

BMJ Open Cancer-related hospitalisations and 'unknown' stage prostate cancer: a population-based record linkage study

Qingwei Luo,^{1,2} Xue Qin Yu,^{1,2} David Paul Smith,^{1,2,3} David Eamon Goldsbury,¹ Claire Cooke-Yarborough,⁴ Manish Indravadan Patel,^{5,6} Dianne Lesley O'Connell^{1,2,7}

To cite: Luo Q, Yu XQ, Smith DP, *et al.* Cancer-related hospitalisations and 'unknown' stage prostate cancer: a population-based record linkage study. *BMJ Open* 2017;**7**:e014259. doi:10.1136/bmjopen-2016-014259

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-014259>).

Received 12 September 2016
Revised 21 November 2016
Accepted 14 December 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Qingwei Luo;
qingwei@nswcc.org.au

ABSTRACT

Objectives: To identify reasons for prostate cancer stage being recorded as 'unknown' in Australia's largest population-based cancer registry.

Design: Prospective population-based cohort.

Setting: New South Wales (NSW) is the most populous state in Australia, with almost one third of the total national population.

Participants: NSW Cancer Registry (NSWCR) records for prostate cancer cases diagnosed in 2001–2009 were linked to the NSW Admitted Patient Data Collection (APDC) for 2000–2010. All patients in this study had a minimum of 12 months follow-up in the hospital episode records after their date of diagnosis as recorded by the NSWCR.

Main outcome measures: Incidence of 'unknown' stage prostate cancer and cancer-specific survival.

Results: Of 50 597 prostate cancer cases, 39.9% were recorded as having 'unknown' stage. Up to 4 months after diagnosis, 77.2% of cases without a hospital-reported cancer diagnosis were recorded as having 'unknown' stage. Among those patients with a hospital-reported cancer diagnosis, stage was 'unknown' for 7.6% of cases who received a radical prostatectomy (RP) and for 34.0% of cases who had procedures other than RP. In the latter group, the factors that were related to having 'unknown' stage were living in disadvantaged areas (adjusted OR (aOR) range: 1.13 to 1.20), attending a private hospital (aOR range: 1.25 to 2.13), having day-only admission for care (aOR=1.23, 95% CI 1.11 to 1.36), or having procedures other than multiple procedures with imaging (eg, biopsy only, aOR range: 1.11 to 1.45).

Conclusions: Over half of 'unknown' stage prostate cancer cases did not have a hospital-reported prostate cancer diagnosis within the 4 months after initial diagnosis. We identified differences in the likelihood of cases being recorded as 'unknown' stage based on socioeconomic status and facility type, which suggests that further investigation of reporting practices in relation to diagnostic and treatment pathways is required.

INTRODUCTION

Data on cancer stage at diagnosis as recorded in population-based cancer registries is an important factor in population-wide cancer

Strengths and limitations of this study

- The first systematic examination of hospital inpatient cancer services and 'unknown' stage at diagnosis using population-based linked records from routinely collected administrative data sets.
- There are numerous advantages to using administrative data, and record linkage between data sets can add further value to these resources, with their population coverage ensuring that the results are representative.
- The Admitted Patient Data Collection captured 90% or more of all radical prostatectomies, only systematically missing a small number of patients treated interstate.
- Our study was restricted to patients with prostate cancer recorded in the NSW Cancer Registry, so findings might not be generalisable to other population-based cancer registries, nor will they be representative of other cancer types.

monitoring. It is a powerful predictor of survival¹ and is important for estimating health service demands,² and for evaluating the effectiveness of cancer screening programmes and other early detection initiatives.³ For prostate cancer, registry-recorded stage information is particularly important in the evaluation of the potential effects of over-diagnosis due to prostate-specific antigen (PSA) testing and, conversely, the benefits associated with a reduction in cases presenting with later stage disease.⁴ Unfortunately, however, a high proportion of 'unknown' stage cases significantly reduces the ability to statistically control for the effect of disease stage on patient outcomes,⁵ limits interpretation of the appropriateness of prostate cancer care based on disease stage and can cause potential bias if cases with 'unknown' stage are excluded from analyses.⁴

The New South Wales Cancer Registry (NSWCR) is the only Australian population-based cancer registry that has routinely collected summary stage at diagnosis, doing so since its inception in 1972. The proportion

of prostate cancer cases with 'unknown' stage in the NSWCR (41% in 1999–2007) has been reported to be much higher than in similar registries in the USA, Germany, Switzerland and Norway.⁴ Investigations focusing on why 'unknown' stage is recorded in the cancer registry will provide a better understanding of the direction of bias in epidemiological studies that use these stage data and may also inform strategies for data quality assurance.

There are a number of possible reasons why a prostate cancer case may be registered as being of 'unknown' stage, including patients not being healthy enough or deciding not to undergo the medical workup required to determine stage,^{3 4 6–8} that there is a determination that staging is not necessary for decisions about treatment to be made,⁸ that economic or social barriers, or a lack of access to comprehensive health services that have the capacity to complete all necessary staging investigations means cases are not fully staged,^{3 6 8} or that stage is known to the treating clinician, but is not recorded in the hospital medical records.⁹ It is also likely that restrictions in defining a valid data source for cancer registry staging can affect the completeness of stage data.^{4 6} There is currently no published research that has systematically examined the reasons why 'unknown' stage at diagnosis is recorded by Australian registries.

By linking administrative data sets for a cohort of prostate cancer cases recorded in the NSWCR, the aim of this descriptive study of the patterns of inpatient hospital cancer services was to identify possible reasons for 'unknown' stage being recorded.

METHODS

Data sources

Cancer registry data

Data for all men with a first diagnosis of prostate cancer in New South Wales, Australia in 2001–2009 (the most recent year for which cancer registry data were available) were identified and extracted from the NSWCR using the International Classification of Diseases for Oncology 3rd Edition (ICD-O-3) topography code of C61.¹⁰ It is a mandatory requirement that all cancers diagnosed in NSW, except for non-melanoma skin cancers, are notified to the NSWCR by institutions, including public and private hospitals, public multipurpose health services, radiation oncology departments, cancer care centres, private day procedure centres, residential aged care facilities and pathology laboratories.¹¹ However, mandatory notifiers do not include the private consulting rooms of individual general practitioners or specialists (including urologists). Month and year of diagnosis were available for analysis. After excluding 430 cases who were notified postmortem, or through death certificate only, 398 cases who were diagnosed and died in the same month, and four cases who were aged 100 years or older at diagnosis,¹² a total of 50 597 prostate cancer cases remained for analysis (see online supplementary resource 1).

Hospital data

The NSW Admitted Patient Data Collection (APDC) contains information on all inpatient separations (discharges, transfers and deaths) from all public, private and repatriation hospitals, and private day facilities in NSW. This information is recorded as 'episodes of care' and includes disease diagnosis codes and procedure codes.¹³ In this study, the APDC data from 2000 to 2010 were linked to the NSWCR records to identify prostate cancer-specific procedures and prostate cancer diagnoses within each episode of care.

Record linkage

Record linkage between NSWCR and APDC records was undertaken by the Centre for Health Record Linkage (CHeReL) using probabilistic linkage and best practice privacy-preserving protocols. Each person in the NSWCR and APDC was assigned a unique project person number to allow matching of individuals across the two data sets. All uncertain matches were reviewed clerically, together with a sample of 'certain' matches and non-matches, with an estimated 0.4% false-positive and <0.5% false-negative linkages.¹⁴ All patients in this study had a minimum of 12 months follow-up in the hospital episode records after their date of diagnosis as recorded by the NSWCR.

Outcome variables

The outcome of interest was stage of disease at diagnosis (referred to as degree of spread at diagnosis in the NSWCR¹⁵), reflecting the highest degree of spread reported within 4 months of diagnosis as ascertained from mandatory notifications from private and public hospitals, pathology laboratories, and inpatient and outpatient treatment facilities.¹¹ Stage recorded in the NSWCR uses a modified classification by the International Agency for Research on Cancer (IARC)¹⁶ similar to that used by SEER,¹⁷ with the stage categories of localised (cancer contained entirely in the prostate gland), regional (cancer extended into tissues surrounding the prostate or to regional lymph nodes), distant (cancer extended beyond regional lymph nodes, to bones or to other distant sites) and 'unknown' (where information in the notifications was insufficient for the cancer registry to assign stage). For some analyses, stage at diagnosis was further grouped into a dichotomous variable, indicating 'unknown' or known stage.

We also investigated survival outcomes by stage at diagnosis and the cancer-related procedures received. Survival status and cause of death were obtained from the NSWCR. People with cancer were matched against death records from the State Registry of Births, Deaths, and Marriages and the National Death Index. All eligible cases were followed up to the end of 2008, the most recent year for which death data were available. The survival time was calculated from the date of prostate cancer diagnosis to the date of death from prostate cancer. Those who did not die from prostate cancer

were censored at the date of death from other causes or at 31 December 2008 if they were still alive. As survival status was not available after 2008, cases diagnosed in 2009 were excluded from the survival analysis.

Hospital health service characteristics

Hospital-reported prostate cancer diagnosis

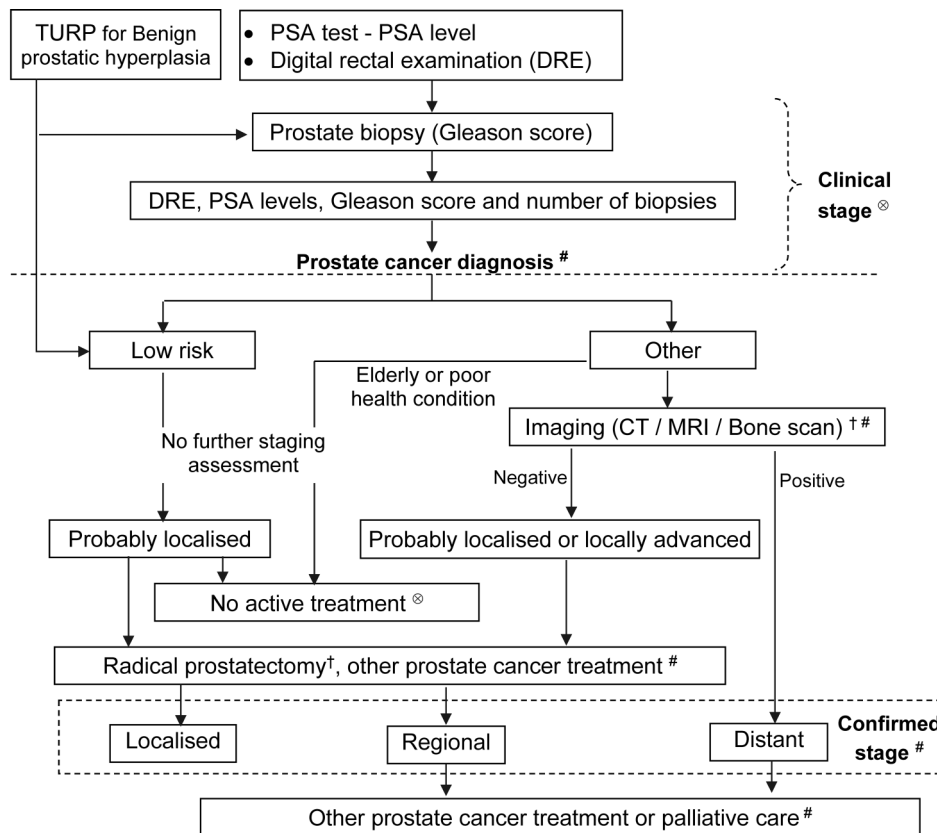
According to the NSW Health Policy Directive, the NSWCR must be notified if a patient presents for a consultation or treatment at any notifying facility in NSW and has a diagnosis of cancer, where the cancer is the principal or an additional diagnosis for the prostate cancer-related episode of care.¹¹ Those cases with a hospital-reported prostate cancer diagnosis, regardless of the care received, should have generated an electronic or paper hospital notification, in a uniform format with specific fields for degree of cancer spread. Therefore, in this study, hospital-reported prostate cancer diagnosis in the APDC was used as a flag that a hospital notification should have been sent to the NSWCR. Hospital notification is considered to be a significant source of stage information, as it should contain this information collected from any of the several different prostate cancer procedures, as well as any clinical stage information collected from urologists' referrals. Up to 55 diagnosis codes using the International Classification of Diseases 10th revision Australian Modification (ICD-10-AM)¹⁸

recorded in each episode of care were scanned for a diagnosis of prostate cancer (C61). In this study, the median number of diagnosis codes recorded at each episode of care was 5 (range 0–44).

Hospital prostate cancer procedures

As a single patient could have multiple episodes of care and multiple procedures recorded in the APDC, we developed a hierarchical classification system (see online supplementary resource 2) to identify the key hospital cancer care activity for prostate cancer treatment experienced. This hierarchical classification system is based on the likelihood that the procedure will provide stage information that could contribute to staging in the cancer registry,¹¹ as well as taking into account general clinical practice for prostate cancer diagnosis and initial treatment, and acknowledging that not all stage information is notifiable to the NSWCR (figure 1). All relevant procedure codes in the Medicare Benefits Schedule-Extended classification of the ICD-10-AM¹⁸ for prostate cancer-related treatment and procedures were identified and categorised into groups: radical prostatectomy (RP), imaging (includes bone scans, abdominal/pelvic MRI and CT), other prostate cancer treatment (includes external beam radiotherapy, brachytherapy and bilateral orchiectomy), other prostate surgery (includes other prostatectomy, cryoablation of the prostate and

Figure 1 Diagnostic and initial treatment pathways for prostate cancer in New South Wales, Australia.



Availability of stage information to the NSW Cancer Registry:

† Must notify

Notifiable depending on the facilities attended

⊗ Not notifiable

endoscopic destruction of a lesion of the prostate).¹⁹ Transurethral resection of the prostate (TURP) postdiagnosis, TURP for benign prostatic hyperplasia before prostate cancer diagnosis and prostatic biopsy do not provide any definitive stage information (see online supplementary resource 2). Up to 51 procedure codes could be recorded for each episode of care. The median number of procedure codes recorded at each episode of care was 5 (range 0–50).

For cases with multiple procedures at the same level in the hierarchy, the hospital type and day stay status of the earliest episode of care was selected for analysis. The time interval in months from the month of prostate cancer diagnosis recorded in the NSWCR to the month of the selected procedure or hospital-reported prostate cancer diagnosis in the APDC was used as a proxy for time to cancer notification to the NSWCR. For some analyses, the procedures were further grouped into mutually exclusive categories (RP, procedures other than RP, no procedure) or five categories (multiple procedures with imaging, imaging only, TURP only, biopsy only, others) based on the nature of the procedures in relation to providing stage information. For cases who had received prostate cancer procedures in hospital, the number of episodes of care with one or more cancer-related procedures up to 4 months after initial diagnosis was categorised into three groups (1, 2 and 3+).

Hospital characteristics

Hospitals were grouped by hospital type (principal referral public, other public, major private, other private hospitals). Day stay status indicates whether or not for that episode of care the patient was admitted and separated from hospital on the same day.¹³ In NSW the procedures for reporting to the NSWCR vary across different health service facilities. Public facilities submit notification forms electronically via the Health Information Exchange database, while private hospitals and procedure centres notify the cancer registry electronically via the Secure Notification Portal or by paper notification forms.¹¹

Comorbidities

Comorbidities were coded according to a slightly adapted version of the Charlson comorbidity index,²⁰ excluding prostate cancer and metastatic tumors. This index was derived from ICD-10-AM diagnostic codes as recorded in any hospital separation within 6 months prior to and 6 months after the prostate cancer diagnosis. For each individual in the linked data set a comorbidity score was calculated by multiplying estimated condition weights by comorbid indicators and summing over the 16 relevant conditions.²⁰

Patients' sociodemographic characteristics

Previous work by the authors⁴ suggested that certain sociodemographic characteristics—age at diagnosis, year of diagnosis, country of birth and socioeconomic status (SES) of place of residence at diagnosis—were

associated with the likelihood of a patient with prostate cancer being recorded as 'unknown' stage at diagnosis, so we included these as study factors. Country of birth was categorised into Australia and other countries according to the English proficiency groups.²¹ Country of birth was obtained from hospital notifications and, therefore, was unknown for those cases without a hospital notification. The Index of Relative Socioeconomic Disadvantage derived from the 2001 Australian census was used as an area-based summary measure of SES.²² An additional area-based summary measure was geographic location of place of residence at diagnosis using the Australian Standard Geographic Classification Remoteness Structure,²³ which has been recognised as a nationally consistent measure of geographic remoteness based on the physical road distance to the nearest town or service centre. Recognising that patients can receive cancer care in neighbouring states in Australia and that there is an agreement between the cancer registries in different states to exchange cancer reporting information, a patient living close to an interstate border was flagged as potentially having treatment interstate. For some analyses, living near state borders was combined with the remoteness of location and grouped into five categories (major cities, inner regional areas near state borders, other inner regional, rural areas near state borders and other rural areas).

Statistical analysis

The analyses were divided into three parts: (a) descriptive data analysis to examine the distribution of cases with 'unknown' stage at diagnosis by hospital prostate cancer diagnosis and procedures, to identify possible explanations for 'unknown' stage in relation to prostate cancer care in hospitals; (b) for cases where no clear explanation was found in part (a), logistic regression was used to examine the associations between study factors and a case being recorded as 'unknown' stage in the NSWCR; (c) cancer specific survival by stage at diagnosis and types of cancer-related procedure using the Kaplan-Meier method. All analyses were performed using STATA (V.13.1) (Stata Statistical Software: Release 13 [program]. 13 version: College Station, Texas, USA: StataCorp LP, 2013).

RESULTS

Of the 50 597 men registered in the NSWCR database with a first diagnosis of prostate cancer between 2001 and 2009, 96.5% were histopathologically confirmed. Half of these men had localised disease, 10.1% had regional or distant spread, and 39.9% had 'unknown' stage of disease (table 1). The median age at diagnosis was 68 years, with a younger median age for cases with localised (66 years) and regional spread (65 years) compared to those with distant (77 years) and 'unknown' stage (72 years). Two-thirds of patients were living in a major city at the time of diagnosis. There were 48 757

Table 1 Characteristics of 50 597 men with prostate cancer and distribution of cancer registry recorded stage at diagnosis in New South Wales, Australia, 2001–2009

Characteristics	Localised	Regional	Distant	Unknown	All stages	
	Per cent within categories				n	(%)*
Year of diagnosis						
2001–2005	48.1	5.8	4.1	42.0	23 885	(47.2)
2006–2009	51.8	6.9	3.3	38.0	26 712	(52.8)
Age at diagnosis						
Median age (years)	66	65	77	72	68	
18–64	61.1	8.7	1.7	28.5	17 490	(34.6)
65–74	49.7	6.6	2.6	41.0	18 656	(36.9)
75–99	37.0	3.3	7.6	52.1	14 451	(28.6)
Country of birth						
Australia	50.7	6.5	3.7	39.1	36 090	(71.3)
High English proficiency countries	52.7	7.3	3.8	36.2	5300	(10.5)
Moderate English proficiency countries	52.1	6.0	3.5	38.5	2998	(5.9)
Low English proficiency countries	47.8	5.7	4.5	42.0	5430	(10.7)
Unknown	8.5	1.5	0.6	89.3	779	(1.5)
Geographical location†						
Major cities	50.9	6.6	3.8	38.7	33 509	(66.2)
Inner regional	48.6	5.8	3.7	41.9	12 916	(25.5)
Rural	47.8	6.2	3.4	42.5	4172	(8.2)
Socioeconomic status‡						
Least disadvantaged	57.8	8.3	2.7	31.1	12 695	(25.1)
Middle group	48.0	6.0	4.1	42.0	24 213	(47.9)
Most disadvantaged	46.5	5.3	4.1	44.2	13 689	(27.1)
Lived near state borders						
No	50.6	6.5	3.7	39.2	46 312	(91.5)
Yes	44.5	5.4	3.5	46.7	4285	(8.5)
Cases linked with APDC						
No	24.1	3.3	1.5	71.1	1840	(3.6)
Yes	51.0	6.5	3.8	38.7	48 757	(96.4)
Comorbidity score‡						
No record in time range	19.3	2.4	1.4	76.9	9027	(17.8)
None	57.8	7.5	3.5	31.2	36 826	(72.8)
1+	48.0	5.6	9.6	36.8	4744	(9.4)
Total	50.0	6.4	3.7	39.9	50 597	(100)

*Percentage within total.

†Geographical location and socioeconomic status were based on the men's place of residence at diagnosis.

‡Diseases in the Charlson comorbidity index excluding prostate cancer and metastases within 6 months before and 6 months after diagnosis.

(96.4%) cases linked with at least one APDC episode up to the end of 2010, and 9.4% had a comorbidity score ≥ 1 recorded in the time period of 6 months prior to 6 months after the prostate cancer diagnosis (table 1).

Figure 2 shows the distribution of cases with 'unknown' stage at diagnosis by hospital-reported prostate cancer diagnosis and procedures. Of the total study cohort, 29.4% of cases (n=14 890) did not have a hospital-reported prostate cancer diagnosis, and the majority of these cases did not receive any cancer-related procedures. Of these 14 890 cases, 77.2% were recorded as 'unknown' stage at diagnosis, which accounted for 57.0% of the total number of cases with 'unknown' stage. Only 30 of these cases received RP within 4 months, and 10 of these 30 cases were recorded as 'unknown' stage.

Within 4 months of the initial diagnosis, 70.6% of the study cohort (n=35 707) had a hospital-reported prostate cancer diagnosis, and 24.3% of these cases were recorded as 'unknown' stage at diagnosis (figure 2). High concordance was found between cancer-related procedures and hospital-reported cancer diagnosis, with 97.3% of cases who received cancer-related procedures also having a hospital-reported prostate cancer diagnosis. In this group of cases, 13 177 received RP and 19 864 received a prostate procedure other than RP within 4 months after diagnosis, and a relatively small number of the patients received RP (n=951) and other procedures (n=192) after 4 months. Among men who received procedures within 4 months of diagnosis, the proportion with 'unknown' stage was the lowest for cases who had an RP (7.6%), and highest for the 19 864 cases with a prostate procedure

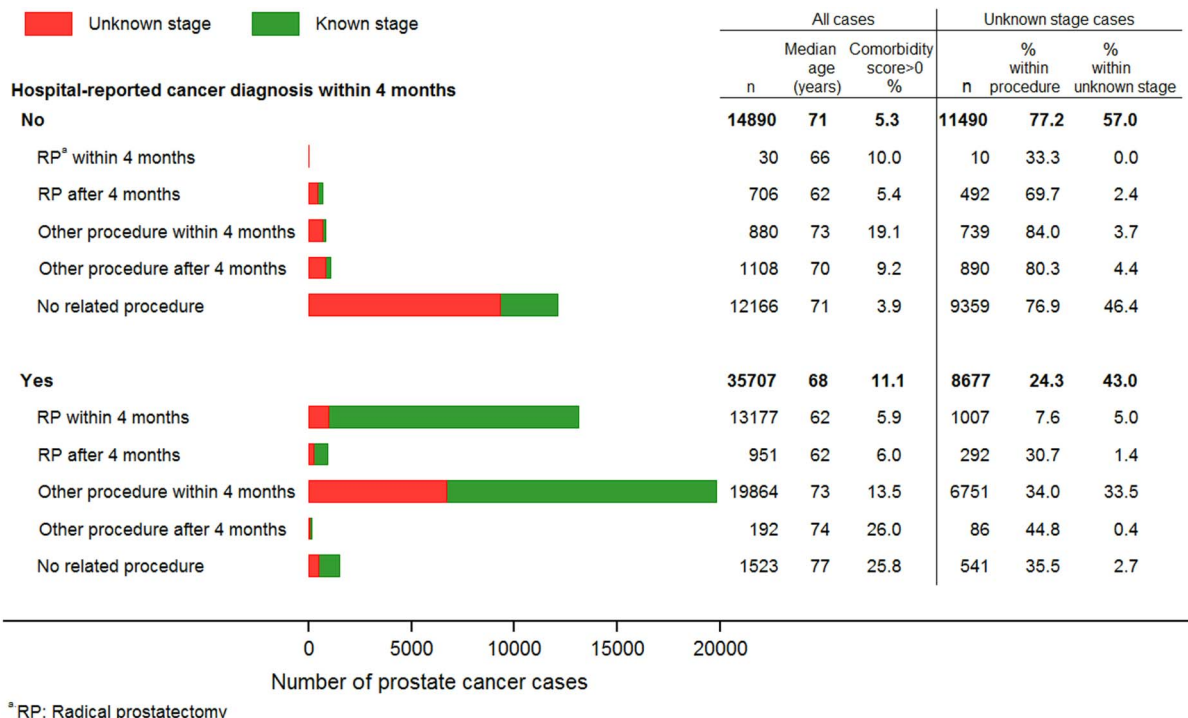


Figure 2 Distribution of men with ‘unknown’ stage at diagnosis by hospital-reported prostate cancer diagnosis and procedures, New South Wales, Australia, 2001–2009 (n=50 597).

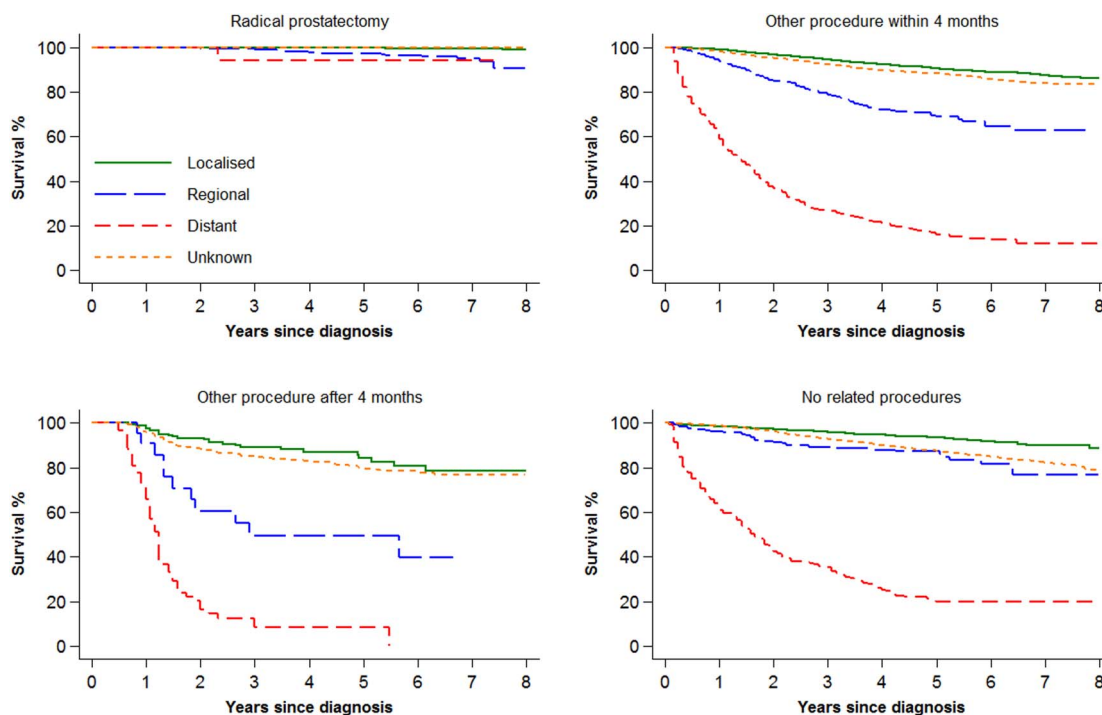


Figure 3 Prostate cancer specific survival by stage at diagnosis recorded in the NSWCR stratified by cancer-related procedures, New South Wales, Australia, 2001–2008, followed up to the end of 2008 (n=43 368).

other than RP (34.0%). The proportion of cases with a comorbidity score ≥ 1 was found to be lower, and median age at diagnosis was younger, for men who had an RP compared with men who had prostate procedures other than RP (figure 2). The Kaplan-Meier survival estimates

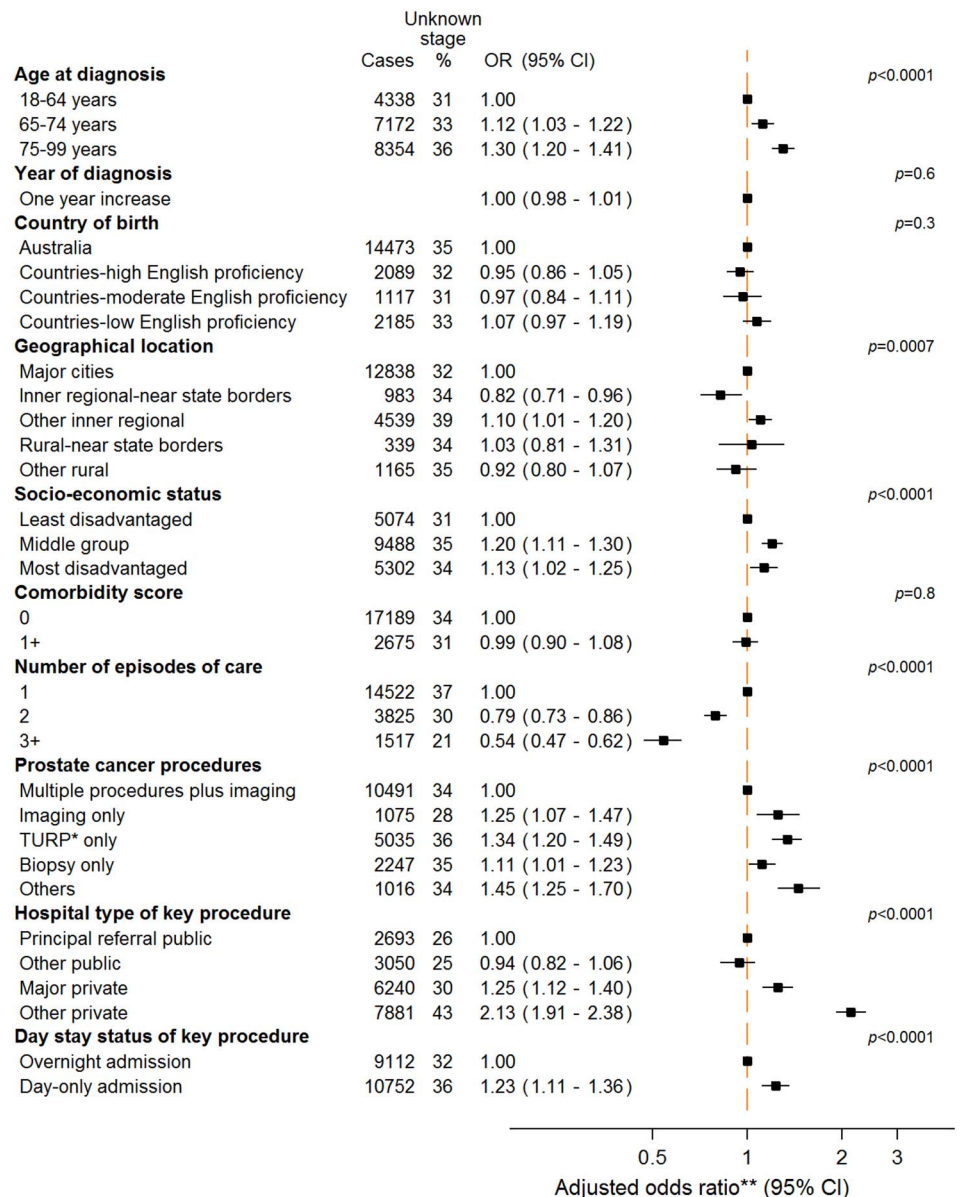
by stage at diagnosis stratified by the type of cancer-related procedure (figure 3) suggested a similar survival outcome for ‘unknown’ stage as localised stage for cases who received cancer-related procedures. For cases who did not receive cancer-related procedures, on average,

the survival outcome for 'unknown' stage was similar to that for regional stage.

The 19 864 cases with a hospital-reported prostate cancer diagnosis and prostate procedures other than an RP within 4 months of diagnosis accounted for a large proportion of this study cohort. The most common procedures received included imaging (58.2%), biopsy (62.6%) and TURP before diagnosis (32.0%), and small proportions with other treatments, TURP post diagnosis and other surgery (see online supplementary resource 3). The procedures are not mutually exclusive and most of the patients who received imaging also received another procedure. About two-thirds of all procedures were performed in private facilities. Most TURP or other prostate surgical procedures had overnight episodes of care, while the majority of biopsy and imaging procedures were recorded as day-only admissions. Results from

bivariable and multivariable logistic regression analysis for these cases are presented in online supplementary resource 4 and figure 4. The odds of a case having 'unknown' stage were significantly higher for men who were older at diagnosis (adjusted OR (aOR) ranged from 1.12 to 1.30), or for cases from more socio-economically disadvantaged areas compared with those from the least disadvantaged areas (aOR ranged from 1.13 to 1.20), or from inner regional areas not close to state borders compared with those who live in major cities (aOR=1.10, 95% CI 1.01 to 1.20). Compared with cases who had undergone multiple procedures plus imaging, all other procedure groups were significantly more likely to be recorded as 'unknown' stage at diagnosis (eg, imaging only, TURP only or biopsy only, aOR ranged from 1.11 to 1.45). Having 'unknown' stage was also more common for cases who were admitted to

Figure 4 Associations between patients' characteristics and 'unknown' stage prostate cancer for patients with a hospital-reported prostate cancer diagnosis and prostate procedures other than radical prostatectomy ≤ 4 months after diagnosis, New South Wales, Australia, 2001–2009 (n=19 864).



*TURP: Transurethral resection of the prostate

**Multivariable model adjusted for all variables listed

private hospitals (aOR ranged from 1.25 to 2.13 compared with principal referral hospital) and those with day-only admission (aOR=1.23 compared with overnight admission, 95% CI 1.11 to 1.36). The proportion of cases with 'unknown' stage fell with the number of episodes of care recorded (aOR ranged from 0.54 to 0.79), and was lower for cases who lived in inner regional areas close to state borders than for those who lived in major cities (aOR: 0.82, 95% CI 0.71 to 0.96).

DISCUSSION

Our findings show that a considerable proportion of the 'unknown' stage prostate cancers arises because of a lack of definitive staging evidence or clinical stage made available to the NSWCR. This particularly occurs when this information has been collated by non-notifying facilities such as private consulting rooms (figure 1). Over half of the total cases with 'unknown' stage prostate cancer in our study did not have a hospital-reported prostate cancer diagnosis and did not receive a hospital-based cancer-related procedure within 4 months after diagnosis. For patients who had a hospital-reported prostate cancer diagnosis within 4 months of the initial diagnosis, RP is the most important prostate cancer procedure for providing definitive stage information that allows the NSWCR to assign a stage. Prostate procedures other than RP provide only limited stage information and the NSWCR rarely assigns stage based on a pathology report from a biopsy or TURP. We found that patients with prostate cancer who had procedures other than RP were older and more commonly had comorbidities than patients who had an RP. These findings suggest that advancing age and the presence of comorbidities were associated with more conservative treatment with less complete diagnostic assessment, which is then associated with 'unknown' stage.^{3 6 7} Among the group of patients who received prostate procedures other than RP and had a hospital-reported prostate cancer diagnosis within 4 months of the initial diagnosis, those who were older at diagnosis, lived in socioeconomically disadvantaged areas, had prostate procedures other than multiple procedures plus imaging, who were admitted to private hospitals, or who had a day-only admission were more likely to have 'unknown' stage recorded in the NSWCR.

The main strength of our study is that it is the first systematic examination of hospital inpatient cancer services and 'unknown' stage at diagnosis using population-based linked records from routinely collected administrative data sets. There are numerous advantages to using administrative data, and record linkage between data sets can add further value to these resources, with their population coverage ensuring that the results are representative. Our study was restricted to patients with prostate cancer recorded in the NSWCR, so findings might not be generalisable to other population-based cancer registries, nor will they be representative of other cancer

types. Our study was limited to inpatient hospital services and does not capture all prostate cancer services. A previous study reported that the APDC captured 90% or more of all RPs, only missing a small number of patients treated interstate, but it underenumerated non-surgical treatment.¹⁹ Comorbidities also tend to be underenumerated in the APDC.^{24 25}

To assign stage, the NSWCR requires a sufficient level of evidence, such as the degree of spread reported in hospital notification forms, or definitive stage information provided in a pathology report of an RP. If this evidence is not available, a patient's stage is recorded as 'unknown'.¹² This is an effective quality control measure, which aims to ensure that the assigned stage is based on the most complete information available. For a cancer such as prostate, however, where over 70% of patients may not receive surgical treatment, or 30% of patients may not receive active treatment within 4 months after diagnosis, the absence of clinical information appears to result in a relatively high proportion of cases being recorded as 'unknown' stage. We found that 95% of patients with prostate cancer with 'unknown' stage received no surgical treatment and that half of these patients received no other active treatment up to 4 months after diagnosis. The relatively good survival and low proportion of cases with comorbidities among this group of patients suggests that it is likely that they were primarily low-risk cases who were diagnosed by a core needle biopsy alone due to elevated PSA levels, and who did not receive any further staging assessment within 4 months of diagnosis, or some of these patients had a bone scan or imaging outside a hospital and then went onto watchful waiting or androgen deprivation therapy (figure 1). For these cases, the only notification to the NSWCR is likely to be the pathology report confirming the cancer diagnosis. The staging system used by the NSWCR does not allow for stage to be assigned from a needle biopsy alone, so low-risk patients diagnosed by a core needle biopsy alone due to elevated PSA, unlike that used in the USA SEER system (classified as T1c stage),²⁶ will be recorded as 'unknown' stage by the NSWCR. Unfortunately, the NSWCR could not adopt this staging system due to the absence of clinical information provided by the notifying institutions.

For patients who received a cancer-related procedure other than RP within 4 months of diagnosis, we found a higher proportion of cases with 'unknown' stage among patients admitted to private hospitals compared to patients treated in public hospitals, after adjusting for the cancer-related procedures received. This may be due to potential issues related to documentation in the cancer notification process, resulting in the available stage information not being received by the NSWCR. A large proportion of biopsy and imaging procedures were conducted during day-only admissions, but even after adjusting for the cancer-related procedures undertaken, patients with day-only admissions were still more likely to be recorded as having 'unknown' stage at diagnosis in

the NSWCR than patients who were admitted overnight. It is possible that this is because the stage information was not available at the time of reporting for those day-only admissions.

We observed some differences in the proportions of cases with 'unknown' stage by place of residence. Cases living in more socioeconomically disadvantaged areas or inner regional areas away from state borders were more likely to be recorded as 'unknown' stage, even after accounting for hospital cancer services and other factors. This is possibly related to the differences between hospital services in different socioeconomic areas, although with the data available to us it is not possible to be sure of the reasons for these differences. We also found that after adjusting for other factors, patients who lived in inner regional areas near state borders were less likely to be recorded as 'unknown' stage than those living in major cities. This may be because these patients were treated at interstate health services, and there is a general agreement between the cancer registries in different states to exchange information for patients who seek treatment from the neighbouring state, but this hypothesis could not be confirmed based on the data available for this study.

Cancer stage as a prognostic indicator is essential for researchers using cancer registry data to study outcome disparities. Our previous research in this area,⁴ and the results of this current study, suggest that because prostate cancer cases with 'unknown' stage consisted of either patients who did not receive definitive staging due to treatment decisions based on older age or the presence of comorbidities, or patients who had low-risk disease diagnosed by a core biopsy only, the 'true' stage of patients recorded as having 'unknown' stage is likely to be a mixture of stages.⁴ The composition of this mixture of stages may vary by the type of cancer-related procedures received; however, the relatively good survival among this group of patients with 'unknown' stage suggests that the majority of these cases are likely to have early stage disease. Also, the USA SEER Cancer Statistics Review reported that in the period 2003–2009, 81% of all prostate cancer cases were diagnosed with localised stage.²⁷ As the patterns of PSA testing in Australia are similar to those in the USA,²⁸ we may expect a similar stage distribution among Australian patients with prostate cancer, which does suggest that the majority of those with 'unknown stage' stage are likely to have localised disease. Understanding the reasons why stage was 'unknown' to the cancer registry is an important step towards understanding the potential biases that may be caused by the incomplete stage information. For example, as the proportion of cases with 'unknown' stage is higher among cases living in more socioeconomically disadvantaged areas, when examining the socioeconomic disparities in patients' outcomes adjusting for incomplete stage data, the estimated differences may be biased. The direction of bias can be towards or away from the null, depending on the variation in the

composition of this mixture of stages by SES for cases with 'unknown' stage recorded in the NSWCR. In addition, complete-case analysis as the default option in most statistical software can provide biased estimates and lead to a considerable loss in statistical power. Multiple imputation is a flexible statistical method for dealing with missing data, and has been increasingly used in epidemiological studies.²⁹ Further research focusing on the validity of using MI for 'unknown' stage and the true stage distribution for cases with 'unknown' stage is warranted.

IMPLICATIONS AND CONCLUSIONS

This study found that hospital cancer services are important determinants of the availability of cancer stage at diagnosis in the cancer registry for prostate cancer. Over half of 'unknown' stage prostate cancer cases do not have a hospital-reported prostate cancer diagnosis within the cancer registry's 4-month reporting window. Men living in more disadvantaged areas or those attending private facilities were more likely to be recorded as having 'unknown' stage, even after adjustment for other factors. The results reflect the nature of the investigative and follow-up pathways for prostate cancer, and reveal a problem for patients with prostate cancer who are managed outside notifying healthcare facilities by treating clinicians or by private facilities. We speculate that to reduce the variation in practice and reporting would improve the completeness of stage data in the NSWCR. The recent establishment of the NSW Prostate Clinical Cancer Registry which intends to collect grade, TNM stage and PSA levels directly from clinicians' notes may alleviate problems in the future when reporting prostate cancer stage information for the NSW population. If successful, this will allow much finer gradation of prognostic categorisation than is currently available in the NSWCR data. Nonetheless, the clinical registry will not revise clinical stage for cases diagnosed prior to 2015, so a clear understanding of 'unknown' stage disease as recorded in the NSWCR remains important. Furthermore, this study provided clues on the direction of the possible bias due to cases with 'unknown' stage in epidemiological studies using this important historical data to investigate survival outcomes. Further studies to explore the composition of 'unknown' stage and develop a valid method to manage this incomplete stage data are necessary and could help us to increase the usage of these valuable cancer registry data.

Author affiliations

¹Cancer Research Division, Cancer Council NSW, Sydney, New South Wales, Australia

²Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia

³Menzies Health Institute Queensland, Griffith University, Gold Coast, Queensland, Australia

⁴New South Wales Cancer Registry, Cancer Institute NSW, Sydney, New South Wales, Australia

⁵Discipline of Surgery, University of Sydney, Sydney, New South Wales, Australia

⁶Department of Urology, Westmead Hospital, Westmead, New South Wales, Australia

⁷School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia

Twitter Follow Qingwei Luo @qingwei

Acknowledgements The authors thank the New South Wales Ministry of Health for providing the data, the New South Wales Centre for Health Record Linkage for conducting record linkage and Clare Kahn for editorial assistance.

Contributors QL designed the study with input from XQY and DLO. QL performed the literature review, statistical analysis and interpretation of the results and drafted the manuscript. DLO and XQY guided the analysis and revised the manuscript. DPS contributed to the interpretation of the results. DEG provided advice on the data analysis. CC-Y provided technical input regarding cancer registry procedures. MP provided advice on clinical practice. DPS, DEG, CC-Y and MIP reviewed the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by the Prostate Cancer Foundation of Australia (YI-0410). DPS was supported by an Australian National Health and Medical Research Council Training Fellowship (App1016598).

Competing interests None declared.

Ethics approval This study was approved by the NSW Population and Health Services Research Ethics Committee (Reference: 2010/04/223). The study uses existing, routinely collected and de-identified health records and does not involve any direct contact with participants. The Ethics Committee waived the need for individual consent from participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

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 and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Yu XQ, O'Connell DL, Gibberd RW, *et al.* A population-based study from New South Wales, Australia 1996–2001: area variation in survival from colorectal cancer. *Eur J Cancer* 2005;41:2715–21.
2. Yu XQ, Smith DP, Clements MS, *et al.* Projecting prevalence by stage of care for prostate cancer and estimating future health service needs: protocol for a modelling study. *BMJ Open* 2011;1:e000104.
3. Klassen AC, Curriero F, Kulldorff M, *et al.* Missing stage and grade in Maryland prostate cancer surveillance data, 1992–1997. *Am J Prev Med* 2006;30(2 Suppl):S77–87.
4. Luo Q, Yu XQ, Cooke-Yarborough C, *et al.* Characteristics of cases with unknown stage prostate cancer in a population-based cancer registry. *Cancer Epidemiol* 2013;37:813–19.
5. Yu XQ, O'Connell DL, Gibberd RW, *et al.* Trends in survival and excess risk of death after diagnosis of cancer in 1980–1996 in New South Wales, Australia. *Int J Cancer* 2006;119:894–900.
6. Gurney J, Sarfati D, Stanley J, *et al.* Unstaged cancer in a population-based registry: prevalence, predictors and patient prognosis. *Cancer Epidemiol* 2013;37:498–504.
7. Merrill RM, Sloan A, Anderson AE, *et al.* Unstaged cancer in the United States: a population-based study. *BMC Cancer* 2011;11:402.
8. Koroukian SM, Xu F, Beard H, *et al.* Complexity of care needs and unstaged cancer in elders: a population-based study. *Cancer Detect Prev* 2007;31:199–206.
9. Liff JM, Chow WH, Greenberg RS. Rural-urban differences in stage at diagnosis. Possible relationship to cancer screening. *Cancer* 1991;67:1454–9.
10. Fritz A, Percy C, Jack A, *et al.*, eds. *International classification of diseases for oncology*. 3rd edn. Geneva, Switzerland: World Health Organisation, 2000.
11. Department of Health New South Wales. *Cancer registry—notifying cancer cases to the NSW central cancer registry*. NSW Department of Health, 2009.
12. Cancer Institute New South Wales. *Cancer survival in New South Wales 200206*. Sydney: Cancer Institute New South Wales, 2012.
13. Centre for Health Record Linkage (CHeReL). *Data dictionaries, NSW admitted patient data collection (APDC)*. Centre for Health Record Linkage, 2014.
14. CHeReL. *Data linkage quality assurance*. Centre for Health Record Linkage (CHeReL), 2014.
15. Cancer Institute NSW. *NSW cancer registry—data dictionary*. Sydney: Cancer Institute NSW, 2016.
16. Esteban D, Whelan S, Laudico A, Parkin D, eds. *Manual for cancer registry personnel*. IARC Technical Report No. 10. Lyon, France: IARC, 1995.
17. Young JJ, Roffers S, Ries L, Fritz A, Hurlburt A, eds. *SEER summary staging manual—2000: codes and coding instructions*. Bethesda, MD, National Cancer Institute, NIH Pub. No. 01-4969, 2001.
18. Australian Consortium for Classification Development. *ICD-10-AM/ACHI/ACS*. The National Centre for Classification in Health, 2016.
19. Goldsbury DE, Smith DP, Armstrong BK, *et al.* Using linked routinely collected health data to describe prostate cancer treatment in New South Wales, Australia: a validation study. *BMC Health Serv Res* 2011;11:253.
20. 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.
21. Department of Immigration and Multicultural and Indigenous Affairs. 2001 Classification of Countries into English Proficiency Groups. *Department of Immigration and Multicultural and Indigenous Affairs (DIMIA) 2003;C01.2.0: 2003*.
22. Australian Bureau of Statistics. *Information paper: 2001 census socioeconomic indexes for areas (SEIFA)*. Canberra: Australian Bureau of Statistics (ABS), 2001.
23. Australian Bureau of Statistics. *ASGC remoteness classification: purpose and use*. Canberra: Commonwealth of Australia, 2003.
24. 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.
25. Goldsbury DE, Armstrong K, Simonella L, *et al.* Using administrative health data to describe colorectal and lung cancer care in New South Wales, Australia: a validation study. *BMC Health Serv Res* 2012;12:387.
26. Beahrs O, Henson D, Huntter R, eds. *AJCC cancer staging manual*. 4th edn. Philadelphia: Lippincott, American Joint Committee on Cancer, 1992.
27. Howlader N, Noone A, Krapcho M, *et al.*, eds. *SEER Cancer Statistics Review, 1975-2010*. Bethesda, MD: National Cancer Institute. http://seer.cancer.gov/csr/1975_2010/ based on November 2012 SEER data submission, posted to the SEER web site, April 2013.
28. Howard K, Brenner AT, Lewis C, *et al.* A comparison of US and Australian men's values and preferences for PSA screening. *BMC Health Serv Res* 2013;13:388.
29. Sterne JA, White IR, Carlin JB, *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.