RESEARCH

Accuracy of coded cause of death data: a study based on primary liver cancer

Hoa Nguyen,[1] Kwang C Yee,[2, 3] Michael Braude,[4] Cristina Moldovan,[5] Fiona Cocker,[2] Andrew J Palmer,[1] and Barbara de Graaff.[1]

[1] Menzies Institute for Medical Research, University of Tasmania; [2] School of Medicine, University of Tasmania; [3] Royal Hobart Hospital, Tasmania; [4] School of Clinical Sciences, Monash University and Monash Health; [5] Icon Cancer Centre, Hobart

Abstract

Introduction: Coded cause of death data from death certificates is important for estimating mortality related to cancers and their treatments. However, coding may be inaccurate or complicated by variation in mode of diagnosis (for example, radiology versus pathology, or in methodology between government agencies). This may have important reporting and funding implications.

Aims: (1) To investigate the level of agreement for cause of death data between the Australian Bureau of Statistics (ABS) and Tasmanian Cancer Registry (TCR), and to compare these with independent, blinded medical practitioner opinions on cause of death. (2) To estimate the impact of different coding practices and resulting cause of death data on cause-specific survival.

Methods: Causes of death were compared between the ABS and TCR, and discrepancies were independently reviewed by specialist medical practitioners. Cohen's Kappa statistics were applied to evaluate the degree of concordance between the ABS, TCR and medical practitioner opinions regarding cause of death. The cumulative incidence function was used to estimate cause-specific survival time in the presence of a competing risk framework according to sex, place of residence, country of birth, and type of PLC.

Results: A minimal level of agreement (Kappa=0.35) was observed when comparing the TCR and ABS cause of death data. Agreement between the TCR and medical practitioners was weak (Kappa=0.51), moderate between the ABS and medical practitioners (Kappa=0.61), strong (Kappa=0.87) between the medical practitioners. These results reflect a greater level of agreement between medical practitioners than between coding agencies. Overall, cause-specific survival time was similar across the TCR, ABS and medical practitioners by sex, place of residence and country of birth, however, regarding type of PLC, only a small difference was observed.

Conclusions: Reporting of mortality data can vary substantially between agencies/institutions. Utilisation of specialist clinician oversight might improve data cohesion and fidelity. Overall, cause-specific survival time was similar across the TCR, ABS and medical practitioners, but a small difference was observed for the type of PLC. As PLC is a low survival cancer, these results may not apply to cancers with better survival such as breast cancer.

Tasman Medical Journal 2022: 4(2); 12-20

Introduction

Primary liver cancer or hepatoma (PLC) is the fourth leading cause of cancer mortality and the sixth most frequent cancer globally.[1] In Australia, age-standardised incidence and mortality rates have increased substantially over the past three decades and are projected to increase further in coming decades.[2-4.] Cancer mortality data play an important role in the estimation of population-based mortality statistics and survival rates.[5] In turn, these data are used in research to develop and evaluate health policies and resource allocation decisions. Therefore, the information needs to be as accurate and complete as possible. In Australia, a death must be registered with the state or territory jurisdictional Registry of Births, Deaths and Marriages as soon as possible.[5,6] The registration is

based on the death certificate completed by the attending medical practitioner or a coroner.[5,6] The certificate includes information on the underlying cause of death and associated cause of death. According to the Australian Institute of Health and Welfare (AIHW), the *underlying cause of death* is defined as "the disease or injury which initiated the train of morbid events leading directly to death".[5] The *associated causes of death* are all causes that contributed to the death, other than the underlying cause of death.[5]

Coding Cause of Death by the Australian Bureau of Statistics

Information from death certificates is provided to the Australian Bureau of Statistics (ABS) monthly, for coding based on the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)[7] (Fig. 1). The ABS also receives information from the National Coronial Information System (NCIS), from a relative or other person familiar with the deceased person, or where the death has occurred in an institution, from an official. In turn, the NCIS collects data from a broad range of sources, including reports from coroners, post-mortem and toxicology reports and police summary of death reports. The coding enables comparison of mortality statistics over time and between different areas. This process produces the national Cause of Death Unit Record File (COD-URF),[8] the mortality dataset including all national information on causes of death registered in Australia by the ABS. The information from the COD-URF is the major source of Australian cancer mortality and survival data published by the AIHW.[5] Cause of death data from the COD-URF is referred to here as the ABS dataset.

Coding Cause of Death by the Tasmanian Cancer Registry

The state/territory Registries of Births, Deaths and Marriages (including Tasmania) also report death certificate data to jurisdictional cancer registries.[5] This includes up to five causes of death, eight antecedent causes of death, and two other important conditions recorded at the time of death. In most cases, the underlying cause of death reported in the death certificate is accepted, unless the cause requires further investigation. For example, if the Tasmanian Cancer Registry (TCR) has no record of an individual having hepatocellular carcinoma and this is recorded as the underlying cause of death, further investigation occurs. In such situations, experienced medical coders in the TCR access the Digital Medical Record (DMR) of each public hospital patient and review the cause of death data in the death certificate in the context of the patient's DMR (this does not occur for patients in

private hospitals). A single underlying cause of death is then recorded: for deaths which cancer was the underlying cause, the TCR records the topography and morphology of the tumour as the cause of death. For non-cancer deaths, no cause of death information is recorded.

The TCR uses the International Classification of Diseases for Oncology, third edition (ICD-O3)[8] for coding the site (topography) and morphology of all cases to determine the underlying cause of death. For the purposes of comparing between different geographic areas of Australia, all cause of death data are converted to the ICD-10[7] (Fig. 1).

Several studies have identified that cause of death information obtained from death certificates may be inaccurate.[9-13] For example, a study in the US reported that 50.8% of reviewed death certificates had at least one error in the reported cause of death.[9] To our knowledge, no studies have investigated this issue in Australia.

The primary aim of this study was to investigate the level of agreement on cause of death between the ABS, TCR, and medical practitioner diagnoses in patients who died from lever cancers. A secondary aim was to understand the impact of different coding practices and resulting cause of death data on estimates of cause-specific survival.

Methods

In previous work on the same dataset, we estimated survival for Tasmanian patients diagnosed with PLC (unpublished). Data on 293 cancer notifications, date and cause(s) of death were taken from the TCR and ABS datasets. We compared these data and found that the stated cause of death differed in 112 cases (48.3%).

Data sources The present cohort was defined as all individuals with a PLC notification to the TCR between 01/01/2007 and 31/12/2015, aged ≥18 years, and deceased. The ABS dataset provided information on date of birth, sex, date of death, age at death, underlying cause of death and up to 20 associated cause of death for each patient. Linkage to the ABS COD-URF was undertaken by the Tasmanian Data Linkage Unit using probabilistic linkage. This approach links records from different datasets using variables such as name, date of birth, address, gender, address. Match weights are applied to each field, and linkage is undertaken based on the greatest probability of matches belonging to the same person.[15]
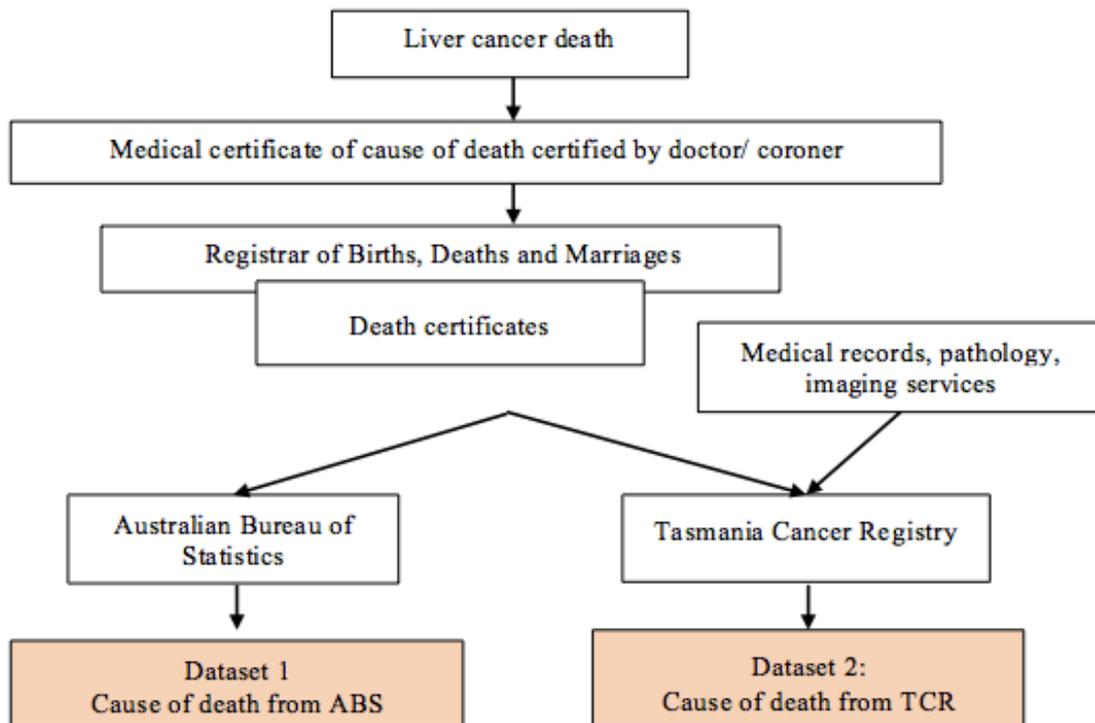
```
                    Liver cancer death

        Medical certificate of cause of death certified by doctor/ coroner

            Registrar of Births, Deaths and Marriages

                    Death certificates

                                         Medical records, pathology,
                                               imaging services

    Australian Bureau of                     Tasmania Cancer Registry
         Statistics

        Dataset 1                                 Dataset 2:
   Cause of death from ABS                  Cause of death from TCR
```

Fig. 1. Diagram of data handling for comparing cause of death data from the Australian Bureau of Statistics and the Tasmanian Cancer Registry

Data handling

Deidentified records of the underlying cause of death in the TCR dataset with the ABS dataset and information from death certificates were compared. For cases in which a discrepancy was identified, an Excel spreadsheet was developed for independent review. Three specialist medical practitioners were involved in this review, including two consultant hepatologists/gastroenterologists (referred as medical practitioner 1 and medical practitioner 2) with expertise in PLC. Where discrepancies remained, a third reviewer, a consultant medical oncologist (referred as medical practitioner 3) made the final determination of the cause of death based on the group of information from the ABS, TCR and death certificates. Each practitioner independently reviewed the available data.

The deidentified spreadsheet included:
1. ABS data (1 underlying cause of death and up to 8 associated causes of death);
2. TCR data (date of birth, sex, age at death, morphology code, topography code, ICD-10 code, underlying cause of death); and
3. Information from death certificates (including up to five causes of death, the disease or condition directly leading to death; up to 8 antecedent causes, and other significant conditions, contributing to the death but not related to the disease, injury or condition causing it).

Statistical analysis Descriptive statistics were used to describe the characteristics of the cohort. Cohen's Kappa (κ) statistics were applied to evaluate the degree of concordance in cause of death between the ABS, TCR and medical practitioner's judgement,[16] and to assess the degree of agreement by chance alone between the different methods of ascertaining cause of death. Values range from 0 to 1, representing no and perfect agreement respectively, and the interpretation of intermediate Cohen's Kappa values is shown in Table 1.[16]

The cumulative incidence function (CIF) was used to estimate deaths caused by non-PLC cases based on the final causes of death provided by the medical practitioners in the presence of competing risks.[17] The CIF curves and the subdistribution hazard ratios (SHR) were generated to describe the incidence of death from the event of interest (liver cancer) over time and the competing risk of death (other causes of death) for each dataset.[17,18] Estimates for the cumulative incidence of death after notification of PLC were undertaken according to sex; place of residence (urban or rural) according to the Australian Statistical Geography Standard 2016;[19] type of PLC based on the TCR data according to the ICD-10 codes for hepatocellular carcinoma (HCC) with the code C22.0, cholangio-carcinoma with the code C22.1 and unspecified types

Kappa	Agreement	% reliable data
0-0.20	None	0-4%
0.21-0.39	Minimal	4-15%
0.40-0.59	Weak	15-35%
0.60-0.79	Moderate	35-63%
0.80-0.90	Strong	64-81%
> 0.90	Almost Perfect	82-100%

Table 1: Interpretation of Cohen's Kappa values[16]

(C22.9). Patients in the last group tend to be diagnosed at an advanced stage, when further investigation to identify the type of PLC is often unwarranted because it has little impact on treatment. Other rare forms of PLC were excluded from statistical analyses due to small numbers (≤5) (C22.2, hepatoblastoma; C22.3, angiosarcoma of liver; C22.4, other sarcomas of liver; and C22.7 other specified carcinomas of liver).[7] Country of birth was coded according to the Standard Australian Classification of Countries 2016.[20] Due to the variability of country of birth data, and therefore the small numbers of patients that could be grouped, this was reported as either Australian- or overseas-born. Where the underlying cause of death was recorded as PLC, this was considered as the event of interest; other causes of death were considered as competing risks.

Data were analysed using Stata 15 (Ver.15, College Station, Texas, USA). Cell counts ≤5 cases were suppressed to prevent any possibility of identification. $P < 0.05$ was deemed statistically significant.

Ethical approval for this study was obtained by the University of Tasmania Human Research Ethics Committee (H0016958).

Results
Between 2007 and 2015, 293 patients were diagnosed with PLC in Tasmania, of whom 239 had died. Of these, seven cases were not matched between the ABS and TCR datasets. Six were marked as deceased in the TCR but not in the ABS, and one marked deceased in the ABS was not found in the TCR. These patients were excluded in the agreement analysis as the data provided no potential for comparison. Thus 232 deceased cases with matches in the TCR and ABS datasets were included in the study.

Characteristics		N (N = 232)	%
Sex	Males	173	74.6
	Females	59	25.4
Residence	Rural	86	37.1
	Urban	146	62.9
Place of birth	Australian born	158	68.1
	Overseas born	30	12.9
	Not stated/ missing	44	19.0
Cancer	HCC	107	46.1
	Cholangiocarcinoma	49	21.1
	Unspecified type	71	30.6
	Other types	5	2.2

Table 2: Tasmanian Cancer Registry (TCR): reported characteristics of deceased cases among liver cancer patients in Tasmania (2007-2015)

The majority of cases were male (74.6%, n=173), most resided in urban areas (62.9%, n=146) and over two-thirds were born in Australia (68.1%, n=158). Based on the TCR data, the most common cause of death was HCC (46.1%, n=107), followed by cholangiocarcinoma (21.1%, n=49). The remaining 30.6% (n=71) were reported as unspecified type of PLC or 'other' (C22.2 hepatoblastoma, C22.3 angiosarcoma of liver, C22.4 other sarcomas of liver, C22.7 other specified carcinomas of liver). Table 2 shows the characteristics

of all included cases. The comparisons regarding the underlying cause of death by the ABS, TCR and medical practitioners were conducted in three stages. Firstly, the 232 ABS and TCR causes of death were reviewed, with discrepancies identified in 48.3% (n=112) of cases, a minimal level of agreement (Kappa=0.35, p<0.001) (Table 3).

The 112 discrepancies were then reviewed independently by medical practitioners 1 and 2. Of these cases, 16 (17.0%) discrepancies remained and were further reviewed by medical practitioner 3 to achieve a consensus. Inter-rater reliability between the TCR and final consensus from the medical practitioners showed weak agreement (Kappa=0.51, p<0.001) and between the ABS and medical practitioners' moderate agreement (Kappa=0.61, p<0.001).

Comparison	Agreement frequency (%)	Kappa
ABS-TCR	120 (51.3%)	0.35[*]
TCR-MP	150 (64.7%)	0.51[*]
ABS-MP	168 (72.4%)	0.61[*]

Table 3: The agreement between the ABS, TCR and the medical practitioners regarding underlying cause of death of liver cancer patients in Tasmania (N=232) MP = Medical Practitioners. [*] p<0.001

The highest inter-rater reliability was observed between medical practitioners 1 and 2 (Kappa = 0.87, p<0.001). This reflects a greater consistency between these experts when deciding the underlying cause of death. Although showing similar trends of agreement across the different comparisons, the Kappa statistics revealed more robust outputs than the percent agreement.

Cumulative incidence of cause specific deaths
The cumulative incidence of death after being diagnosed with PLC is presented according to (1) sex (males/females), (2) place of residence (urban/rural), (3) country of birth (Australia/overseas), and (4) type of PLC (unspecified/HCC/cholangiocarcinoma) (Table 4).

The calculations from three different sources of information (TCR, ABS and medical practitioners) returned many similar SHR results. First, the cumulative incidence of death trended higher in males (Fig. 2A). Across the datasets the SHR were 0.74 (95%CI, 0.52-1.1, TCR data), 0.77 (0.53-1.12, ABS data; and 0.74 (0.51-1.01, medical practitioners' data).

Second, the cumulative incidence of death with urban as the reference group was largely similar for rural and urban residence, with each of the three datasets providing similar results (Fig. 2B): the SHR being 0.90 (95%CI: 0.68-1.20), 0.93 (0.69-1.26), and 0.93 (0.69-1.25), respectively from the TCR, ABS and medical practitioners data. The different estimations based on the three datasets (i.e., cause of death datasets from the TCR, ABS and consensus from the medical practitioners), consistently showed that the cumulative incidence of death was higher in patients born in Australia than those overseas-born (Fig. 2C), with the SHR being 0.68 (95%CI: 0.57-0.82) based on the TCR data, 0.63 (95%CI: 0.52-0.77) based on the ABS data, and 0.69 (95%CI: 0.58-0.83) based on the medical practitioner data. All these estimates were statistically significant (p<0.001).

Variables		Reference group	SHR (95% CI)		
			Tasmanian Cancer Registry	Australian Bureau of Statistics	Medical Practitioners
Sex	Female	Male	0.74 (0.52-1.1)	0.77 (0.53-1.12)	0.74 (0.51-1.01)
Residence	Rural	Urban	0.90 (0.68-1.20)	0.93 (0.69-1.26)	0.93 (0.69-1.25)
Place of birth	Overseas	Australia	0.68 (0.57-0.82)[***]	0.63 (0.52-0.77)[***]	0.69 (0.58-0.83)[***]
	HCC		0.48 (0.33-0.70)[***]	0.45 (0.31-0.67)[***]	0.64 (0.42-0.96)[*]
Cancer	Cholangio-carcinoma	"Unspecified"	0.50 (0.33-0.77)[**]	0.40 (0.26-0.63)[***]	0.69 (0.44-1.08)

Table 4: Hazard Ratios by sex, place of residence, country of birth, and type of liver cancer from the ABS, TCR, Medical Practitioners' data. Note: [***] p<0.001; [**] p<0.01; [*] p<0.05

Figure 2: Cumulative incidence function curves illustrate deaths from liver cancer patients and other reasons by gender (2-A), place of residence (2-B), country of birth (2-C), and type of liver cancer (2-D) from ABS, TCR, Medical Practitioners' data
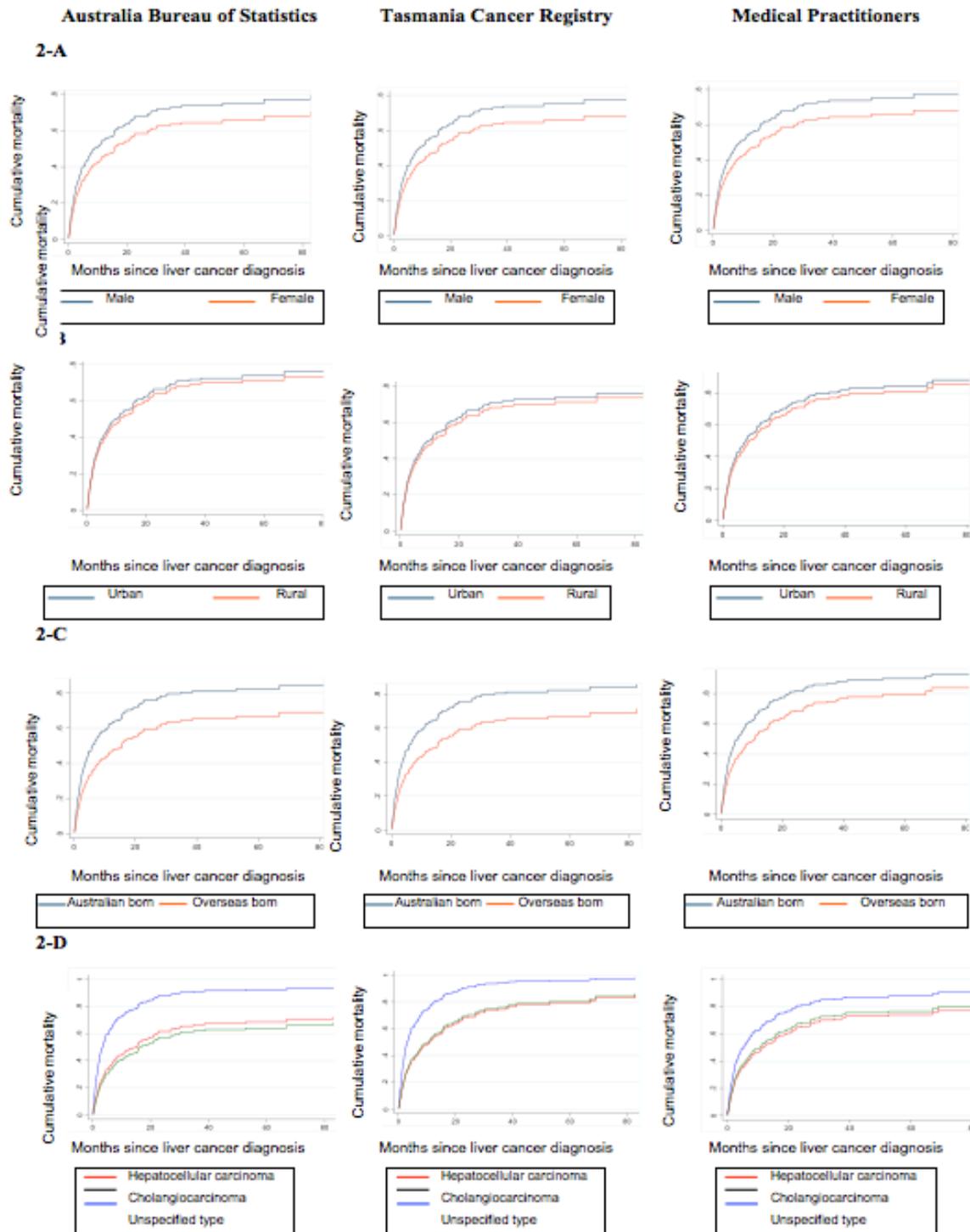
Lastly, the type of PLC was the only factor that contributed to inconsistent estimates in the SHR (Table 4) and survival time across the datasets (Fig. 2D). The SHR (reference group: 'Unspecified type') were 0.48 (95%CI: 0.33-0.70) and 0.50 (0.33-0.77) for HCC and cholangiocarcinoma groups respectively based on the TCR data. The SHR based on ABS data were similar with 0.45 (95% CI: 0.31-0.67) and 0.40 (0.26-0.63) for HCC and cholangiocarcinoma respectively, and for the medical practitioner data the SHR were 0.64 (0.42-0.96) and 0.69 (0.44-1.08) for HCC and cholangiocarcinoma respectively. The cumulative incidence of death was highest in cases with unspecified PLC in all datasets. This largely reflects the nature of this unspecified group, that is, diagnosed at an advanced stage with no subsequent clinical investigations conducted to inform treatment. However, while significant differences were found

when estimating the SHR in the ABS and TCR datasets, the estimation based on the medical practitioner data was not statistically significant (p=0.11) between the patients with cholangiocarcinoma and HCC.

Discussion

Cause of death data derived from death certificates is of critical importance in estimating mortality related to cancers and their treatments. This study describes both the processes of recording the cause of death and the impact on cause-specific survival using a competing risk framework for a cohort of PLC patients. Our results showed that the overall concordance between the ABS and TCR was minimal (Kappa = 0.35), with discrepancies present for nearly half the cases. When cause of death was based on death certificates only, the results between the ABS and specialists showed moderate agreement (Kappa = 0.61). The discrepancies identified between the TCR and ABS were largely due to the different sources of information used: the TCR, death certificates; the DMR, pathology and imaging services; and the ABS the National Coronial Information System, relative or other person familiar with the deceased person, or the institution at which the death occurred.

To our knowledge, no previous studies have been published on the accuracy of cause of death data and the impact on survival estimates for PLC. A study using a similar approach for breast cancer, compared cause-specific survival time from death certificates with coded cause of death from the Geneva Cancer Registry. This Registry codes cause of death based on death certificate data along with all available clinical information. The authors reported a high level of agreement (Kappa = 0.82), and 8.8% of cases were misclassified. Overall cause-specific survival time was not affected, but differences were observed for some groups for example patients aged over 80 years.[10] In our study, 71 cases were diagnosed with morphology code 8003 (Neoplasm, malignant) based on the ICD-O3.[8] In the absence of histology data to accurately describe the morphology of a tumour, the TCR coded cases as C22.*, with the 8003 morphology code recorded as unspecified liver cancer (C22.9). For cases thus coded, a high rate of discrepancies in cause of death (54/71 cases), between the TCR and ABS was observed. These cases contributed to the minimal agreement between the ABS and TCR. A Victorian study[21] highlighted the issue of PLC cases being coded as 'unspecified liver cancer' due to missing morphology code. That study reported that, following a review of the medical records, 75.9% of 'unspecified liver cancer' cases were subsequently recoded as HCC. A limitation of our study was that unlike the TCR we did not have access to the medical records from public hospitals in Tasmania or results from pathology, hospital laboratories or imaging departments. Thus, we could not draw strong conclusions regarding the accuracy of coded causes of death data from TCR.

Our results using the competing risks framework did not identify sex or remoteness to have an impact on the cumulative incidence of death. We did, however, observe that the cumulative incidence of death was highest in cases with unspecified liver cancer. This potentially reflects clinical practice given that patients diagnosed at an advanced stage of PLC may have more undifferentiated imaging characteristics, and unlike people treated with curative intent, are less likely to obtain a histological diagnosis of PLC subtype.

In contrast to our results, a US study reported cumulative incidence of breast cancer deaths was overestimated due to misclassification of the cause of death,[22] but comparisons of mortality between breast cancer and PLC using a competing risk framework require careful consideration. Breast cancer is associated with substantially longer survival in both the US and Australia (5 year survival 91.3%[23] and 90.8%[24] respectively) compared to PLC (19.2%[23] and 18.5%,[2] respectively). Additionally, breast cancer patients are generally younger than PLC patients, and the highest incidence of breast cancer is between 70 and 74 years of age and for PLC patients in the range 80-84.[3] Also, it carries a higher risk of death from other causes of death in comparison to PLC cases. In our study, the majority of PLC cases died from the event of interest (liver cancer) in all three datasets (89.2% for the TCR, 78.0% for the ABS, and 74.6% based on the consensus from medical practitioners) whereas in the breast cancer study, the number of deaths due to the event of interest (breast cancer) represented a small proportion of all deaths (12.2%).[22] These different results suggest that the cumulative incidence function may not show substantial impacts on cause-specific survival for cancers with short survival time such as PLC. However, the methodology is likely to be applicable to high-survival cancers in Australia such as colorectal and breast cancer, to examine the impact of different coded causes of death on estimates of cause-specific survival.

Strengths and Limitations

The use of linked health data is useful as a means of studying many health conditions, outcomes and service provision in a cost-effective manner.[26] In our study, the TCR provided access to quality data, based on using probabilistic linkage methods to reduce potential errors.[15] The cumulative incidence function was used to estimate deaths caused by non-cancer events in the

presence of competing risks. This method is preferred to the Kaplan-Meier and net survival methods, which usually overestimate the absolute risk of cause-specific survival and ignore competing causes.[17,18,27] The competing risk methods provided the appropriate framework to analyse the interplay between deaths from PLC and other competing risk factors based on different coding practices.[18,27,28]

A limitation of this study was that we could not draw strong conclusions regarding the accuracy of cause of death data from the TCR or ABS. Access to individual medical records would have been required for this. Future research should focus on better understanding of coding practices and the accuracy of the different approaches. Cause-specific survival estimates have been shown to provide lower survival rates compared to all-cause estimates for breast cancer.[29] In this context, robust cause of death data will support more accurate cause-specific survival estimates for patients and clinicians. Whilst this may not be as relevant for low survival cancers such as liver PLC, this may be particularly useful for cancers with longer survival and earlier age at diagnosis such as breast cancer.

In addition, whilst we observed strong agreement between the medical practitioners, our study design limited these reviews to cases in which a discrepancy was observed between the ABS and TCR. Therefore, we cannot comment on the level of agreement between medical practitioners for cases in which no discrepancy between the ABS and TCR was observed.

Conclusions

This is the first study to evaluate the impact of different coded causes of death on estimates of cause-specific survival for PLC. The cumulative incidence of death was similar across the TCR, ABS, and medical practitioners, but differed according to the type of PLC. As PLC is a low-survival cancer, such results may be different to cancers with better survival such as breast cancer. Coding causes of death is a complex process, with all systems having some level of inaccuracy. However, greater levels of harmonisation should be pursued, with the aim of providing the most accurate data as possible. Utilisation of specialist clinician oversight might improve data cohesion and fidelity.

Corresponding Author: Dr Barbara de Graaff, Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania 7000, Australia
Email: barbara.degraaff@utas.edu.au

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 2018; 68(6): 394-424.
2. Australian Institute of Health and Welfare. Colorectal and other digestive tract cancers. Canberra, 2018.
3. Australian Institute of Health and Welfare (AIHW) 2018 Cancer Data in Australia. Australia Cancer Incidence and Mortality (ACIM) books: liver cancer. Canberra, 2018.
4. Cocker F, Chien Yee K, Palmer AJ, de Graaff B. Increasing incidence and mortality related to liver cancer in Australia: time to turn the tide. Aust N Z J Public Health 2019; 43(3): 267-73.
5. Australian Institute of Health and Welfare. Deaths data Australian Institute of Health and Welfare 2018. Available from: https://www.aihw.gov.au/about-our-data/our-data-collections/national-mortality-database/deaths-data.
6. Tasmania Government. The Tasmanian Registry of Births, Deaths and Marriages 2020 [cited 2020 10 June]. Available from: https://www.justice.tas.gov.au/
7. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 2010.
8. World Health Organization. International Classification of Diseases for Oncology Edition T, 2013.
9. McGivern L, Shulman L, Carney JK, Shapiro S, Bundock E. Death certification errors and the effect on mortality statistics. Public Hlth Rep 2017; 132(6): 669-75.
10. Schaffar R, Rapiti E, Rachet B, Woods L. Accuracy of cause of death data routinely recorded in a population-based cancer registry: impact on cause-specific survival and validation using the Geneva Cancer Registry. BMC cancer 2013; 13(1): 609.
11. Filippatos G, Andriopoulos P, Panoutsopoulos G, Zyga S, Souliotis K, Gennimata V et al. The quality of death certification practice in Greece. Hippokratia 2016; 20(1): 19.
12. Middleton D, Anderson R, Billingsly T, Virgil NBM, Wimberly Y, Lee R. Death certification: issues and interventions. Open J Prev Med 2011; 1(3): 167-70.
13. Yin D, Morris CR, Bates JH, German RR. Effect of misclassified underlying cause of death on survival estimates of colon and rectal cancer. J Nat Cancer Inst 2011; 103(14): 1130-3.
14. Australian Bureau of Statistics. Causes of Death, Australia: Doctor Certified Deaths, Summary Tables: Australian Bureau of Statistics; 2020 Available from: https://www.abs.gov.au/statistics/health/causes-death/causes-death-australia-doctor-certified-deaths-summary-tables/latest-release#doctor-certified-mortality-data
15. Tasmanian Data Linkage Unit. What is Data Linkage? 2021 [15/02/2022]. Available from: https://www.menzies.utas.edu.au/research/research-centres/data-linkage-unit/what-is-data-linkage
16. McHugh ML. Interrater reliability: the Kappa statistic. Biochemia Medica 2012; 22(3): 276-82.

17. Dutz A, Löck S. Competing risks in survival data analysis. Radiotherapy and Oncology 2019; 130: 185-9.

18. Mariotto AB, Noone A-M, Howlader N, Cho H, Keel GE, Garshell J *et al*. Cancer survival: an overview of measures, uses, and interpretation. J Nat Cancer Inst Monographs 2014; 2014(49): 145-86.

19. Australian Bureau of Statistics. Australian Statistical Geography Standard (ASGS): Volume 4 - Significant Urban Areas, Urban Centres and Localities, Section of State. Canberra 2016.

20. Australian Bureau of Statistics. Standard Australian Classification of Countries. Canberra 2016.

21. Hong TP, Gow P, Fink M, Dev A, Roberts S *et al*. Novel population-based study finding higher than reported hepatocellular carcinoma incidence suggests an updated approach is needed. Hepatology 2016; 63(4): 1205-12.

22. Hinchliffe SR, Abrams KR, Lambert PC. The impact of under and over-recording of cancer on death certificates in a competing risks analysis: a simulation study. Cancer Epidemiol 2013; 37(1): 11-9.

23. American Cancer Society (2020). Cancer Facts & Figures 2020. Available from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf

24. Australian Institute of Health and Welfare. Cancer in Australia 2019. Canberra, 2019.

25. Australian Institute of Health and Welfare. Cancer Data in Australia. Canberra, 2020.

26. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. Ann Rev Public Hlth 2011; 32: 91-108.

27. Howlader N, Mariotto AB, Woloshin S, Schwartz LM. Providing clinicians and patients with actual prognosis: cancer in the context of competing causes of death. J Nat Cancer Inst Monographs 2014; 2014(49): 255-64.

28. Lambert P, Dickman P, Nelson C, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. Statistics in Medicine 2010; 29(7-8): 885-95.

29. Howlader N, Ries LA, Mariotto AB, Reichman ME, Ruhl J, Cronin KA. Improved estimates of cancer-specific survival rates from population-based data. J Nat Cancer Inst 2010; 102(20): 1584-98.