JIM Review

doi: 10.1111/joim.13139

Cancer survival statistics for patients and healthcare professionals – a tutorial of real-world data analysis

S. Eloranta[1], K. E. Smedby[1,2], P. W. Dickman[3] & T. M. Andersson[3]

From the [1]Department of Medicine, Division of Clinical Epidemiology, Karolinska University Hospital,Karolinska Institutet, Stockholm, Sweden; [2]Department of Medicine, Division of Hematology, Karolinska University Hospital,Karolinska Institutet, Stockholm, Sweden; and [3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Abstract. Eloranta S, Smedby KE, Dickman PW, Andersson TM (Karolinska University Hospital, Karolinska Institutet; Karolinska Institutet, Stockholm, Sweden). Cancer survival statistics for patients and healthcare professionals – a tutorial of real-world data analysis (Review). *J Intern Med* 2021; **289**: 12–28. https://doi.org/10.1111/joim.13139

Monitoring survival of cancer patients using data collected by population-based cancer registries is an important component of cancer control. In this setting, patient survival is often summarized using net survival, that is survival from cancer if there were no other possible causes of death. Although net survival is the gold standard for comparing survival between groups or over time, it is less relevant for understanding the anticipated real-world prognosis of patients. In this review, we explain statistical concepts targeted towards patients, clinicians and healthcare professionals that summarize cancer patient survival under the assumption that other causes of death exist. Specifically, we explain the appropriate use, interpretation and assumptions behind statistical methods for competing risks, loss in life expectancy due to cancer and conditional survival. These concepts are relevant when producing statistics for risk communication between physicians and patients, planning for use of healthcare resources, or other applications when consideration of both cancer outcomes and the competing risks of death is required. To reinforce the concepts, we use Swedish population-based data of patients diagnosed with cancer of the breast, prostate, colon and chronic myeloid leukaemia. We conclude that when choosing between summary measures of survival it is critical to characterize the purpose of the study and to determine the nature of the hypothesis under investigation. The choice of terminology and style of reporting should be carefully adapted to the target audience and may range from summaries for specialist readers of scientific publications to interactive online tools aimed towards lay persons.

Keywords: cancer, epidemiology, biostatistics, death risk.

Introduction

On the surface, estimates of cancer patient survival, the proportion of patients who survive a given period of time subsequent to diagnosis, are easy to calculate and communicate. There is, however, considerable nuance in how such measures can be interpreted and in what context they are of relevance. Net survival, the perhaps most common measure used to report survival associated with cancer, is, for example, interpreted in the hypothetical situation where the cancer under investigation is the only possible disease that can kill the patients [1-3]. In this review, we will explain why this is a practical measure in many situations, but also describe other measures of patient survival that have a valid interpretation in the real world. The latter include estimates of survival in the presence of competing risks and loss in life expectancy, measures that illustrate cancer patient survival in a clinically meaningful manner. These have become increasingly popular amongst medical researchers but confusion sometimes remains with respect to assumptions underlying the statistics and their correct interpretation. To reinforce the theoretical aspects related to each concept, we will demonstrate their appropriate use through examples from national register-based cancer data from Sweden.

The role of cancer patient survival in cancer control

Our focus throughout this tutorial will be on cancer patient survival estimated from data

collected by population-based cancer registries. Population-based cancer survival is a key measure of the overall effectiveness of health systems in management of cancer, and descriptive studies that compare survival between countries and regions, or between population groups defined by socio-economic status, race or ethnicity, or health insurance are being used to drive improvement in health services [4, 5]. The International Cancer Benchmarking Partnership [6, 7], CONCORD [8-10] and the EuroCare [11] projects are examples of ongoing international collaborations that report net survival and that aim to provide an evidence base for global cancer control and international comparisons of the effectiveness of healthcare systems that can serve as a basis for policy and practice changes towards reducing the burden of cancer.

The statistical measures used in cancer control research typically focus on the net effect of cancer on patient survival, meaning that they lack a real-world interpretation as deaths unrelated to cancer are assumed to not exist [2, 12]. Whilst this assumption may seem odd at first glance, it is precisely the quantity of most relevance when comparing groups with different noncancer mortality. Assume for example that our interest is in comparing colon cancer survival across countries with large demographic differences (e.g. age, income-level and education). Because the demographic differences are known to be associated with life expectancy, incorporating noncancer mortality in the estimation would yield unfair comparisons of cancer survival. To isolate the effect of cancer on the risk of dying in each country, deaths due to other causes are considered a nuisance that should be 'eliminated' in the estimation process [13]. In practice, this elimination process is done via censoring in the survival analysis [14-16].

What statistics are useful for healthcare professionals and patients?

Throughout the review, we will focus on statistical measures that are useful for healthcare professionals and patients in the real-world setting. In this context, we mean measures that contribute to the understanding of the anticipated prognosis for a cancer patient with certain clinical characteristics and that provide results that can be used, for example in risk communication between doctor and patient, or in resource allocation and operational planning [17]. Cancer control research that aims to identify or evaluate strategies to prevent and detect cancer, or to improve quality of life amongst affected patients, is without doubt also useful and have great impact for healthcare professionals and patients. However, with respect to their potential value for risk communication in the clinical setting, a distinction between survival measures designed for the hypothetical (net survival) and the real world is necessary [18, 19].

First, we will present key measures for estimating survival in the presence and absence of competing risks, explain the assumptions that are required in the estimation and how results are interpreted. Next, we introduce loss in life expectancy which is an appealing approach to quantifying the impact of cancer on the remaining lifetime with clear merits for communication with nonspecialists. We then explain the concept of conditional survival and discuss for whom these measures are relevant. Throughout the review, we use population-based data from the national cancer register in Sweden for patients diagnosed with prostate cancer, colon cancer, breast cancer and chronic myeloid leukaemia (CML), as illustrative examples and to reinforce the theory.

Statistical methods used in cancer patient survival studies

The hypothetical world—net survival

Net survival is the probability of being alive at a certain time following diagnosis in the hypothetical scenario where the cancer of interest is the only possible cause of death. Two common frameworks for estimating net survival are cause-specific survival and relative survival. In the cause-specific framework, each death must be classified as either (a) death due to cancer, or (b) something other than cancer. This classification is typically made based on routinely recorded information in cause of death registers or medical records. In the statistical analysis, cancer deaths are denoted as events and deaths due to any other causes are treated as censored observations [15] and standard methods (e.g. Kaplan–Meier) can be applied to estimate cause-specific survival.

Relative survival provides an alternative to cause-specific survival that does not require cause of death information. Whilst several estimation methods for net survival in a relative survival framework have been suggested [12, 20, 21], a common way is to calculate the ratio between the observed all-cause survival amongst the cancer patients, and the expected survival in a comparable group in the

Fig. 1 *Comparison of cause-specific survival estimated using the Kaplan–Meier method and relative survival for men with colon- and prostate cancer diagnosed in the year 2012.*

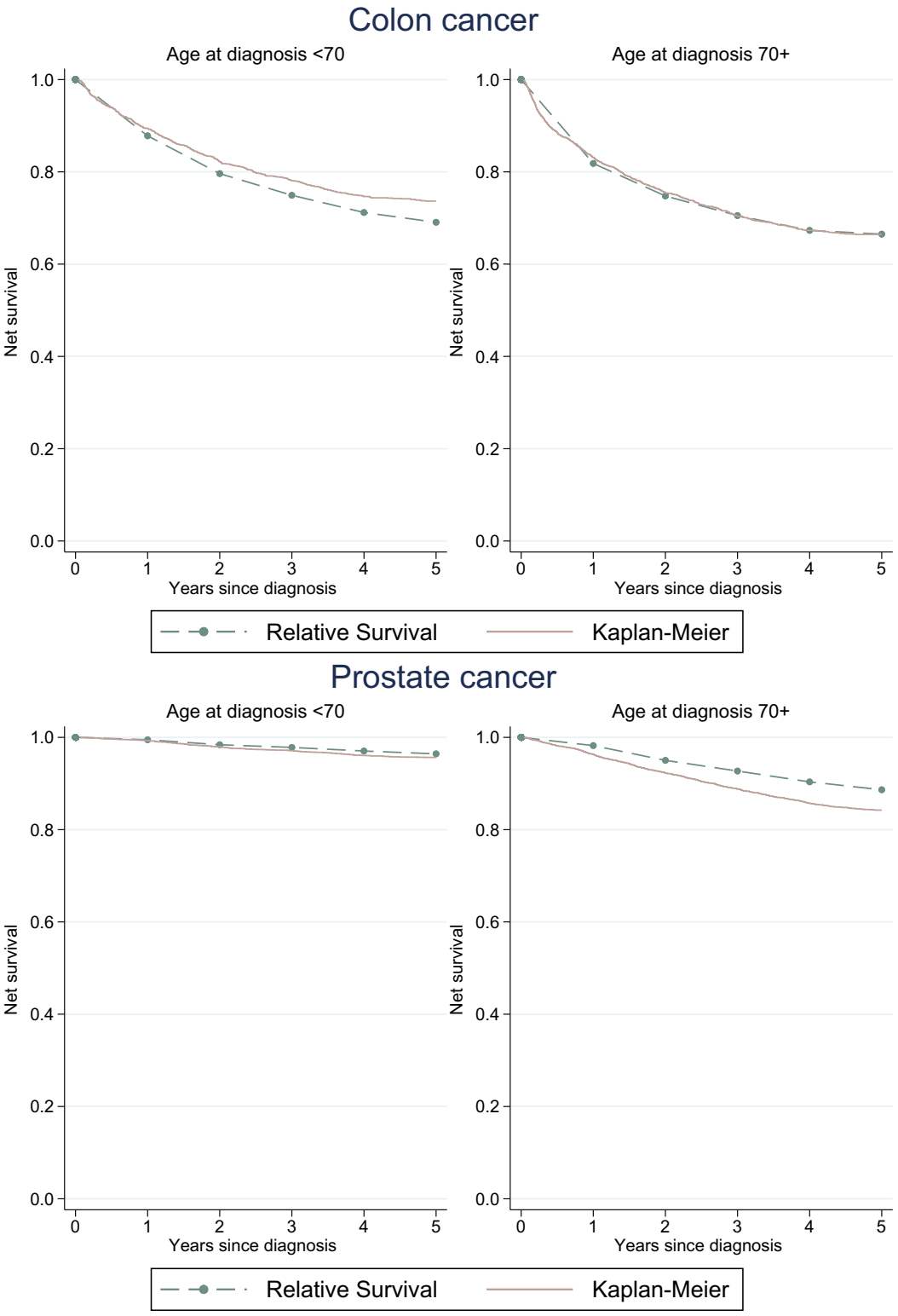
general population. If, for example, the one-year relative survival ratio is 1.0, the survival proportion in the group of cancer patients coincides with that in the general population, suggesting that there is no *excess mortality* associated with the cancer during the first year of follow-up. A relative survival ratio lower than 1.0 means that cancer patients have inferior survival compared to the general population, and we typically assume that the survival deficit is due to excess mortality from the cancer. As a consequence, the excess mortality thus captures both the direct and indirect mortality associated with the cancer. The mortality that is directly related to the cancer includes all causes of death that would typically be classified as death due to cancer on the death certificate. Indirect mortality due to cancer encompasses death due to treatment toxicity, suicides and late effects such as cardiovascular disease, second malignancies or infections. These are events that are unlikely to be classified as cancer deaths on the death certificate and that would therefore not be captured in a cause-specific analysis. As cause of death information is not always readily available or reliable, relative survival has become the preferred approach for obtaining estimates of cancer patient survival in population-based investigations, as well as by cancer registries worldwide.

Key assumptions for cause-specific and relative survival

For comparison of the two methods, Fig. 1 shows Kaplan–Meier estimates of cancer-specific survival, as well as relative survival for patients diagnosed with colon cancer (upper panel) and prostate cancer (bottom panel) in Sweden in 2012. Survival up to five years for two different age groups (<70 years and 70+ years at diagnosis) has been plotted for each cancer type and estimation method. From the upper part of the figure, we see that five years after a colon cancer diagnosis, men who belong to the oldest age group have a net survival of 66%. That is, if colon cancer was the only thing that could kill the patients 66% of all men in this age group would still be alive five years after the diagnosis. Whilst the five-year Kaplan–Meier and relative survival aim to estimate the same underlying quantity, the estimates are nevertheless not identical. There are for example slight differences in the estimates for <70-year-old colon cancer patients and substantial differences for the 70+ year-old men with prostate cancer (lower part of the figure). The observed differences are related to the statistical assumptions that each of the methods requires. Whilst the Kaplan–Meier method assumes accurate cause of death classification, it is not always clear from routinely recorded cause of death information if a death is attributable to the cancer or not. In our colon cancer example, the survival estimated by the Kaplan–Meier method is slightly higher than the relative survival for the youngest age group. A possible explanation for the small discrepancy could be the contribution of postoperative mortality or treatment toxicity, factors that are likely to contribute to indirect mortality and thus to reduced relative survival.

For prostate cancer, cause-specific survival is instead lower than the relative survival. The difference is particularly pronounced for men who are 70+ years at diagnosis. One possible mechanism explaining the difference could be that older men with a diagnosis of prostate [22, 23] cancer are likely to have prostate cancer recorded on their death certificates even when prostate cancer was not the leading cause of death [24]. Whilst this is probably the main reason for the observed difference in this example, another possible explanation is related to the so called exchangeability assumption for relative survival. This assumption essentially means that had the patients not been diagnosed with cancer, they should be comparable to the general population with respect to noncancer mortality. Men with localized prostate cancer have been shown to have lower mortality from noncancer causes than men of the same age in the general population which might reflect a more health conscious lifestyle [25]. Using the general population for comparison of all-cause survival will therefore likely underestimate the excess mortality associated with prostate cancer, or conversely, relative survival will be overestimated. Stratifying the population life tables on additional factors (e.g. education, social class) can reduce the magnitude of the problem but will not eliminate it [22, 23].

A third assumption that is applicable to both methods is that of independence between cancer deaths and other causes of deaths. For this



assumption to be satisfied there can be no shared factors associated with the two types of deaths other than those factors already adjusted for in the analysis. For example, age is associated with survival for many types of cancer. Age is also associated with survival from causes unrelated to cancer, such as cardiovascular disease, and should therefore be adjusted for in the analysis in applications where net survival is the measure of interest. Examples of other factors that are typically discussed in the context of the independence assumption and controlled for in the analysis are sex, calendar period and socio-economic factors.

The real world—survival in the presence of competing risks

The assumption that patients are immune from competing causes of death in cause-specific and relative survival analysis is often not understood and can yield misleading conclusions if the approach to estimation is inappropriate for the study context and hypothesis [26, 27]. For patients and healthcare professionals who seek information to understand or to communicate the prognosis associated with a specific cancer, measures that acknowledge the concomitant risk for competing events are more informative than net survival [28-30]. A competing event is any event that occurs during follow-up and prevents or substantially alters the outcome of primary interest to occur. For example, if the primary outcome of the study is death due to prostate cancer, and a patient dies in a traffic accident, the death due to the accident is a competing event. The difference between cancer-specific survival estimated in the absence, versus the presence, of competing events can be clearly demonstrated for diseases that have an indolent course and where deaths due to other causes are common amongst the cancer survivors [17, 26]. Figure 2 shows the probability of death within the first 10 years of diagnosis amongst men diagnosed with prostate cancer at age 75 years in 2007. The net probability of death (dashed line) is simply the complement to the net survival probability (i.e. 1—net survival). The shaded areas represent the probabilities of death due to prostate cancer in the presence of death due to other causes. Results from competing risks analyses are often presented in this manner so that the risk of death due to the cancer can be assessed in relation to the competing deaths as well as to the chance of still being alive (white area). Thus, for a 75-year-old man with prostate cancer, the 10-year risk of death from the disease is 18% in a world where it is not possible to die from any other cause than prostate cancer. When recognizing that other causes of death are also possible, the risk of death due to the disease is instead 11%. The net probability of death hence overestimates the real-world risk of dying from the prostate cancer since the reality is that competing causes of death will lead to the death of some men before they 'have the opportunity' to die from their disease. In this example, the 10-year risk of death due to other causes than prostate cancer was 41%.

Survival in the presence of competing risks is by no means a novel concept but there has been a surge in interest in these methods from the medical community over the past twenty years, not least in cancer epidemiology. Competing risks methodology has been developed for both the cause-specific [14, 16] and relative survival setting [31, 32]. The cause-specific approach has similar data requirements as standard cause-specific analyses. Each cause of death must be classified as related or unrelated (competing) to the cancer based on death certificate information. In the relative survival framework, separating deaths due to the disease from competing events is done implicitly in the estimation process as the deaths that are unrelated to the cancer are cancelled out when contrasting the overall mortality of the patients to the expected mortality in the general population. All deaths that occur in the general population, and that contribute to the expected mortality rate, are assumed to be unrelated to prostate cancer and can therefore contribute to estimation of the risk associated with competing events.

In Fig. 3, probabilities of death from prostate cancer and competing causes during 5 years of follow-up for men diagnosed in Sweden in 2012 at ages 55, 65 and 75 years are shown. As expected, age at diagnosis is strongly associated with risk of death due to other causes than prostate cancer. The risk of death due to prostate cancer is, however, not as strongly associated with age. However, the latter observation should not be used to draw conclusions about differences in disease severity by age. It rather reflects that men with prostate cancer who are older at diagnosis are more likely to die from competing causes before they 'have a chance' to die from their prostate cancer compared to men who are younger.

In this example, 2% of the men who were 55 years at diagnosis are predicted to have died from their

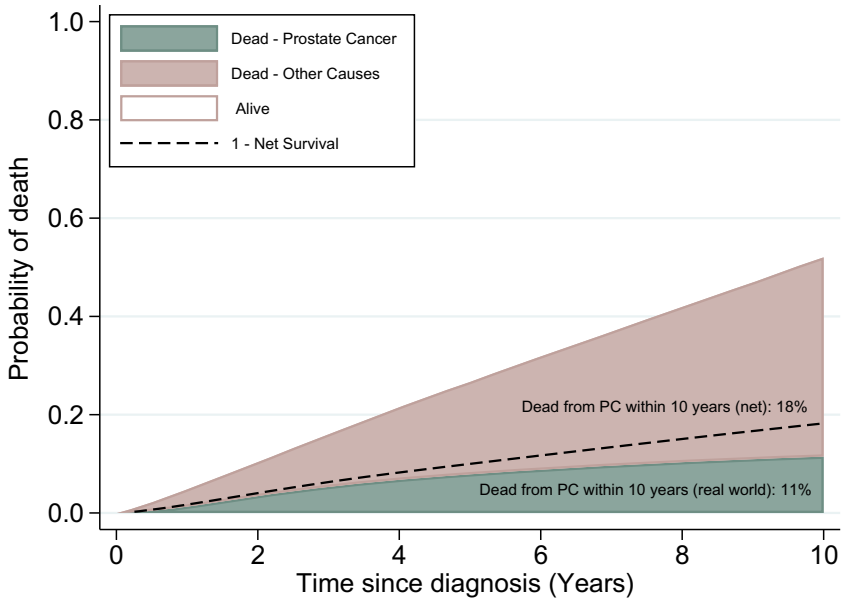


Fig. 2 *Probabilities of death due to prostate cancer and deaths due to other causes in 75-year-old men diagnosed in 2007 in the presence and absence of competing risks, respectively.*

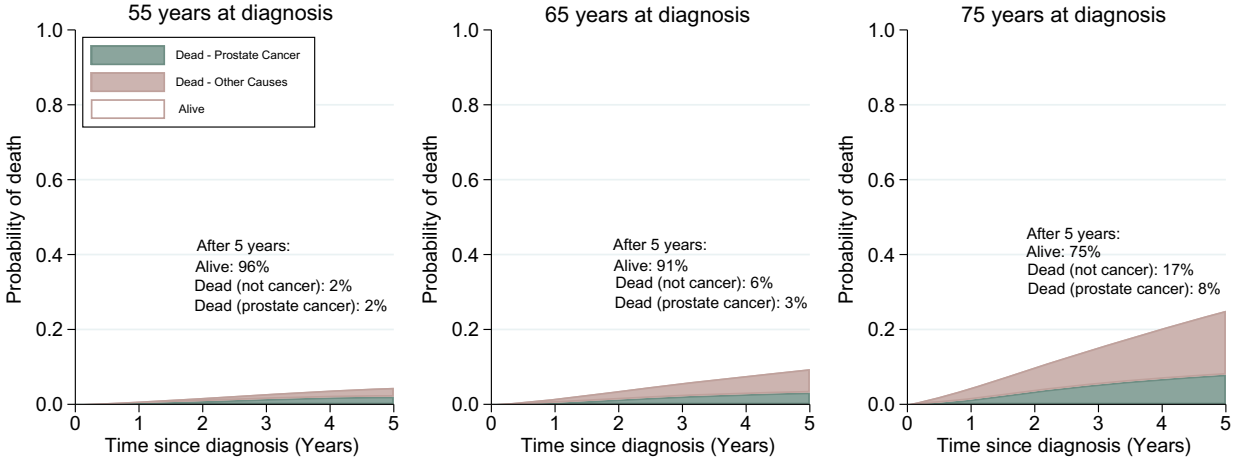


Fig. 3 *Probabilities of death due to prostate cancer and deaths due to other causes amongst 55-, 65 and 75-year-old men diagnosed in 2012.*

cancer within the first 5 years after the diagnosis, as compared to 8% of the 75-year-olds (green curve). Amongst the 55-year-olds, the corresponding proportion is 96% as deaths due to other causes are less common. The reality for those diagnosed at age 75 is nevertheless that 75% will not have died from any cause within 5 years of their cancer diagnosis.

Figure 4 shows corresponding probabilities of death for women diagnosed with breast cancer, colon cancer and CML. For all three cancers, the proportion of patients that have died within 5 years after diagnosis increases with increasing age, and for younger ages, almost all deaths are due to the cancer. However, age differences in the proportion that have died due to cancer differ between the cancers. For breast cancer, the proportion that have died due to any cause at 5 years is 7% for patients aged 55 at diagnosis, 9% for patients aged 65 and 22% for patients aged 75 at diagnosis. The proportion that have died due to breast cancer is 5%, 5% and 10% for patients aged 55, 65 and 75 at diagnosis, respectively. These figures reflect the

successive improvement of long-term outcomes in breast cancer in recent decades, due to advances in detection and treatment [33]. Nevertheless, due to the high breast cancer incidence, death due to breast cancer remains the second most common cause of death due to malignancy amongst women in high income countries [34].

For colon cancer patients, the majority of deaths within 5 years are due to cancer for all three selected ages, the proportion that have died due to any cause is 32%, 34% and 40%, and due to cancer is 31% for all three ages at diagnosis reflecting both a worse prognosis than the other two cancer forms and the fact that cancer recurrence occurs mainly within the first years following diagnosis [35]. Most deaths due to colon cancer happen in the first years after diagnosis, after which the proportion that die due to colon cancer levels off. The proportion of CML patients who died 5 years after diagnosis is 7%, 14% and 28% for patients aged 55, 65 and 75 at diagnosis, respectively. The proportion who died due to CML is 5%, 10% and 17% for the same ages. Unlike for colon cancer, this proportion continues to increase during the whole 5-year follow-up period.

In the examples presented so far, patients have only been stratified with respect to age at diagnosis, calendar year and sex. Whilst these data provide a broad overview of the anticipated prognosis of cancer patients, the aim of competing risks analyses is often to quantify absolute risks for the purpose of risk communication, healthcare planning or health economic investigations. To this end, results for broader patient groups may not provide sufficient granularity to facilitate the understanding of how the prognosis varies within clinically relevant subgroups. A more realistic

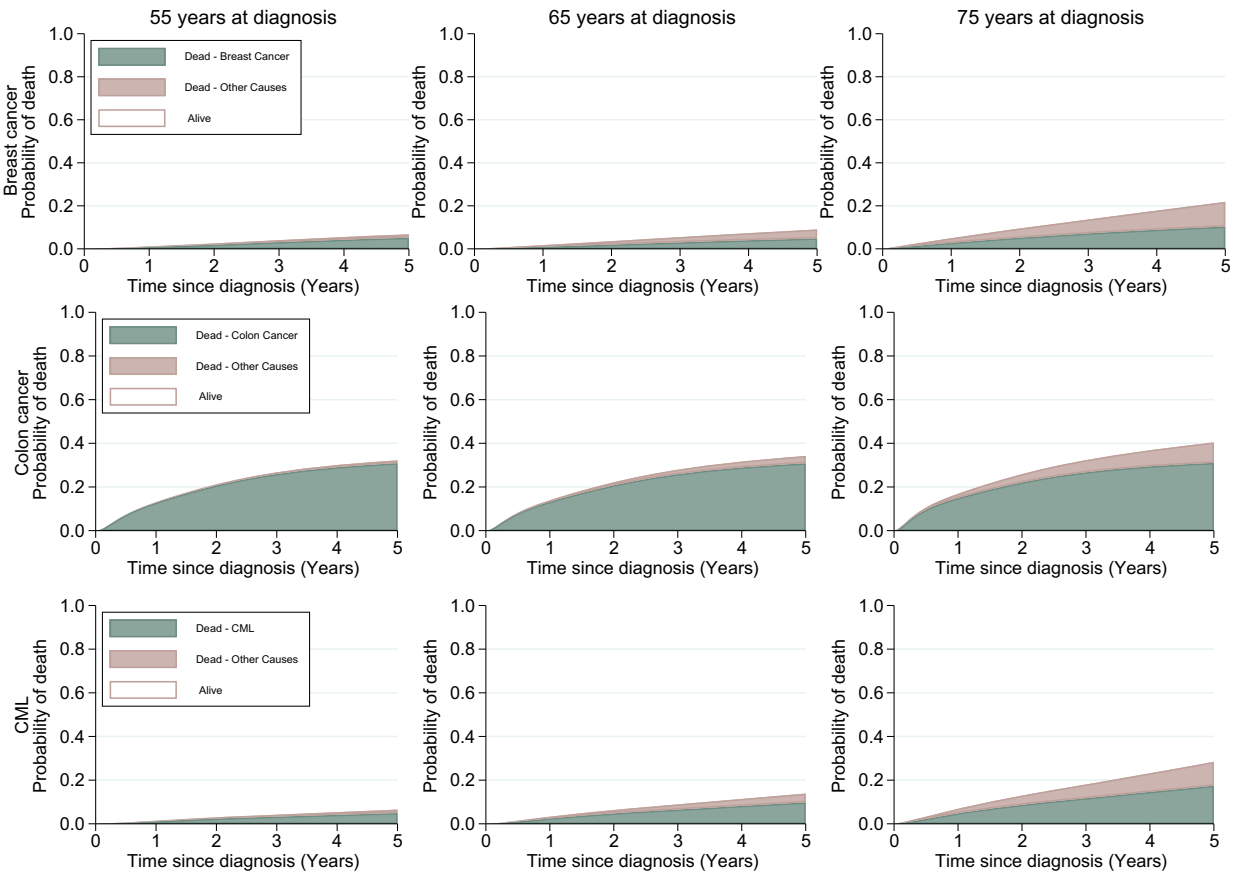


Fig. 4 *Stacked probabilities of death due to breast cancer, colon cancer and Chronic Myeloid leukaemia (CML) and deaths due to other causes amongst 55-, 65 and 75-year-old patients diagnosed in 2012.*

picture would include consideration of other known prognostic disease factors. These will vary by disease and could include:

- Risk group classifications based on tumour size, spread, prostate-specific antigen and Gleason score and initial treatment strategy for prostate cancer.
- Hormone receptor and HER-2 status, number of affected lymph nodes and adjuvant therapy for women with breast cancer.
- Disease localization, neoadjuvant and adjuvant radio and/or chemotherapy and type of surgery for patients with colon cancer.
- Prognostic scores, for example the Hasford score calculated using age, spleen size and laboratory values for CML

Prediction models that incorporate detailed clinical and biological data and aim to aid medical decision making have been developed and implanted in clinical practice with supporting online tools [36, 37]. Examples of such validated tools include PREDICT (https://breast.predict.nhs.uk/), used to support treatment selection in early breast cancer [38], and its counterpart PREDICT prostate (https://prostate.predict.nhs.uk/) for men with nonmetastatic prostate cancer [39]. The underlying statistical models in these specific tools predict all-cause survival, cancer-specific survival in the presence of competing risks, and risk for treatment side-effects (PREDICT prostate only) according to patient demographics, clinical and biological characteristics of the disease, detection mode. Other interactive tools that are available for physicians for risk counselling include the Memorial Sloan Kettering Cancer Center prediction tools that encompass a range of sites, including colorectal, bladder, endometrial, gastric, liver, lung, melanoma, sarcoma, renal cell carcinoma, prostate and breast cancer (https://www.mskcc.org/nomograms).

Loss in life expectancy

Survival in the presence of competing risks provides an easy-to-interpret summary measure of cancer patient survival within a fixed time period after diagnosis (up to 10 years in the examples above). Other measures that quantify survival across the entire remaining life span can provide useful insight into the prognosis and burden of disease in the longer perspective. One such measure, that is also interpretable in the real world, is loss in life expectancy [40, 41]. Life expectancy, in general, is a measure of average expected survival time. For human populations, it is often reported from birth, but can be estimated and reported according to any combination of year of birth, attained age, sex or other demographic factors. For the general population, such data can be obtained via publicly available country-specific population life tables [42]. In cancer patient survival studies, life expectancy is typically estimated from the date of cancer diagnosis until death (irrespective of cause of death) and gives a prediction of the number of life years that remain for patients after they have been diagnosed with cancer.

Contrasting the remaining life expectancy amongst cancer patients to that in the general population gives an estimate of the loss in life expectancy (LLE) due to cancer [40, 43-45]. The LLE is interpreted as the number of life years a cancer patient is expected to lose, on average, due to the cancer diagnosis. Despite the simplicity with respect to its theoretical definition, LLE has been used little in research practice as a tool to summarize cancer patient survival. However, due to recent methodological developments that facilitate its estimation, the LLE has gained popularity also in observational cancer studies [46-50]. For example, the LLE has been used to efficiently summarize the continual survival improvements amongst patients with CML [48]. In Fig. 5, the life expectancy of a female CML patient aged 70 at diagnosis across calendar time is shown along with the life expectancy for females of the same age in the general population. The difference between the two lines (indicated by the yellow arrow) is the estimated number of years lost due to the disease. It can be seen that a patient diagnosed in 1993 had a remaining life expectancy of 4.2 years, whereas the corresponding time was 16.4 years for a comparable female in the general population, giving a LLE of 12.2 years. Similarly, for a female of the same age who was diagnosed in 2010 the life expectancy was 14.7 years, compared to 18.2 years for females without CML, giving a LLE of only 3.5 years.

The improvements in the life expectancy of CML patients over time by far exceed the improvements seen in the general population. Importantly, for the latter calendar period the life expectancy of the patients has approached that in the general population. The marked improvements shown here are

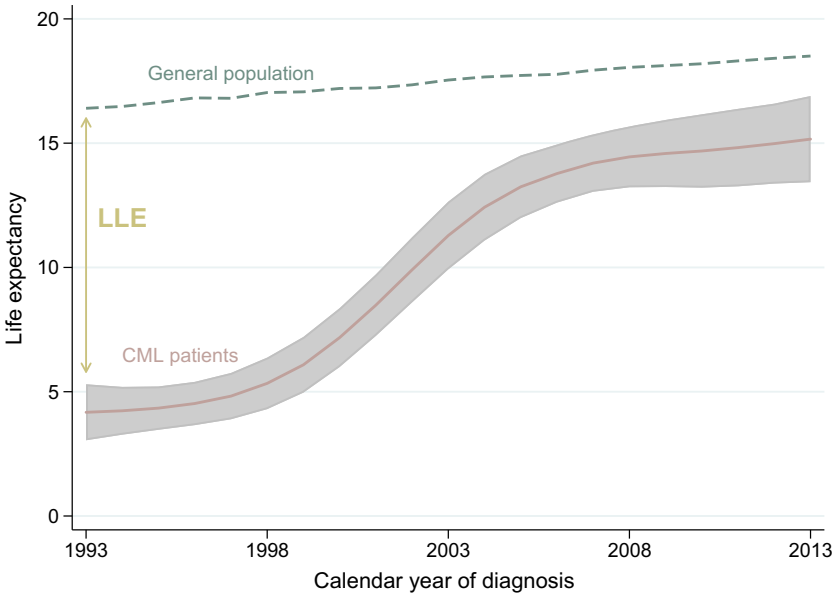


Fig. 5 *Illustration of loss in life expectancy for women aged 70 years at diagnosis of chronic myeloid leukaemia (CML) in Sweden.*

likely due to the introduction of imatinib mesylate and other new tyrosine kinase inhibitors along with allogeneic stem cell transplantation for which efficacy has previously been shown in clinical trials [51, 52].

The LLE varies greatly across age, since younger individuals in the general population have a longer life expectancy, and therefore more life years to lose. In Fig. 6, the LLE across age at diagnosis is presented for prostate cancer patients, female breast cancer patients, female colon cancer patients and female CML patients who were diagnosed in 2007. As anticipated, LLE is consistently higher at young ages (50–60 years) compared with older ages (70 years and above) although with substantial variation in numbers between the different cancers. High losses are observed amongst young patients with aggressive cancer types such as colon and breast cancer. In prostate cancer, losses are lower in young ages, likely reflecting the indolent nature of many diagnosed cases and early detection through PSA-screening [53].

Whilst survival is generally better for younger cancer patients for most cancer sites, the impact of cancer on the life expectancy is by definition greater for younger patients compared to older. For this reason, the LLE is not a relevant measure for drawing conclusions about the prognostic effect of age on survival. An alternative measure that allows more relevant comparisons across age is the proportion of life lost due to cancer. It is calculated by dividing the LLE by the expected remaining life years and answers the question: *What proportion of my expected remaining life may be lost due to the cancer?*

The proportion of life lost is presented in Fig. 7 for prostate cancer patients, female breast cancer patients, female colon cancer patients and female CML patients, diagnosed in 2007. For prostate-, breast- and colon cancer, where there are great differences in the LLE across age, the proportion of life lost is fairly constant across age. Prostate and breast cancer patients loose approximately 10-20% of their remaining life expectancy, and colon cancer patients approximately 30-40% of their remaining life expectancy. For CML patients, the LLE is more similar across age, whereas instead the proportion of life lost varies across age.

Since the loss in life expectancy is calculated across the entire remaining life span, the method relies on extrapolation of survival beyond observable data. Restricted mean survival is a related measure that instead estimates the loss in life expectancy within a predefined, limited observation window, for example five or ten years. The restricted time window obviates the need to make assumptions about survival trends in the cancer patients and in the general population into the nonobservable future [54]. For this reason, restricted mean survival can be a useful summary

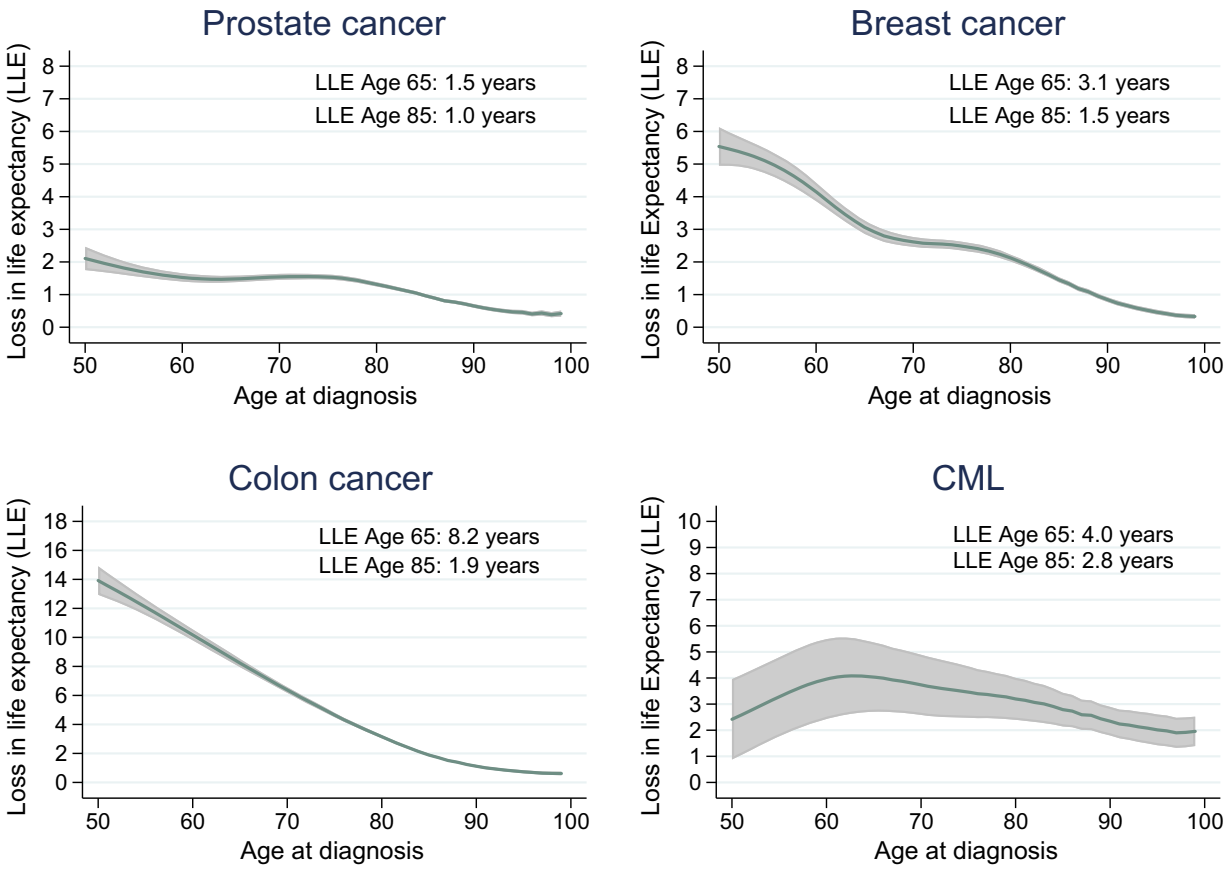


Fig. 6 *Loss in life expectancy by age at diagnosis for men with prostate cancer, female patients with breast cancer, colon cancer and chronic myeloid leukaemia (CML) and diagnosed in 2007 in Sweden.*

measure for young patients for whom extrapolation of survival 30–40 years into the future would otherwise imply a high degree of uncertainty in the estimates. Hodgkin lymphoma is a malignancy well-known for affecting young individuals and for whom restricted mean survival has been applied to show loss in life expectancy within the first five years after diagnosis [55].

Updating the prognosis for survivors—conditional survival

In the examples discussed so far, survival has been presented from diagnosis and gives estimates of the pretreatment prognosis. When a cancer patient has survived the first period after diagnosis, the overall survival estimates at diagnosis no longer apply. From a patient perspective, it is then of more interest to study conditional estimates, for example *What is the survival among patients that have survived the first year, or the first five years*? [56, 57]. To this end, all measures that have been introduced in this tutorial can also be presented as conditional estimates. Conditional estimates of LLE are shown in Fig. 8 for prostate cancer patients, female breast cancer patients, female colon cancer patients and female CML patients diagnosed in 2007. For example, a woman diagnosed with colon cancer aged 65 years in 2007 is expected to lose 8.2 life years from diagnosis. However already after having survived 1 year from diagnosis, the LLE has decreased to 5.9 life years, and after 5 years the remaining LEL is 1.8 years. A 75-year-old female patient is expected to lose 4.7 life years, but this decreases to 0.7 at 5 years after diagnosis, indicating that if a 75-year-old patient survives 5 years after diagnosis, the life expectancy is similar to cancer-free women of the same age.

The summarized information in Fig. 8 is quite dense which is a necessity for scientific

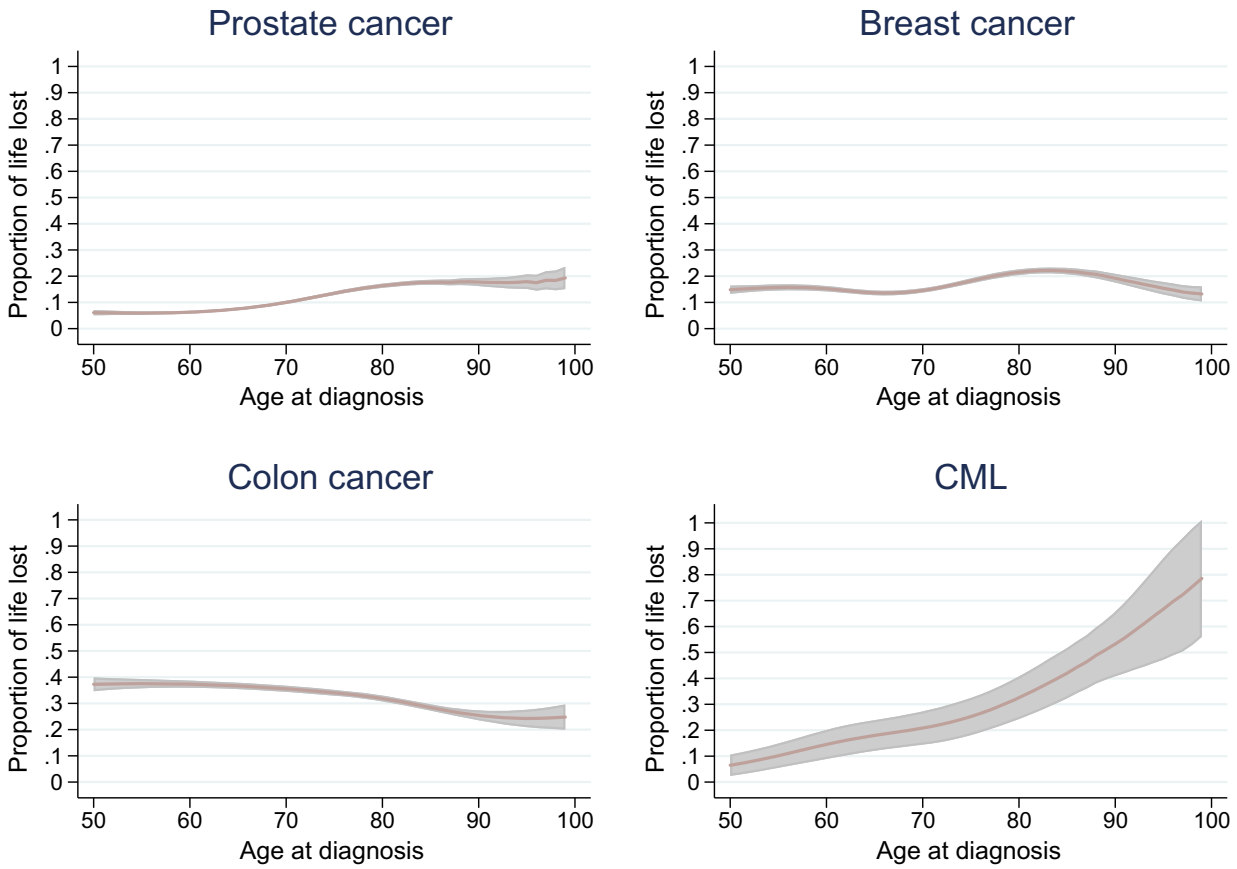


Fig. 7 *Proportion of life lost by age at diagnosis for men with prostate cancer, female patients with breast cancer, colon cancer and chronic myeloid leukaemia (CML) and diagnosed in 2007 in Sweden.*

publications and mostly useful for readers with a high degree of information literacy. InterPreT (https://interpret.le.ac.uk/) is an interactive cancer survival prediction tool that uses population-based cancer data from England and Sweden to illustrate variation in survival by patient characteristics. Both net survival and survival in the presence of competing risks are reported, and conditional survival estimates are available through dynamic illustrations. To aid communication with nonspecialist target groups, it may also be useful to present summary statistics in a more simplified manner. Infographics are graphical tools that are helpful to visualize prognostic information and that are sometimes used to improve patient understanding. Figure 9 shows life expectancy of a female colon cancer patient diagnosed in 2007 at age 65 and corresponding life expectancy of a 65-year-old in the general population. In this figure, each bar represents an expected remaining life year at the time of diagnosis (left) and 5 years after diagnosis (right), respectively. Specifically, the yellow bars represent the remaining life years in women with colon cancer and the sum of the yellow and blue bars represent the remaining life years in women in the general population (assumed free from colon cancer).

This illustration clearly visualizes that the discrepancy in life expectancy at diagnosis between women with colon cancer and women without cancer is 8 years. Amongst five-year survivors the expected survival in the general population is 18 years, the conditional survival of women with a history of colon cancer is 16 years, which gives a difference of 2 life years lost due to the cancer.

Efficient tools for risk communication, such as infographics, are increasingly important in times

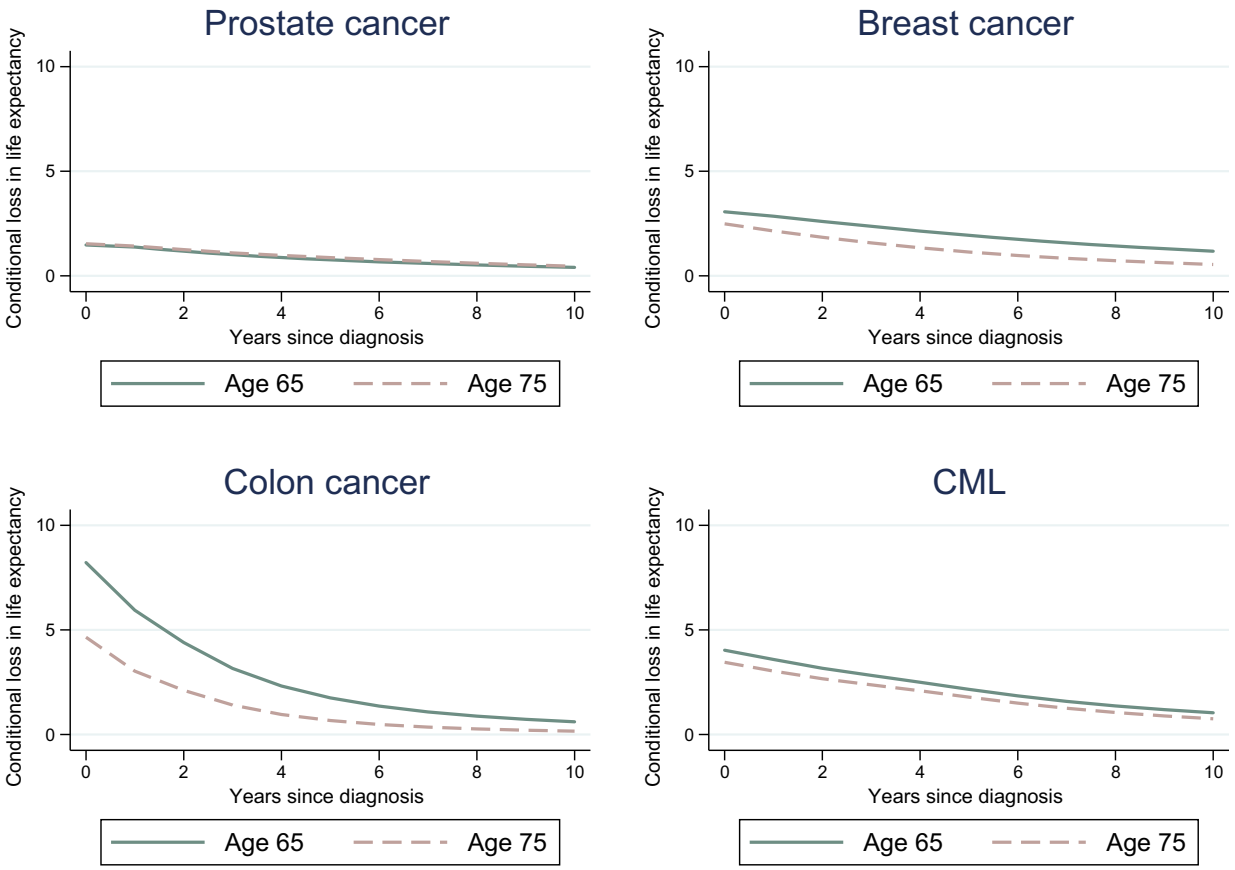


Fig. 8 *Conditional estimates of loss in life expectancy for men with prostate cancer, female patients with breast cancer, colon cancer and chronic myeloid leukaemia (CML) and diagnosed in 2007 in Sweden.*

when patients actively search web-based resources to learn more about their anticipated disease course. To meet this information need, it is imperative that data for this purpose are recent, accurate and with the source clearly stated. In all situations where results are presented for lay audiences, disclaimers should accompany the data stating the conditions for interpretation, and that the most appropriate information for the individual should be discussed with the treating physician.

Discussion and conclusion

Using population-based data from Sweden, we have described two key statistical measures that are useful for describing and communicating the real-world survival experience of cancer patients; cancer survival in the presence of competing risks and expected life years lost due to the cancer. These measures are used increasingly in the scientific literature as tools for prediction or to understand the impact of a cancer diagnosis on the remaining life expectancy. In contrast to net statistical measures that are commonly reported in cancer control research and descriptive epidemiology, the measures highlighted in this tutorial are produced with the interest of risk counselling between healthcare professionals and patients in mind. We also described the concept of conditional survival, that is updated survival estimates that are particularly relevant for survivors as they reflect the inherent change in the anticipated prognosis as patients live through and beyond their cancer diagnosis. Although the fundamental idea of these measures and concepts are not novel from a theoretical point of view, examples in the clinical literature suggest there are misconceptions with respect to what measures are appropriate to report given the research question at hand [58, 59].

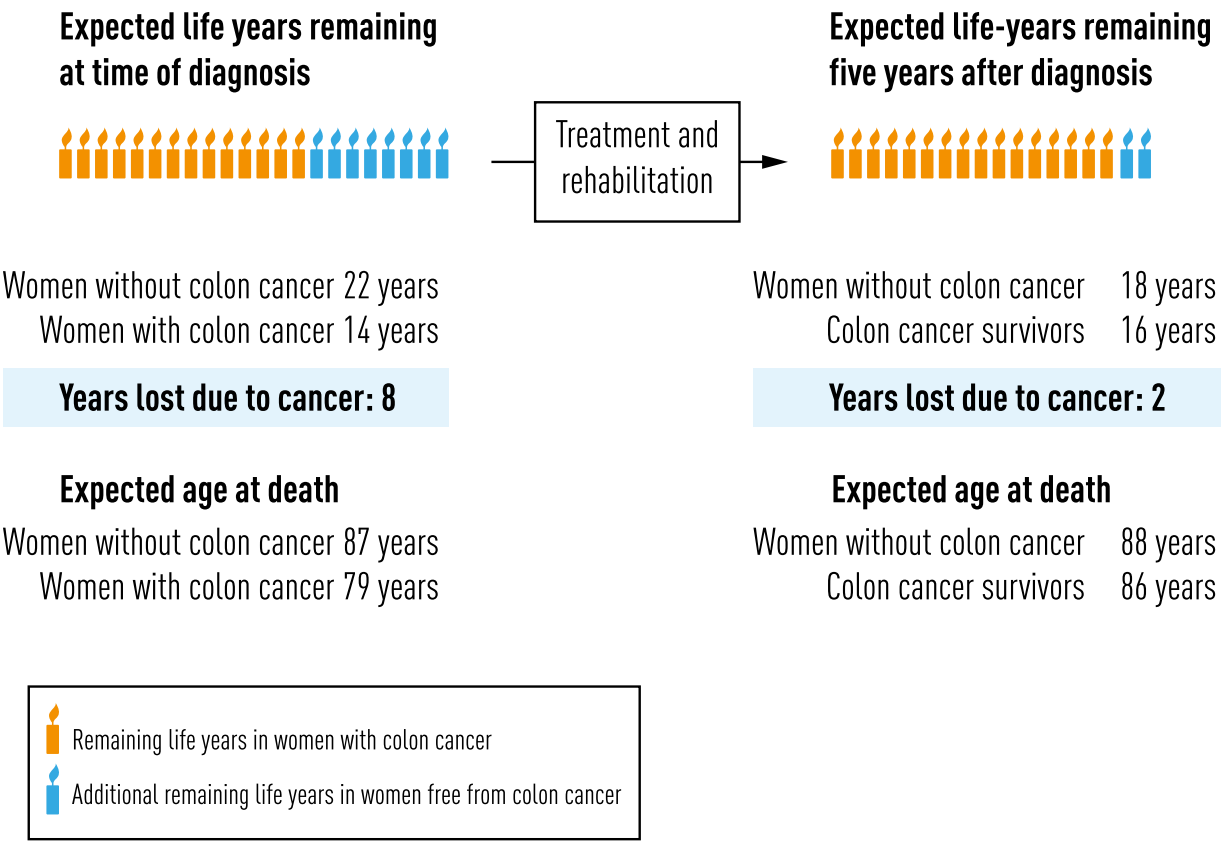


Fig. 9 *Example illustrating the life expectancy and loss in life expectancy of a female colon cancer patient diagnosed in Sweden in 2007 at age 65, at diagnosis and conditional on surviving 5 years.*

Plausible reasons for the said confusion are failure to convey the purpose of the study and the nature of the hypothesis (hypothetical versus real-world) but also preferences in estimation methods and their associated terminology. The former reason is a delicate matter that requires careful consideration and thought at the design phase of any investigation. Broadly speaking, studies that aim to understand how some underlying biological mechanism, hypothesized to drive differences in cancer outcome between groups, should be performed in the net survival setting (absence of competing risks). Conversely, for investigations where absolute risk comprehension is the goal (whether for healthcare planning or risk communication) estimating the impact of competing risks in parallel to the cancer outcome is necessary.

If one is interested in comparing treatment efficacy, randomized controlled trials (RCT) remain the gold standard study design. RCTs provide estimates of the best achievable survival since patients enrolled in trials are typically younger and healthier than the average patient. In the context of trials, relative survival is an inappropriate summary measure of survival since the general population does not provide a representative comparison group for patients who are selected for RCTs. Population-

based studies, on the other hand, provide estimates of patient survival amongst all patients, including those not eligible for trial inclusion.

With respect to terminology, different estimation methods can be used to target the same underlying theoretical quantity but the language used for reporting varies. For example, cause-specific survival, relative survival, actuarial survival and cumulative incidence are all terms used in population-based cancer patient survival research to report estimates of net survival. Likewise, terms reflecting survival in the presence of competing risks include crude survival, (cause-specific) cumulative incidence and real-world risk. We argue that 'Survival in the absence of competing risks', and 'Survival in the presence of competing risks' are intuitive and less prone to ambiguity when communicating with nonspecialist target groups and chose this nomenclature for this tutorial. However, we also recognize a need for consistent terminology within specialist fields. For example, the term 'cause-specific survival' (reflecting net survival) has been criticized for being ambiguous since during follow-up, every patient has either died of one of many possible causes or survived all possible causes of death [13]. Whilst this is a semantic issue with 'cause-specific survival', we do not think the term should be avoided since is the established technical term for what is a very common descriptive statistic in cancer control research and descriptive epidemiology.

We also briefly touched upon the idea of generalizability of results to individual patients through increasing use of detailed clinical data coupled with online tools to facilitate uptake of elaborate models amongst physicians. This is a nontrivial topic that requires ability to balance the model-based simplifications of the clinical reality with other patient-specific factors, and an understanding that this area of research is a fast moving target. Individual-level prognostication that integrates clinical, biological and imaging data to facilitate patient tailored treatment strategies is increasingly called for in the era of precision medicine. And although beyond the scope of this review, machine learning algorithms, particularly deep learning methods including artificial neural networks and support vector machines, have emerged in various fields of oncology [60, 61]. These methods are inherently data-driven and have been appraised for their predictive accuracy which stems from their ability to capture complex nonlinear correlations in high dimensional mixed data [62]. However, several challenges remain before artificial intelligence and machine learning tools are likely to be widely used in clinical practice. Important barriers for wider uptake and grounds for more research include lack of interpretability of the prediction models, as machine learning models are typically not causal, and insufficient external validation of the predictive accuracy [62, 63]. This is nonetheless an active area of research that will undoubtedly continue to create interest in the clinical community for many years to come and gain more acceptance as a complementary tool to support medical decision making once intelligible and generalizable models become available.

To conclude, statistical methods for competing risks, loss in life expectancy and conditional survival have become increasingly popular amongst medical researchers, and in this review we have discussed when these methods are appropriate to use, under what assumptions the results are valid and how results are interpreted. An important take home message for when choosing between summary measures of survival is to carefully determine the nature of the hypothesis under investigation. Whilst traditional measures of net survival are highly relevant in cancer control and suitable in aetiologic research, measures that account for competing risks and/or provides updated estimates of survival as the lives of cancer survivors progress are more relevant for patients and healthcare professionals. Reporting of such statistics may range from dense data summaries for specialist readers of scientific publications to interactive online tools that allow for a higher level of granularity of the data, as well as adaptation towards lay persons.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Source of funding

Therese Andersson's contribution to this work was funded via the Swedish Cancer Society (grant number: 19 0102 Pj) and the Swedish Research Council (grant number: 2019-01965).

Author Contribution

Sandra Eloranta: Conceptualization (equal); Data curation (supporting); Formal analysis (equal);

Investigation (lead); Methodology (equal); Project administration (lead); Visualization (lead); Writing-original draft (lead); Writing-review & editing (equal). **Karin Ekstrom Smedby:** Conceptualization (equal); Investigation (supporting); Methodology (supporting); Supervision (supporting); Visualization (supporting); Writing-review & editing (equal). **Paul Dickman:** Conceptualization (equal); Data curation (lead); Formal analysis (supporting); Investigation (supporting); Methodology (equal); Visualization (supporting); Writing-review & editing (equal). **Therese M-L Andersson:** Conceptualization (equal); Data curation (supporting); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (supporting); Writing-original draft (supporting); Writing-review & editing (equal).

References

1 Belot A, Ndiaye A, Luque-Fernandez MA *et al.* Summarizing and communicating on survival data according to the audience: a tutorial on different measures illustrated with population-based cancer registry data. *Clin Epidemiol* 2019; **11:** 53–65.
2 Dickman PW, Adami HO. Interpreting trends in cancer patient survival. *J Intern Med* 2006; **260:** 103–17.
3 Mariotto AB, Noone AM, Howlader N *et al.* Cancer survival: an overview of measures, uses, and interpretation. *J Natl Cancer Inst Monogr* 2014; **2014:** 145–86.
4 Brewster DH, Coebergh JW, Storm HH. Population-based cancer registries: the invisible key to cancer control. *Lancet Oncol* 2005; **6:** 193–5.
5 Coleman MP. Cancer survival: global surveillance will stimulate health policy and improve equity. *Lancet* 2014; **383:** 564–73.
6 Arnold M, Rutherford MJ, Bardot A *et al.* Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol* 2019; **20:** 1493–505.
7 Coleman MP, Forman D, Bryant H *et al.* Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011; **377:** 127–38.
8 Allemani C, Matsuda T, Di Carlo V *et al.* Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018; **391:** 1023–75.
9 Allemani C, Weir HK, Carreira H *et al.* Global surveillance of cancer survival 1995–2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* 2015; **385:** 977–1010.
10 Coleman MP, Quaresma M, Berrino F *et al.* Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 2008; **9:** 730–56.
11 Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T, Este've, J. (Eds.), *Survival of cancer patients in Europe: the EURO-CARE study*. Lyon: IARC Scientific Publication. 1995.
12 Perme MP, Stare J, Esteve J. On estimation in relative survival. *Biometrics* 2012; **68:** 113–20.
13 Pohar Perme M, Esteve J, Rachet B. Analysing population-based cancer survival - settling the controversies. *BMC Cancer* 2016; **16:** 933.
14 Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007; **26:** 2389–430.
15 Geskus RB. *Data analysis with competing risks and intermediate states*. New York: Chapman & Hall, 2015.
16 Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival data analysis. *Br J Cancer* 2004; **91:** 1229–35.
17 Eloranta S, Adolfsson J, Lambert PC *et al.* How can we make cancer survival statistics more useful for patients and clinicians: an illustration using localized prostate cancer in Sweden. *Cancer Causes Control* 2013; **24:** 505–15.
18 Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol* 2012; **41:** 861–70.
19 Dignam JJ, Kocherginsky MN. Choice and interpretation of statistical tests used when competing risks are present. *J Clin Oncol* 2008; **26:** 4027–34.
20 Lambert PC, Dickman PW, Rutherford MJ. Comparison of different approaches to estimating age standardized net survival. *BMC Med Res Methodol* 2015; **15:** 64.
21 Seppa K, Hakulinen T, Pokhrel A. Choosing the net survival method for cancer survival estimation. *Eur J Cancer* 2015; **51:** 1123–9.
22 Bright CJ, Brentnall AR, Wooldrage K, Myles J, Sasieni P, Duffy SW. Errors in determination of net survival: cause-specific and relative survival settings. *Br J Cancer* 2020; **122:** 1094–101.
23 Forjaz de Lacerda G, Howlader N, Mariotto AB. Differences in cancer survival with relative versus cause-specific approaches: an update using more accurate life tables. *Cancer Epidemiol Biomarkers Prev* 2019; **28:** 1544–51.
24 Fall K, Stromberg F, Rosell J, Andren O, Varenhorst E, South-East Region Prostate Cancer G. Reliability of death certificates in prostate cancer patients. *Scand J Urol Nephrol* 2008; **42:** 352–7.
25 Cho H, Mariotto AB, Mann BS, Klabunde CN, Feuer EJ. Assessing non-cancer-related health status of US cancer patients: other-cause survival and comorbidity prevalence. *Am J Epidemiol* 2013; **178:** 339–49.
26 Berglund A, Garmo H, Tishelman C, Holmberg L, Stattin P, Lambe M. Comorbidity, treatment and mortality: a population based cohort study of prostate cancer in PCBaSe Sweden. *J Urol* 2011; **185:** 833–9.
27 Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol* 2013; **66:** 648–53.
28 Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med* 2012; **31:** 1089–97.
29 Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009; **170:** 244–56.

30 Chappell R. Competing risk analyses: how are they different and why should you care? *Clin Cancer Res* 2012; **18:** 2127–9.
31 Cronin KA, Feuer EJ. Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Stat Med* 2000; **19:** 1729–40.
32 Lambert PC, Dickman PW, Nelson CP, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Stat Med* 2010; **29:** 885–95.
33 DeSantis CE, Ma J, Gaudet MM *et al.* Breast cancer statistics, 2019. *CA Cancer J Clin* 2019; **69:** 438–51.
34 Hashim D, Boffetta P, La Vecchia C *et al.* The global decrease in cancer mortality: trends and disparities. *Ann Oncol* 2016; **27:** 926–33.
35 Shah MA, Renfro LA, Allegra CJ *et al.* Impact of patient factors on recurrence risk and time dependency of oxaliplatin benefit in patients with colon cancer: analysis from modern-era adjuvant studies in the Adjuvant Colon Cancer End Points (ACCENT) database. *J Clin Oncol* 2016; **34:** 843–53.
36 Muhlbauer V, Berger-Hoger B, Albrecht M, Muhlhauser I, Steckelberg A. Communicating prognosis to women with early breast cancer - overview of prediction tools and the development and pilot testing of a decision aid. *BMC Health Serv Res* 2019; **19:** 171.
37 Rabin BA, Gaglio B, Sanders T *et al.* Predicting cancer prognosis using interactive online tools: a systematic review and implications for cancer care providers. *Cancer Epidemiol Biomarkers Prev* 2013; **22:** 1645–56.
38 Wishart GC, Azzato EM, Greenberg DC *et al.* PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res* 2010; **12:** R1.
39 Thurtle DR, Greenberg DC, Lee LS, Huang HH, Pharoah PD, Gnanapragasam VJ. Individual prognosis at diagnosis in nonmetastatic prostate cancer: Development and external validation of the PREDICT Prostate multivariable model. *PLoS Medicine* 2019; **16:** e1002758.
40 Andersson TM, Dickman PW, Eloranta S, Lambe M, Lambert PC. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Stat Med* 2013; **32:** 5286–300.
41 Viscomi S, Pastore G, Dama E *et al.* Life expectancy as an indicator of outcome in follow-up of population-based cancer registries: the example of childhood leukemia. *Ann Oncol* 2006; **17:** 167–71.
42 Barbieri M, Wilmoth JR, Shkolnikov VM *et al.* Data resource profile: the human mortality database (HMD). *Int J Epidemiol* 2015; **44:** 1549–56.
43 Andersson TM, Rutherford MJ, Lambert PC. Illustration of different modelling assumptions for estimation of loss in expectation of life due to cancer. *BMC Med Res Methodol* 2019; **19:** 145.
44 Hakama M, Hakulinen T. Estimating the expectation of life in cancer survival studies with incomplete follow-up information. *J Chronic Dis* 1977; **30:** 585–97.
45 Rutherford MJ, Andersson TM, Bjorkholm M, Lambert PC. Loss in life expectancy and gain in life years as measures of cancer impact. *Cancer Epidemiol* 2019; **60:** 168–73.
46 Andersson TM, Dickman PW, Eloranta S, Sjovall A, Lambe M, Lambert PC. The loss in expectation of life after colon cancer: a population-based study. *BMC Cancer* 2015; **15:** 412.
47 Baade PD, Youlden DR, Andersson TM *et al.* Temporal changes in loss of life expectancy due to cancer in Australia: a flexible parametric approach. *Cancer Causes Control* 2016; **27:** 955–64.
48 Bower H, Bjorkholm M, Dickman PW, Hoglund M, Lambert PC, Andersson TM. Life expectancy of patients with chronic myeloid leukemia approaches the life expectancy of the general population. *J Clin Oncol* 2016; **34:** 2851–7.
49 Ekberg S, Jerkeman M, Andersson PO *et al.* Long-term survival and loss in expectancy of life in a population-based cohort of 7114 patients with diffuse large B-cell lymphoma. *Am J Hematol* 2018; **93:** 1020–8.
50 Jakobsen LH, Bogsted M, Brown PN *et al.* Minimal loss of lifetime for patients with diffuse large b-cell lymphoma in remission and event free 24 months after treatment: a Danish population-based study. *J Clin Oncol* 2017; **35:** 778–84.
51 Rousselot P, Cony-Makhoul P, Nicolini F *et al.* Long-term safety and efficacy of imatinib mesylate (Gleevec(R)) in elderly patients with chronic phase chronic myelogenous leukemia: results of the AFR04 study. *Am J Hematol* 2013; **88:** 1–4.
52 Sanchez-Guijo FM, Duran S, Galende J *et al.* Evaluation of tolerability and efficacy of imatinib mesylate in elderly patients with chronic phase CML: ELDERGLI study. *Leuk Res* 2011; **35:** 1184–7.
53 Schroder FH, Hugosson J, Roobol MJ *et al.* Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 2014; **384:** 2027–35.
54 Jakobsen LH, Andersson TM, Biccler JL, El-Galaly TC, Bogsted M. Estimating the loss of lifetime function using flexible parametric relative survival models. *BMC Med Res Methodol* 2019; **19:** 23.
55 Biccler JL, Glimelius I, Eloranta S *et al.* Relapse risk and loss of lifetime after modern combined modality treatment of young patients with Hodgkin lymphoma: a Nordic lymphoma epidemiology group study. *J Clin Oncol* 2019; **37:** 703–13.
56 Hieke S, Kleber M, Konig C, Engelhardt M, Schumacher M. Conditional survival: a useful concept to provide information on how prognosis evolves over time. *Clin Cancer Res* 2015; **21:** 1530–6.
57 Shack L, Bryant H, Lockwood G, Ellison LF. Conditional relative survival: a different perspective to measuring cancer outcomes. *Cancer Epidemiol* 2013; **37:** 446–8.
58 Bhaskaran K, Rachet B, Evans S, Smeeth L. Re: Helene Hartvedt Grytli, Morten Wang Fagerland, Sophie D. Fossa, Kristin Austlid Tasken. Association between use of beta-blockers and prostate cancer-specific survival: a cohort study of 3561 prostate cancer patients with high-risk or metastatic disease. *Eur Urol* 2013; **64:** e86–e7. http://dx.doi.org/10.1016/j.eururo.2013.01.007.: beta-blockers and prostate cancer survival--interpretation of competing risks models.
59 Grytli HH, Fagerland MW, Fossa SD, Tasken KA. Association between use of beta-blockers and prostate cancer-specific survival: a cohort study of 3561 prostate cancer patients with high-risk or metastatic disease. *Eur Urol* 2014; **65:** 635–41.
60 Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine learning in oncology: a clinical appraisal. *Cancer Lett* 2020; **481:** 55–62.
61 Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2015; **13:** 8–17.

62 Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med* 2016; **375:** 1216–9.
63 Frohlich H, Balling R, Beerenwinkel N *et al.* From hype to reality: data science enabling personalized medicine. *BMC Med* 2018; **16:** 150.

Correspondence: Sandra Eloranta, Department of Medicine, Karolinska Institutet, Clinical Epidemiology Unit T2, Karolinska University Hospital, Stockholm SE- 17176, Sweden.
(e-mail: sandra.eloranta@ki.se).

