Estimating the change in life expectancy after a diagnosis of cancer among the Australian population

Peter D Baade,1,2,3 Danny R Youlden,1 Therese M-L Andersson,4 Philippa H Youl,1,2,3 Michael G Kimlin,1,2,5 Joanne F Aitken,1,2 Robert J Biggar2,5

ABSTRACT

Objectives: Communication of relevant prognostic information is critical in helping patients understand the implications of their cancer diagnosis. We describe measures of prognosis to help communicate relevant prognostic information to improve patients’ understanding of the implications of their cancer diagnosis.

Setting: Australia-wide population-based cancer registry cohort.

Participants: 870 878 patients aged 15–89 years diagnosed with invasive cancer between 1990 and 2007, with mortality follow-up information to December 2010.

Primary and secondary outcome measures: Flexible parametric models were used to estimate loss of life expectancy (LOLE), remaining life expectancy (RLE) and 10-year cumulative probability of cancer-specific death (1-relative survival).

Results: On average, Australians diagnosed with cancer at age 40 years faced losing an average of 11.2 years of life (95% CI 11.1 to 11.4) due to their cancer, while those diagnosed at 80 years faced losing less, an average of 3.9 years (3.9 to 4.0) because of higher competing mortality risks. In contrast, younger people had lower estimated cumulative probabilities of cancer-specific death within 10 years (40 years: 21.5%, 21.4% to 22.1%) compared with older people (80 years: 55.4%, 55.0% to 55.9%). The patterns for cancer-specific death within 10 years (40 years: 21.5%, 21.4% to 22.1%) compared with older people (80 years: 55.4%, 55.0% to 55.9%). The patterns for individual cancers varied widely, both by cancer type and by age within cancer type.

Conclusions: The LOLE and RLE measures provide complementary messages to standard relative survival estimates (expressed here in terms of cumulative probability of cancer-specific death). Importantly, relative survival per se underplays the cumulative probability of cancer-specific death.

Strengths and limitations of this study

- Large population-based national cohort of patients with cancer with up to 21 years follow-up.
- Use of flexible parametric models to more easily capture the shape of the underlying hazards function and extrapolate future survival estimates.
- No additional information about clinical characteristics and management strategies.
- The use of population mortality rates may be less valid for some patients with cancer due to shared risk factors for other causes of deaths.

INTRODUCTION

As the number of people diagnosed with cancer increases and their average survival lengthens,1 improving the quality and relevance of survival information at the time of diagnosis and during follow-up is critical. While the majority of patients with cancer want information on their survival prospects and life expectancy,2,3 conveying this information can be challenging. For example, most patients have difficulty in understanding probabilities.4 Access to a wider range of measures regarding prognosis may assist in tailoring the communication of mortality risk to individual patients.

Various methods of expressing net survival outcomes have been used, including relative5 and cause-specific survival.6,7 Recently, conditional survival has been reported,8–10 which is the expected survival after a patient has already lived for a certain time. While each of these measures of survival is important, there are limitations in the application of probabilities or percentages. For example, a 10-year survival estimate of 75% may mean little to someone without an understanding of their expected lifespan. In addition, all measures of net survival reflect the fairly uninformative probability (to the patient) of surviving assuming the cancer under consideration is the only cause of death.
Another indicator of prognosis is the impact that a cancer diagnosis has on life expectancy,11 expressed either as the difference between the life expectancy among the general population and that of people diagnosed with cancer (loss of life expectancy or LOLE) or the remaining life expectancy (RLE) among patients with cancer. These approaches address the question “On average, how much does my life expectancy change now that I have been diagnosed with cancer?” While this measure has intuitive appeal, its widespread use has been limited by the lack of suitable methodology and software to extrapolate the observed survival curve beyond the available follow-up period.11 With the recent development of new statistical methods,11 these estimates of LOLE and RLE have potential for wider dissemination and clinical application.

In this article, we present population-based estimates of LOLE and RLE due to cancer for patients diagnosed in Australia, alongside a measure representing the standard relative survival estimates.

METHODS
Cancer cohort
Approval for the use of the data was provided by all of the individual state and territory cancer registries through the Australian Institute of Health and Welfare (AIHW), with the exception of the Australian Capital Territory (ACT). Residents of the ACT account for about 1.2% of all cancers diagnosed in Australia.1 No external funding was obtained for this study.

Notification of all invasive cancers (excluding keratinocyte skin cancers) to Australian state and territory cancer registries is required by law. De-identified case records for the whole of Australia (excluding ACT) were obtained from the AIHW, to which the population state and territory cancer registries supply data on an annual basis.12 All patients with cancer diagnosed from 1990 to 2007 were included in the cancer cohort, with follow-up of mortality status to 31 December 2010. Variables provided in the data extract were unique identifier, month and year of diagnosis, month and year of death (if applicable), days between diagnosis and either death or date of last follow-up (whichever came first), sex, age at diagnosis, type of cancer, and basis of diagnosis. We restricted our cancer cohort to patients aged 15–89 years at diagnosis due to different cancer classifications being used for children13 and for consistency with our recent conditional survival analysis.8 Mortality status was obtained through the routine annual linkage of cancer records with the National Death Index in Australia. The incidence data up to 2007 were the most recent to have been matched to the National Death Index; therefore, patients diagnosed after 2007 would be likely to have more incomplete mortality information.

In addition to all invasive cancers combined (International Classification of Diseases (ICD)-O-3 codes C00-C80), we restricted the cancer-specific analyses to stomach cancer (C16), colorectal cancer (C18-C20, C218), lung cancer (C33-C34), melanoma of the skin (C44, M872-M879), breast cancer among females (C50) and prostate cancer (C61). These major types of cancer also serve to illustrate a range of survival outcomes. Cases diagnosed on the basis of death certificate only or when the recorded date of diagnosis was after the date of death were excluded from the cohort.

Statistical methods
At-risk period
Since the mortality follow-up period extended beyond the diagnostic study period, we applied the hybrid period survival method described by Brenner and Rachet.14 Under this approach, members of the cancer cohort were considered to be “at risk” of dying during the 5-year period between 1 January 2006 and 31 December 2010. However, as the follow-up period extends beyond the period from which incident cases are accrued, there are fewer years of cases available for the period 2008–2010 during the first 3 years of follow-up. The “at-risk” window is, therefore, altered to ensure that each follow-up interval consists of 5 years of cases. For the first year of follow-up, we extended the “at-risk” window to 2003–2007, for the second year to 2004–2008, for the third year to 2005–2009 and then to 2006–2010 for subsequent years of follow-up. For those who died before 31 December 2010, observed survival time was measured from the date of diagnosis to the date of death, with survival being censored at 31 December 2010 for those patients still alive at that date.

Relative survival
We used relative survival to approximate disease-specific survival because it does not require any information on the cause of death.5 Relative survival is calculated by dividing the observed probability of all-cause survival among individuals diagnosed with a specific cancer by the expected probability of survival within the age-sex-year-matched general population of Australia. Expected survival was obtained from the Ederer II method15 using Australian population life tables. To allow a more direct comparison with RLE and LOLE, we have reported the cumulative probability of death, which is equal to one minus the relative survival.

LOLE and RLE
The theoretical concepts of LOLE and RLE are illustrated in figure 1. The RLE for a hypothetical cohort of patients with cancer is represented by the area under the observed survival curve (lower red solid line), with survival time on the x-axis. The area under the expected survival curve (upper green dashed line) represents the life expectancy of the age-sex-year-matched general population. The difference between these two measures, that is, the blue shaded area between the two curves, represents the LOLE from the time of diagnosis onwards.
An important difference between relative survival and either RLE or LOLE is that relative survival estimates compare the observed and expected survival at only one point on the survival curves (according to time after diagnosis). In Figure 1, 10-year relative survival is calculated by dividing the value at point B by the value at point A. In contrast, RLE and LOLE consider the differences between all points on the two survival curves, thus using a greater range of information, including the length of the survival curve, in order to provide a more complete picture of the changed mortality risks experienced by those diagnosed with cancer.

Flexible parametric models
Owing to the limited time period covered by Australian (and most international) population-based cancer registries, calculating the LOLE requires the extrapolation of both the expected and observed survival curves beyond the available follow-up period. To address this concern, we used methods developed by Andersson et al which used flexible parametric models to model the observed survival curves and then extrapolated these to estimate the survival curves for additional follow-up intervals.

Flexible parametric survival models use restricted cubic splines for the baseline and can, therefore, more readily capture the shape of the underlying hazard function compared with more traditional methods, such as the Cox proportional hazards model. We imposed 7 degrees of freedom (or 6 internal ‘knots’) for the baseline function and 4 degrees of freedom (3 internal ‘knots’) for the time-varying effects, and used restricted cubic splines with 4 degrees of freedom and a time-varying component to model the effect of age.

Extrapolation of the observed curves using these models requires making assumptions for what happens beyond the last observed follow-up interval. A previous study has demonstrated that assuming a linear trend in the log cumulative excess hazard fitted the observed survival well; therefore, for simplicity we followed the same approach with our models.

Estimates for relative survival, cumulative probability of death, RLE and LOLE were predicted from the flexible parametric models for single years of age and then reported for each cancer type at 40, 50, 60, 70 and 80 years of age. Median follow-up time was calculated using the reverse Kaplan-Meier method. The proportion of life expectancy lost due to the cancer diagnosis was calculated by dividing the LOLE by the population life expectancy.

All statistical analyses were performed using Stata/SE V.12.1 for Windows (StataCorp, Texas, USA). Flexible parametric survival models were fitted using the stpm2 package.

RESULTS
A total of 1 474 744 Australians were diagnosed with cancer between 1990 and 2007. Of these 359 (0.02%) were excluded due to missing or negative follow-up times and 14 709 (1%) were excluded due to the diagnosis being based on death certificate only. Of the remaining patients, 870 878 (59%) were alive at some point during the at-risk period and so formed the final cohort for this study. Key characteristics of these patients with cancer are shown in Table 1. Breast and prostate cancer accounted for the largest proportion of cases (16% each) followed by melanoma (14%) and colorectal cancer (13%). The median age at diagnosis was 65–69 years for most specified types of cancer but was about 10 years younger for melanoma and breast cancer. Not considering the sex-specific breast and prostate cancers, the majority of cases for each type of cancer were diagnosed among males. The median time of follow-up among the cohort ranged from 5.7 to 6.3 years for the different cancer types. The proportion of patients who died during the ‘at-risk’ period ranged from 14% for melanomas to 77% for lung cancer.
The effect that age had on the 10-year cumulative probability of death varied by cancer type. For all cancers combined, lung cancer, melanoma and, to a lesser extent, stomach cancer, the curve increased with age (table 2 and figure 2). In contrast, breast cancer and prostate cancer showed a U-shaped curve in mortality, while the curve for colorectal cancer was almost flat until it reached the older age groups.

Conversely, the effect that age had on LOLE was substantially more pronounced and consistent. Australians diagnosed with cancer at younger ages had higher LOLE estimates than those diagnosed at an older age (table 2 and figure 3). LOLE depends on prognosis and, more importantly, on population life expectancy, which reduces with increasing age (table 2). To the extent that both change with age, LOLE patterns will be impacted. Using lung cancer as an example, a person diagnosed when at 40 years of age (10 years probability of death: 78.2%) faced losing on average 33 years of life expectancy (RLE=10 years), while a person aged 80 years at diagnosis (10 years probability of death: 92.0%) faced losing 7 years of life expectancy (RLE=1 year). When we calculated the LOLE as a proportion of population life expectancy, it increased steadily with age for all cancers combined (table 2). However, when examined within specific cancer types, proportional life expectancy was much more stable for each cancer type, indicating that the pattern seen with all cancer types resulted from the change in the mix of cancers seen at different ages. Differences between individual cancers by age reflected variation in prognosis by type of cancer, with modest variations by age within type (table 2).

**DISCUSSION**

The communication of relevant prognostic information is critical in helping patients understand the implications of their cancer diagnosis. Most patients with cancer wish to have accurate information regarding their prognosis and life expectancy delivered in a way that is relevant to them and easy to understand. By applying emerging statistical models, we have demonstrated the complementary messages provided by standard relative survival statistics, and LOLE or RLE information. Importantly, relative survival tends to underplay the greater impact that a cancer diagnosis has at a younger age and, when presented in isolation, may provide a misleading impression of future mortality burden.

The measures of LOLE and RLE, which are expressed in years, place greater importance on cancers diagnosed in the early-middle years of life than relative survival. This focus provides a better quantitative understanding of the impact that a cancer has on the subsequent lives of patients and the overall cancer burden. With some exceptions (breast and prostate in this study), relative survival estimates suggest that people diagnosed with cancer do equally well between 40 and 70 years of age, although people diagnosed at older ages tend to do worse. However, by incorporating the implications on life expectancy, the LOLE and RLE measures provide an important different perspective, highlighting the greater impact of a cancer diagnosis on the absolute LOLE among younger people.

In spite of these new measures, relative survival retains a high importance when disseminating information about the burden of cancer. Relative survival (or cumulative probability of death) compares the survival in different ages in the hypothetical scenario that cancer was the only possible cause of death (under some assumptions). It is, therefore, the best way of making comparisons in relation to age equity in quality of cancer care that are free of differences due to other causes of mortality.

There are important differences in LOLE and the contextually similar years of life lost (YLL) which is often used to quantify premature mortality. First, YLL requires accurate cause of death information, which is avoided in LOLE since it is based on relative survival. More importantly, YLL is based only on people dying from the condition within a specific time period, irrespective of when they were diagnosed, thus ignoring the experience of the entire cohort diagnosed with the condition. In contrast, LOLE focuses on a defined cancer cohort and estimates the loss in expectation of life for that cohort, or for recently diagnosed patients if a (hybrid) period approach is used, irrespective of when or if they die due to cancer.
While relative survival and LOLE present different pictures in terms of age-specific impacts, comparisons of the mortality burden across cancer types show similar case histories. Cancers with a low relative survival (and high cumulative probability of death) have consistently higher LOLE. For example, the age-specific LOLE for melanoma (high survival, low cumulative probability of death) was much lower than the corresponding estimates for lung cancer and stomach cancer (low survival, high cumulative probability of death).

Both relative survival and LOLE are based on comparing the observed mortality among the cancer cohort with the population mortality and then attributing any differences to the diagnosed cancer. This may not always be the case, because people diagnosed with some cancers typically have a higher prevalence of risk factors.
(such as smoking in patients with lung cancer) and comorbidities. However, as this limitation applies to both measures it would have minimal impact on any comparison between the two. The impact of factors, such as comorbidities, could be assessed for the RLE and LOLE if information on comorbidities was available for the patients with cancer, and if population mortality rates stratified on comorbidities were available.

We have purposefully used the hybrid period survival method to focus on the survival predictions for recently diagnosed patients, while also accessing the longer follow-up of people diagnosed during the 1990s to provide information about deaths in longer term survival cases; this assists in providing more certainty for the extrapolation of the survival curves. This is different to the cohort survival approach, in which the survival experience of people diagnosed during the 1990s would contribute to both short-term and long-term survival. Since survival for most cancers has increased over time,1 the long-term survival of patients diagnosed recently may be higher than those diagnosed in the early 1990s; in other words, these LOLE estimates will overstate the true values for recently diagnosed patients if this cohort method is used.

However, these results still rely on extrapolated information; therefore, caution is needed in case this does not reflect the observed results 30 or 40 years post-diagnosis. It is also important to keep in mind that the CIs presented for the LOLE and RLE only take uncertainty in the model parameters into account and not uncertainty in the extrapolation or model selection. In their original development of this method, Andersson et al11 compared predicted survival outcomes for Swedish patients diagnosed in the 1960s against actual survival up to 2010 and found that the components of the flexible parametric method were robust, particularly when...
at least some of the cohort had more than 10 years of follow-up, as is the case in our study. Finally, these estimates are based on the averaging of population data and will vary from patient to patient depending on their specific circumstances. The same limitation applies with routinely published relative survival estimates. Further work is needed to examine the effect of clinical characteristics on LOLE and RLE, particularly stage at diagnosis, and how these measures change the longer the patient survives their initial diagnosis (analogous to conditional survival estimates). In addition, there is potential to investigate the changing economic impact of LOLE among younger people due to cancer for their surviving dependents in light of the advancement of the retirement age in Australia.

In conclusion, given the increasing numbers of people being diagnosed with cancer and the improving survival outcomes for most types of cancer, providing patients with relevant risk information is paramount. LOLE and RLE are an important addition to existing measures that give a more complete picture of the impact of a cancer diagnosis.

Contributors PDB and RJB were involved in conception of this work. PDB, RJB and TM-LA were involved in design of this work. PDB was involved in acquisition of data, analysis and drafting the manuscript. PDB, DRY, TM-LA, PHY, MGK, JFA and RJB were involved in interpretation of data, revising the manuscript and final approval of the published version.

Funding PDB was supported by an Australian National Health and Medical Research Council Career Development Fellowship (1005334). PHY is funded by a National Health and Medical Research Council Early Career Fellowship (1054038).

Competing interests None declared.

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
15. Ederer F, Heise H. Instructions to IBM 650 programmers in processing survival computations. Methodological Note No. 10, End Results Section, Bethesda, MD: National Cancer Institute, 1959.
17. Royston P, Lambert PC. Flexible parametric survival analysis using Stata: beyond the Cox model. College Station, TX: Stata Press, 2011.