Cancer Epidemiology 42 (2016) 133–139

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

An empirical comparison of methods for predicting net survival

Tor Åge Myklebust*, Bjarte Aagnes, Bjørn Møller

Department of Registration, Cancer Registry of Norway, Institute of Population-based Cancer Research, P.O. Box 5313 Majorstuen, 0304 Oslo, Norway

ARTICLE INFO

Article history:
Received 17 September 2015
Received in revised form 24 March 2016
Accepted 8 April 2016
Available online xxx

Keywords:
Net survival
Prediction
Prediction error
Flexible parametric models
Non-parametric estimator

ABSTRACT

Background: Providing accurate predictions of long-term net survival for recently diagnosed cancer patients is challenging due to the lack of follow-up. The aim of this study was to empirically compare predictions of net survival obtained from a flexible parametric excess hazard model to predictions obtained using the period and hybrid approaches.

Methods: The Cancer Registry of Norway was used to identify patients diagnosed with cancer during the period 1953–2008. Patients were then followed up for survival until the end of 2013. Net survival was calculated for 23 different cancer sites at 5, 10 and 15 years after diagnosis for each patient cohort. Observed net survival was estimated using the PP estimator. Predicted net survival was obtained omitting the most recent follow-up years using three approaches: a flexible parametric excess hazard model (FPM), the period approach, and the hybrid approach. All estimates were age-standardized to the age distribution of the cohort for which predictions were made. Prediction errors were calculated as the absolute difference between observed and predicted net survival.

Results: Average absolute prediction error across all cancer sites was smallest for FPM for 5-year, 10-year and 15-year net survival. FPM and the hybrid approach gave better predictions of 10- and 15-year net survival than the period approach. The period and hybrid approaches tended to over-estimate net survival for cancer sites with poor survival, and under-estimate net survival for cancer sites where survival has increased over time. Uncertainty in the predictions was considerably smaller when FPM was used compared to the other approaches.

Conclusions: FPM should be considered for predicting net survival when follow-up is incomplete.

1. Introduction

Cancer registries routinely publish estimates of net survival, and studies comparing cancer survival between countries or regions, and across time periods, are largely based on such estimates [1–4]. A net survival curve shows the proportion of patients still alive at a given time, assuming the cancer of interest is the only possible cause of death. A rigorous theoretical explanation of net survival has been given by Perme et al. [5]. Unfortunately, for recently diagnosed patient cohorts, only short-term observed net survival is available because of the short follow-up time. Therefore, predictions are needed to estimate net survival for recently diagnosed patients. Throughout this article the term prediction is used for situations when follow-up is incomplete, whereas the term estimation is used when follow-up is complete.

Estimates of net survival can be obtained using the cohort approach, which requires complete potential follow-up on all patients, meaning that 5-year net survival can be estimated only for patients diagnosed at least 5 years ago. However, as treatment regimens change, estimates from the cohort approach quickly become outdated. Alternatively, one could estimate net survival by the so-called period approach which was introduced in 1996 [6]. The period approach fixes an observation window, and net survival is estimated by left truncation at the start of the window and right censoring at the end of the window. Table 1 illustrates how the period approach may be used to predict net survival up to 10 years for the cohort of patients diagnosed in 2008–2012. Here, patients diagnosed in the period 1998–2007 who are still alive are considered to be at risk from January 1st 2008 and to death or to the censoring date December 31st 2012. In situations where there is delayed recording of incident cases, a commonly implemented period analysis would not make use of the follow-up information available after the last year of recorded cases. Table 1 illustrates this by showing that the period approach does not use any of the follow-up information in 2013. A natural way to solve this would be to shift the observation window 1 year forward to 2009–2013. Doing this, all conditional survival estimates, apart from the first year, contain contributions from 5 potential years of

* Corresponding author.
E-mail address: Tor.age.myklebust@kreftregisteret.no (T.Å. Myklebust).

http://dx.doi.org/10.1016/j.canep.2016.04.006

Table 1
Illustration of how the period and hybrid approaches utilize survival experience when predicting the 10-year net survival curve for patients diagnosed in 2008–2012, and with follow-up to the end of 2013. The colored area corresponds to the *period approach*, and the framed area where cell contents are in boldface and italics corresponds to the *hybrid approach* (For interpretation of the references to color in this table, the reader is referred to the web version of this article.).

Year of diagnosis \ Year of follow-up	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
1998	1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10					
1999		1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	***10***				
2000			1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	***9–10***	***10***			
2001				1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	***8–9***	***9–10***	***10***		
2002					1	1–2	2–3	3–4	4–5	5–6	6–7	***7–8***	***8–9***	***9–10***	***10***	
2003						1	1–2	2–3	3–4	4–5	5–6	***6–7***	***7–8***	***8–9***	***9–10***	***10***
2004							1	1–2	2–3	3–4	4–5	***5–6***	***6–7***	***7–8***	***8–9***	***9–10***
2005								1	1–2	2–3	3–4	***4–5***	***5–6***	***6–7***	***7–8***	***8–9***
2006									1	1–2	2–3	***3–4***	***4–5***	***5–6***	***6–7***	***7–8***
2007										1	1–2	***2–3***	***3–4***	***4–5***	***5–6***	***6–7***
2008											***1***	***1–2***	***2–3***	***3–4***	***4–5***	***5–6***
2009												***1***	***1–2***	***2–3***	***3–4***	***4–5***
2010													***1***	***1–2***	***2–3***	***3–4***
2011														***1***	***1–2***	***2–3***
2012															***1***	***1–2***

diagnosis. In this situation the observation window should be widened for the first year to also include patients diagnosed in 2008 (see Table 1). This subtle change in implementation is referred to as the hybrid approach, and was introduced in 2004 [7]. The hybrid approach does not restrict follow-up to the last year of diagnosis. Instead, the estimate of net survival is obtained by letting the time at which individuals become at risk (the start of the observation window) differ according to the time of diagnosis.

Alternatively, predicted net survival can be obtained by fitting a model including historical data, predicting net survival using the estimated parameter values. This is a useful approach if the assumptions of the underlying model are met. In this study, predictions of net survival were obtained from a flexible parametric cumulative excess hazard model (FPM).

Several empirical studies have concluded that the period and hybrid approaches are useful for predicting net survival [8–10]. Studies have also concluded that model-based predictions of net survival are accurate [11,12]. To our knowledge, no empirical comparison of predictions obtained from the period and hybrid approaches as well as predictions obtained from flexible parametric models have been done. The aim of this study was to empirically compare predictions of net survival obtained from a flexible parametric excess hazard model to predictions obtained using the period and hybrid approaches.

2. Material and methods

2.1. Data material

We included all diagnoses of cancer reported to the Cancer Registry of Norway between 1953 and 2008. Cancer cases were grouped into 23 categories based on topography, according to the annual report, *Cancer in Norway* (http://www.kreftregisteret.no/no/Generelt/Publikasjoner/Cancer-in-Norway/Cancer-in-Norway-2013/). Cancers diagnosed at autopsy were excluded from the analyses. A total of 453,202 cancers among 417,138 men and 419,386 cancers among 388,227 women were included in the analysis. Data from the Cancer Registry of Norway is linked regularly to the Norwegian Population Register, using the national personal identification number, to obtain dates of death and confirm continued residency in Norway. At the time of this analysis, follow-up was available up to and including 31 December 2013.

2.2. Statistical analysis

Cohorts were constructed using all available 5-year periods of diagnosis. Since a minimum of 15 years of follow-up is required to estimate the FPM, only cohorts with at least 15 years of historical data were eligible for analysis, meaning that the first cohort consisted of patients diagnosed in the period 1968–1972. Observed 5-, 10- and 15-year net survival was calculated for each cohort where follow-up was complete, and was estimated using the non-parametric estimator proposed by Perme et al. [5]. (When analyzing 10- and 15-year net survival, fewer cohorts have complete follow-up compared to analyzing 5-year net survival. Hence, the number of cohorts differs for each of the three analyses.) Predicted 5-, 10- and 15-year net survivals were obtained by censoring follow-up 1 year after the last year of diagnosis for each cohort, and estimated by fitting a flexible parametric cumulative excess hazard model and using the period and hybrid approaches. The PP estimator was used for the two non-parametric approaches. All period approach estimates were obtained using a 5-year observation window, corresponding to the 5-year periods of diagnosis. Estimates of net survival from the hybrid approach were achieved by letting the date at which patients became at risk differ according to year of diagnosis, as explained in the introduction.

The PP estimator is based on the inverse probability weighting (IPW) method. Within a time-interval each patient survival experience is weighted by the inverse of the cumulative probability of being alive at the beginning of the interval. Thus, older patients carry larger weights. When used in the context of the period and hybrid approaches the IPW weights are used only after left truncation, i.e., a patient only contributes to the net survival for the period he or she is assumed to be at risk. The IPW