however, routine data processes were delayed which affected availability of cancer incidence data, with 2009 the latest year available at the time of extraction.

The APDC (referred to here as 'hospital data') is a compilation of patient demographic, admission and discharge information, and diagnosis and procedure codes for admissions to all NSW public and private hospitals. Diseases, injuries, procedures and treatments in admitted episodes of care are coded using ICD-10-AM and the Australian Classification of Health Interventions (ACHI). The coding of diagnoses and procedures is

carried out according to the Australian Coding Standards for ICD-10-AM and ACHI. This national coding framework has been in place since the late 1990s.[23]

The APEDDR is a statutory public health and disease register maintained by the NSW Ministry of Health. It contains linked hospital (APDC) data and is the potential data source for timelier cancer incidence data in NSW, with data available up to December 2013 at the time of extraction.

Two separate probabilistic linkages (the registry to hospital data; and the hospital data in the APEDDR) were performed by the Centre for Health Record Linkage (http://www.CHeReL.org.au) with an estimated false-positive rate of 3 per 1000 and using a best practice privacy-preserving protocol that separates personal identifiers from the analysis data sets.[24] De-identified data sets were provided to the researchers by the data custodians.

Hospital coding of pancreatic and periampullary cancer—comparison to cancer registry data

We extracted admissions with a pancreatic or periampullary cancer diagnosis recorded in a primary or secondary diagnosis field between 2005 and 2009 in the linked hospital admission data of all people with a cancer case on the cancer registry between 1994 and 2009 (figure 1). We obtained from the registry month and year of cancer diagnosis, sex, age at diagnosis, remoteness of residence,[25] tumour histology type, primary site, extent of disease (the furthest extent from notifications within 4 months of diagnosis[26]) and best basis of diagnosis (the highest level of verification from notifications within 4 months of diagnosis) for pancreatic and periampullary cancer cases diagnosed between 2005 and 2009. People who underwent pancreatectomy (ACHI block 978) were identified from the linked hospital data. Resections in the one calendar month prior to diagnosis (to allow for minor inaccuracies in the month of diagnosis) or any time after diagnosis were included.

For the hospital data, we allocated a primary cancer site using diagnosis codes recorded in admissions for pancreatic and periampullary cancer. Inconsistencies in the recording of primary site codes across different admissions for the same person required the development of an algorithm to allocate site. For pancreas primary sites (C25), the most specific site in the pancreas was recorded. For example, if the site was 'C25.9 pancreas, not otherwise specified' in one admission and 'C25.0 head of pancreas' in another, then the latter more specific site was allocated. When a periampullary cancer (C17.0, C24) was recorded in one admission and a pancreatic primary site in another, the primary site was allocated as periampullary.

Ascertaining incident cases from linked hospital data

All admissions from 1 July 2000 (the earliest available data) to 31 December 2013 with a pancreatic or periampullary cancer diagnosis recorded in a primary or secondary diagnosis field were extracted from the linked hospital data in the APEDDR. Non-NSW residents and people aged <18 years at their first cancer admission were excluded.

Analysis
Linked cancer registry and hospital data

Sensitivity of case ascertainment from the hospital data was calculated as the proportion of cases on the registry with an admission with a pancreatic or periampullary cancer diagnosis recorded in the hospital data (true positives) for the 2005–2009 period. Sensitivity was calculated by patient demographics (age at diagnosis (grouped by 10-year age groups) and remoteness of residence (grouped as major city, inner regional, outer regional, remote and very remote)) and tumour characteristics (primary site, histology type, extent of disease, best basis of diagnosis, resection status) to assess if sensitivity varied by these characteristics. A PPV was estimated as the proportion of people with a pancreatic or periampullary cancer diagnosis recorded in the hospital data who had an incident case recorded on the registry for the 2005–2009 period.

We examined the diagnoses recorded for false-negative and false-positive cases in the hospital and

Figure 1 Flow chart of case ascertainment from the linked registry and hospital data.
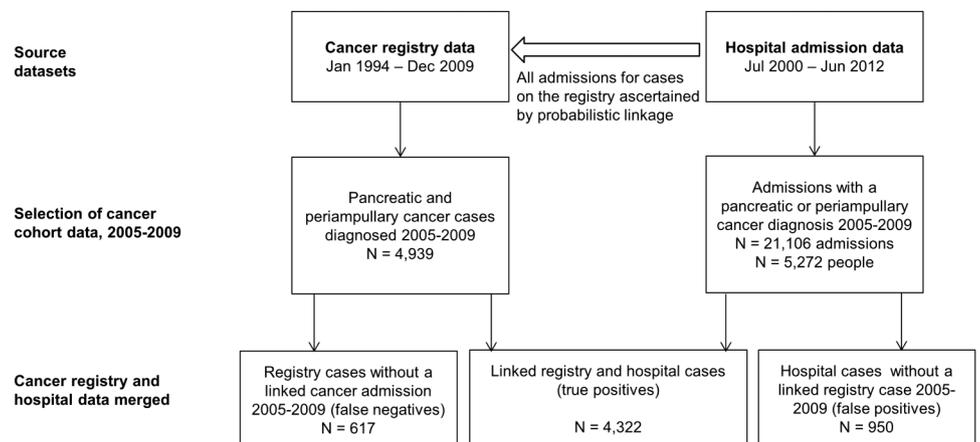
registry data, respectively, to evaluate misclassification of cancer cases. For false negatives, we examined diagnoses in hospital admissions from the month prior to the registry diagnosis date or any time after in the 2005–2009 period. For false positives, we ascertained cases on the registry diagnosed between 2005 and 2009, except for pancreatic and periampullary cancers for which we ascertained cases on the registry diagnosed between 1994 and 2004. Exact binomial CIs were calculated for sensitivity and PPV (SAS/STAT V.12.1, SAS Institute, Cary, North Carolina, USA).

Linked hospital data
We examined the effect of a look-back period to distinguish incident cases from prevalent cases ascertained from the hospital data in the APEDDR by calculating the number of people with a 'first' admission for pancreatic or periampullary cancer with look-back periods increasing by increments of 6 months up to a period of 5 years using 2009 and 2013 as index years.

RESULTS
Accuracy of hospital diagnoses
Overall, the sensitivity of case ascertainment of registry recorded cases of pancreatic and periampullary cancer cases for the 2005–2009 period from the hospital data was 87.5% (4322/4939) and was highest (97.9%, 715/730) for people who underwent resection (table 1). The lowest sensitivity (9.6%, 18/187) was for cases notified to the registry by death certificate only. Most cases (71.1%, 133/187) with death certificate-only notifications were among people 80 years or older, and this age group had lower sensitivity (83.2%, 1271/1528) for hospital ascertainment than younger age groups. Lower sensitivity was observed for people residing outside of major cities, with inner and outer regional areas having the lowest sensitivity (83.3%). Further exploration of the geographic variation in case ascertainment found that sensitivity ranged from 55.0% to 92.6% across regional and remote administrative health districts (not shown).

Sensitivity was highest for localised (95.9%, 805/839) cancers, followed by cancers with regional (93.3%, 846/907) and distant (89.5%, 1863/2082) extent of disease and was lowest for cases where the extent was recorded as unknown on the registry (72.7%, 808/1111). Across primary tumour sites, sensitivity was higher for pancreatic (88.6%, 3733/4211) and duodenal cancers (87.1%, 162/186) compared with extrahepatic bile duct and ampullary cancers (78.8%, 427/542), particularly those of the cholangiocarcinoma histology type (67.9%, 131/193).

Of the cases on the registry without an admission with pancreatic or periampullary cancer recorded (false negatives, n=617), 42.9% (n=265) of people did not have a hospital admission in the month prior to diagnosis or any time after in the 2005–2009 period (table 2). The number of false negatives was affected by only including admissions up to the end of 2009, with the

sensitivity of case ascertainment for cases diagnosed in 2009 lower (85.6%, 860/1005) than the preceding years. People diagnosed in 2009 who had their first admission for cancer in 2010 would be classified as false negatives. Of false-negative cases with an admission (n=352), diagnoses recorded in hospital admissions included cancer of ill-defined or unspecified site (22.7%, 80/352), intrahepatic bile duct carcinoma (13.9%, 49/352) and non-cancer diagnoses (table 2).

A total of 5272 people had an admission with pancreatic or periampullary cancer recorded in the linked hospital to registry data in the 2005–2009 period, giving a PPV of 82.0% (4322/5272; 95% CL 80.9% to 83.0%). Of the false positives (n=950), 50.5% (n=480) had a pancreatic or periampullary cancer case recorded on the registry diagnosed prior to 2005 (table 2). For these cases, the pancreatic or periampullary cancer diagnoses recorded on the hospital admission was correct; however, the cases were prevalent cases rather than incident cases in the 2005–2009 period. Of the other false positives, one-fifth (22.8%, 107/470) of people had an intrahepatic bile duct carcinoma and another fifth (21.1%, 99/470) had an ill-defined or unspecified primary site cancer case recorded on the registry. Nine false positives (1.9%, 9/470) had a pancreatectomy recorded. These estimates of false positives in the hospital data only include people with a linked invasive cancer case in the registry data; therefore, the number of false positives may be underestimated.

We compared the classification of the primary site allocated from the hospital data with the site recorded on the registry (table 3). One per cent (41/3945) of cases allocated as a pancreatic cancer using hospital diagnoses were true periampullary cancer cases, whereas 12.5% (106/847) of cases allocated as periampullary using hospital diagnoses were true pancreatic cancer cases. The misclassification of periampullary cancers was greater since any person with an admission with a periampullary cancer diagnosis recorded was allocated as periampullary regardless of if they also had a pancreatic cancer diagnosis recorded in another admission.

Ascertaining incident cases from hospital data
Around one-quarter of cases ascertained from the linked hospital data (APEDDR) were misclassified as an incident case without any look-back period (figure 2). Two years of look-back was sufficient to distinguish the majority (98%) of incident cases from prevalent cases, which was similar for pancreatic and periampullary cancers or pancreatic cancers only. For example, 1.8% (21/1193) of people classified as having an 'incident' pancreatic or periampullary cancer case in 2013, using a 2-year look-back period, were identified as prevalent cases when a 5-year look-back period was used. The number of cases ascertained from the APEDDR for the 2005–2009 period, using all available data to July 2000 as a look-back period, was 4970, which is within 1% of the number of cases recorded on the registry for this period (n=4939).

Table 1 Sensitivity of the ascertainment of incident cancer cases on the registry from hospital data by demographic and tumour characteristics, 2005–2009

Variable	Registry cases	Ascertained in the hospital data*	Sensitivity (%)	95% CL
Year of diagnosis				
2005	973	850	87.4	(85.1 to 89.4)
2006	926	827	89.3	(87.1 to 91.2)
2007	1034	907	87.7	(85.6 to 89.7)
2008	1001	878	87.7	(85.5 to 89.7)
2009	1005	860	85.6	(83.2 to 87.7)
Age at diagnosis (years)				
<50	246	221	89.8	(85.4 to 93.3)
50 to 59	591	519	87.8	(84.9 to 90.3)
60 to 69	1087	973	89.5	(87.5 to 91.3)
70 to 79	1487	1338	90.0	(88.3 to 91.5)
80+	1528	1271	83.2	(81.2 to 85.0)
Remoteness of residence				
Major city	3447	3078	89.3	(88.2 to 90.3)
Inner regional	1091	909	83.3	(81.0 to 85.5)
Outer regional	366	305	83.3	(79.1 to 87.0)
Remote and very remote	35	30	85.7	(69.7 to 95.2)
Primary site				
Pancreas (C25)	4211	3733	88.6	(87.7 to 89.6)
Extrahepatic bile duct and ampulla (C24)	542	427	78.8	(75.1 to 82.2)
Duodenum (C17.0)	186	162	87.1	(81.4 to 91.6)
Histology type				
Adenocarcinoma	2629	2388	90.8	(89.7 to 91.9)
Cholangiocarcinoma	193	131	67.9	(60.8 to 74.4)
Neuroendocrine	144	118	81.9	(74.7 to 87.9)
Other	45	40	88.9	(75.9 to 96.3)
Unspecified	1928	1645	85.3	(83.7 to 86.9)
Extent of disease				
Localised	839	805	95.9	(94.4 to 97.2)
Regional	907	846	93.3	(91.4 to 94.8)
Distant	2082	1863	89.5	(88.1 to 90.8)
Unknown	1111	808	72.7	(70.0 to 75.3)
Best basis of diagnosis†				
Histopathology	2761	2510	90.9	(89.8 to 92.0)
Cytology	481	407	84.6	(81.1 to 87.7)
Clinical	1507	1387	92.0	(90.6 to 93.4)
Death certificate only	187	18	9.6	(5.8 to 14.8)
Resection status				
No resection	4209	3607	85.7	(84.6 to 86.7)
Resection	730	715	97.9	(96.6 to 98.8)
Overall	4939	4322	87.5	(86.6 to 88.4)

*Registry cases with a diagnosis of pancreatic or periampullary cancer recorded in an admission. We did not measure the concordance of variables between the two data sources.
†Cases (n=3) notified by postmortem only are not shown. Cases notified to the New South Wales Central Cancer Registry (NSW CCR) by death certificate only were unavailable for 2009.[22] Death certificate-only notified cases were 4.8% (187/3934) of cases for the 2005–2008 period.

DISCUSSION

We found sensitivity for the ascertainment of pancreatic and periampullary cancer cases from the hospital data of 87.5% for the 2005–2009 period. The accuracy of hospital coding varied by tumour primary site and histology, with higher sensitivity of case ascertainment for pancreatic (88.6%) and duodenal cancers (87.1%) compared with extrahepatic bile duct and ampullary cancers (78.8%) and with lower sensitivity for extrahepatic cholangiocarcinomas (67.9%). Misclassification of pancreatic and periampullary cancers in the hospital data was often to closely related sites, for example, intrahepatic bile duct carcinoma, or to less specific sites such as cancers of ill-defined or unspecified primary sites. Whereas hospital coders might only have information from a particular admission available, coders at a cancer

Table 2 Diagnoses (International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification, ICD-10-AM) recorded for false negatives and false positives, 2005–2009

		N (%)
	*Hospital diagnosis**	
False negatives (N=617)	Intrahepatic bile duct carcinoma (cholangiocarcinoma) (C22.1)	49 (7.9)
	Gallbladder cancer (C23)	10 (1.6)
	Cancer of ill-defined or unspecified site (C26, C76, C80)	80 (13.0)
	Secondary cancer of the liver, other or unspecified digestive organ (C78.7, C78.8)	23 (3.7)
	Neoplasms with in situ, benign, uncertain or unknown behaviour of the pancreas, small intestine, biliary tract, ill-defined or unspecified digestive organs†	29 (4.7)
	Calculus of the bile duct and other biliary tract diseases (K80.3-5, K83)	54 (8.8)
	Acute pancreatitis and other diseases of the pancreas (K85, K86)	11 (1.8)
	Other diagnosis	96 (15.6)
	No hospital admission	265 (42.9)
	Registry diagnosis	
False positives (N=950)	Pancreatic or periampullary cancer diagnosed prior to 2005	480 (50.5)
	Intrahepatic bile duct carcinoma (cholangiocarcinoma) (C22.1)	107 (11.3)
	Gallbladder cancer (C23)	47 (4.9)
	Cancer of ill-defined or unspecified site (C26, C76, C80)	99 (10.4)
	Other cancer	217 (22.8)

*People had more than one diagnosis category recorded in admissions. The category was assigned hierarchically to avoid double counting people.
†D01.5, D01.7, D01.9, D13.2-9, D37.2, D37.6, D37.7, D37.9.

Table 3 Primary site (International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification, ICD-10-AM) of registry and hospital cases, 2005–2009

	Hospital primary site			
	False negatives*	Periampullary cancer	Pancreatic cancer	Total
Registry primary site				
False positives†	0	193	277	470
Periampullary cancer	112	548	41	701
Pancreatic cancer	240	106	3627	3973
Total	352	847	3945	5144

*People who did not link to a hospital admission are excluded from the false negatives (n=265).
†People who linked to a pancreatic or periampullary cancer diagnosed prior to 2005 are excluded from the false positives (n=480).

registry often have multiple sources of information from diagnostic procedures and treatment which can enable more accurate coding of tumour characteristics.

Several aspects of the coverage of the hospital data affected the ascertainment of cancer cases. In particular, only admissions in NSW hospitals were available. Variation in sensitivity across regional and remote administrative health areas most likely reflects patterns of interstate patient outflows since cases will not be ascertained from NSW hospital data for people who have all of their inpatient treatment outside of NSW. The population-level coverage of the hospital data must be assessed with geographic areas with insufficient coverage unsuitable for cancer case ascertainment from hospital data. Conversely, the NSW cancer registry achieves complete population coverage since NSW residents treated in other jurisdictions are notified to the NSW cancer registry by the statutory population-based cancer registry in that jurisdiction, with Australia having full population coverage by statutory cancer registries. Another factor affecting case

ascertainment is the inclusion in the registry of cases notified by death certificate only, where the sensitivity of ascertainment from hospital data was low (9.6%). The accuracy of diagnostic information may be lower for cases when a death certificate is the only source of information, but nevertheless they are recorded by population-based cancer registries to capture comprehensive incidence data for a population.[26]

Pancreatic and periampullary cancers are good candidates for ascertainment from hospital admission data since diagnosis, treatment or symptom management is likely to require hospitalisation during the course of the disease. Sensitivity was highest (97.9%) for the minority of people (15%) who underwent curative resection. Hospitalisation to relieve biliary or gastric obstruction by stenting or bypass surgery or to manage pain and nutrition is commonly required in the management of these cancers.[27 28] Chemotherapy, which is indicated for people with unresectable disease,[28] is mostly delivered in an outpatient setting in NSW and therefore is largely not
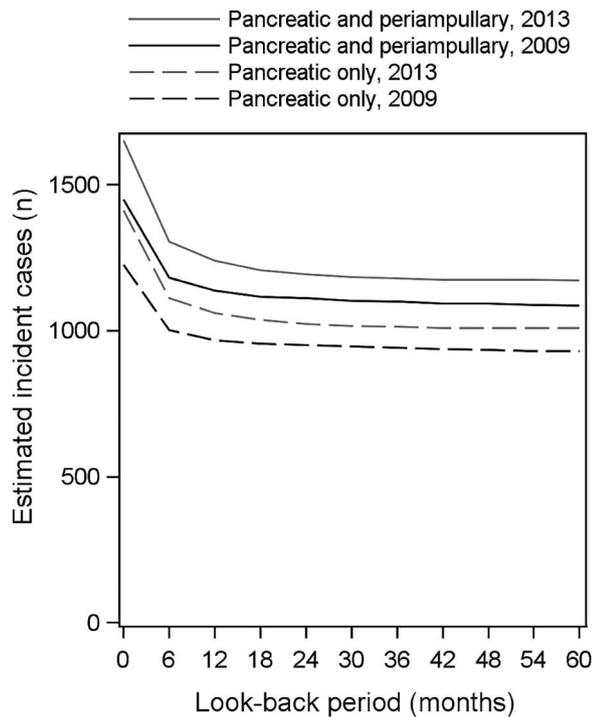
Figure 2 Effect of a look-back period on the estimates of incident cases from hospital admission data, 2009 and 2013 index years.

captured in hospital admission data. This may lead to underascertainment of cases from hospital admission data when chemotherapy is the only therapy required for management of the cancer. Multiple hospitalisations for the management of these cancers mean that prevalent cases need to be distinguished from incident cases in the hospital data. The PPV estimated in this study (82.0%) was impacted by this pattern of multiple hospitalisations. On examination, half of the false positives in the hospital data were true pancreatic or periampullary cancer cases but were people with a prevalent rather than an incident case for the 2005–2009 period. The PPV could be improved by the use of a minimum 2-year look-back period to identify a person's first admission for cancer, which we found was sufficient to distinguish the majority of incident cases from prevalent cases.

We were able to calculate a PPV; however, it may be an underestimate since we could only identify people with a hospital admission for pancreatic or periampullary cancer who had a cancer case recorded on the NSW cancer registry between 1994 and 2009. Therefore, false positives in the hospital data without any cancer registry-recorded case were not ascertained, which is a limitation of our study. We do not expect the underestimation of false positives to be substantial, however, since the number of incident cases ascertained from the APEDDR was similar to the number of cases recorded on the registry for the 2005–2009 period.

Few studies have examined the validity of ascertaining pancreatic cancer cases from hospital data, with none

examining periampullary cancers. Our study compares favourably to a study in the USA which reported sensitivity for pancreas cancer of up to 86% using inpatient Medicare claims data.[29] A NSW study reported sensitivity for pancreas cancer of 94.7% and a PPV of 80.9% for admission data from public hospitals in one administrative health district.[20] A strength of our study was that it used population-based registry data as the 'gold standard' and included both public and private hospital data. Without coverage of public and private hospitals, we expect there would have been substantial underascertainment of cases, as was found by another study.[11] Our study is relevant to jurisdictions with administrative hospital data with population coverage and standardised coding in place. Incident breast, colorectal and lung cancer case data obtained from administrative hospital records have been found to be of sufficient quality for informing health services research in other jurisdictions.[8 9 12]

Internationally, pancreatic cancer has been the subject of health system performance improvement programmes[16–19] since studies have identified underuse of curative resection[30 31] and variation across hospitals in morbidity, mortality and survival outcomes following surgery, particularly in relation to hospital volume.[32] Programmes have generally established recommended minimum hospital volumes with the aim of increasing access to expert multidisciplinary care and improving patient outcomes. Measuring if these minimum volumes are met, changes to the per cent of people receiving curative surgery and changes in outcomes are key components of monitoring the implementation of these programmes.[17–19] Our study demonstrates that NSW hospital data are of sufficient quality to inform aspects of the development and monitoring of a service improvement programme for pancreatic surgery for cancer. For example, the data are adequate for measurement of hospital volume of pancreatic cancer surgery, overall postsurgical outcomes (when linked to death registry data) and resection rates for population groups with good coverage in the hospital data. Hospital admission data, however, lack a date of cancer diagnosis and detailed clinical information, such as tumour size and vascular involvement, which are required to measure postdiagnosis survival and perform risk adjustment of outcomes with minimum residual confounding for case complexity.

CONCLUSION

Pancreatic and periampullary cancer cases can be ascertained from hospital admission data where coding standards are applied and there is population coverage. The pattern of hospitalisation for these cancers mean that linked hospital data, in which multiple admissions for the same person can be identified, and a sufficient look-back period are required to distinguish incident cases from prevalent cases. Our study indicates that hospital-derived case and resection data have sufficient validity to inform

aspects of health system performance planning and monitoring for pancreatic cancer. However, case ascertainment differs across population subgroups and the clinical variables are limited. Cancer registry data with population-level coverage and clinical information are required for some health system performance measures.

Author affiliations
[1]Cancer Institute NSW, Sydney, New South Wales, Australia
[2]Centre for Population Health Research, University of South Australia, Adelaide, South Australia, Australia
[3]Cancer Council Australia, Sydney, New South Wales, Australia

REFERENCES

1. Veillard J, Huynh T, Ardal S, *et al*. Making health system performance measurement useful to policy makers: aligning strategies, measurement and local health system accountability in Ontario. *Healthc Policy* 2010;5:49–65.
2. Parkin DM. The evolution of the population-based cancer registry. *Nat Rev Cancer* 2006;6:603–12.
3. Hiatt RA, Tai CG, Blayney DW, *et al*. Leveraging state cancer registries to measure and improve the quality of cancer care: a potential strategy for California and beyond. *J Natl Cancer Inst* 2015;107:pii:djv047.
4. Zanetti R, Schmidtmann I, Sacchetto L, *et al*. Completeness and timeliness: cancer registries could/should improve their performance. *Eur J Cancer* 2015;51:1091–8.
5. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol* 1996;25:435–42.
6. Australian Institute of Health and Welfare. *An AIHW framework for assessing data sources for population health monitoring: working paper. Cat no PHE 180.* Canberra: AIHW, 2014.
7. Couris C, Schott A-M, Ecochard R, *et al*. A literature review to assess the use of claims databases in identifying incident cancer cases. *Health Serv Outcomes Res Methodol* 2003;4:49–63.
8. Baldi I, Vicari P, Di Cuonzo D, *et al*. A high positive predictive value algorithm using hospital administrative data identified incident cancer cases. *J Clin Epidemiol* 2008;61:373–9.
9. Yuen E, Louis D, Cisbani L, *et al*. Using administrative data to identify and stage breast cancer cases: implications for assessing quality of care. *Tumori* 2011;97:428–35.
10. Kemp A, Preen DB, Saunders C, *et al*. Ascertaining invasive breast cancer cases; the validity of administrative and self-reported data sources in Australia. *BMC Med Res Methodol* 2013;13:17.
11. Bernal-Delgado EE, Martos C, Martínez N, *et al*. Is hospital discharge administrative data an appropriate source of information for cancer registries purposes? Some insights from four Spanish registries. *BMC Health Serv Res* 2010;10:9.
12. Quantin C, Benzenine E, Hagi M, *et al*. Estimation of national colorectal-cancer incidence using claims databases. *J Cancer Epidemiol* 2012;2012:1–7.
13. Australian Institute of Health and Welfare. *Cancer in Australia: an overview, 2014. Cancer series no 90 Cat no CAN 88.* Canberra: AIHW, 2014.
14. U.S. Cancer Statistics Working Group. United States cancer statistics: 1999–2012 incidence and mortality web-based report. *Secondary United States cancer statistics: 1999–2012 incidence and mortality web-based report* 2015. https://nccd.cdc.gov/uscs/toptencancers.aspx.
15. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, *et al*. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013;49:1374–403.
16. Simunovic M, Urbach D, Major D, *et al*. Assessing the volume-outcome hypothesis and region-level quality improvement interventions: pancreas cancer surgery in two Canadian provinces. *Ann Surg Oncol* 2010;17:2537–44.
17. Massarweh NN, Flum DR, Symons RG, *et al*. A critical evaluation of the impact of leapfrog's evidence-based hospital referral. *J Am Coll Surg* 2011;212:150–9.e1.
18. Gooiker GA, Lemmens VE, Besselink MG, *et al*. Impact of centralization of pancreatic cancer surgery on resection rates and survival. *Br J Surg* 2014;101:1000–5.
19. de Cruppé W, Malik M, Geraedts M. Minimum volume standards in German hospitals: do they get along with procedure centralization? A retrospective longitudinal data analysis. *BMC Health Serv Res* 2015;15:279.
20. Stavrou E, Pesa N, Pearson SA. Hospital discharge diagnostic and procedure codes for upper gastro-intestinal cancer: how accurate are they? *BMC Health Serv Res* 2012;12:331.
21. Pancreatic Section of the British Society of Gastroenterology, Pancreatic Society of Great Britain and Ireland, Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland, *et al*. Guidelines for the management of patients with pancreatic cancer, periampullary and ampullary carcinomas. *Gut* 2005;54(Suppl 5):v1–16.
22. Currow D, Thomson W. Cancer in New South Wales: incidence report 2009. Sydney: Cancer Institute NSW, 2014.
23. Roberts RF, Innes KC, Walker SM. Introducing ICD-10-AM in Australian hospitals. *Med J Aust* 1998;169(Suppl):S32–5.
24. Kelman CW, Bass AJ, Holman CD. Research use of linked health data—a best practice protocol. *Aust N Z J Public Health* 2002;26:251–5.
25. Australian Bureau of Statistics. 1259.0.30.004 - Australian standard geographical classification (ASGC) remoteness structure (RA) digital boundaries, Australia, 2006. Canberra: ABS, 2007. http://www.abs.gov.au/ausstats%5Cabs@.nsf/0/9A784FB979765947CA25738C0012C5BA?Opendocument
26. Esteban D, Whelan S, Laudico A, Parkin DM, eds. *Manual for cancer registry personnel. IARC Technical Report no 10.* Lyon: IARC. 1995. http://www.iarc.fr/en/publications/pdfs-online/treport-pub/treport-pub10/index.php
27. House MG, Choti MA. Palliative therapy for pancreatic/biliary cancer. *Surg Clin North Am* 2005;85:359–71.
28. NCCN. NCCN clinical practice guidelines in oncology: Pancreatic adenocarcinoma v.2.2014. National Comprehensive Cancer Network. https://www.nccn.org/professionals/physician_gls/pdf/pancreatic.pdf
29. Cooper GS, Yuan Z, Stange KC, *et al*. The sensitivity of Medicare claims data for case ascertainment of six common cancers. *Med Care* 1999;37:436–44.
30. Bilimoria KY, Bentrem DJ, Ko CY, *et al*. National failure to operate on early stage pancreatic cancer. *Ann Surg* 2007;246:173–80.
31. Sharp L, Carsin AE, Cronin-Fenton DP, *et al*. Is there under-treatment of pancreatic cancer? Evidence from a population-based study in Ireland. *Eur J Cancer* 2009;45:1450–9.
32. Gooiker GA, van Gijn W, Wouters MWJM, *et al*. Systematic review and meta-analysis of the volume–outcome relationship in pancreatic surgery. *Br J Surg* 2011;98:485–94.