Cancer Epidemiology 39 (2015) 465–472

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Cause-specific or relative survival setting to estimate population-based net survival from cancer? An empirical evaluation using women diagnosed with breast cancer in Geneva between 1981 and 1991 and followed for 20 years after diagnosis

Robin Schaffar [a,b,*], Bernard Rachet [b], Aurélien Belot [b], Laura Woods [b]

[a] *Geneva Cancer Registry, Global Health Institute, University of Geneva, Geneva, Switzerland*

[b] *Cancer Research UK Cancer Survival Group, Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom*

ARTICLE INFO

Article history:
Received 6 January 2015
Received in revised form 2 April 2015
Accepted 5 April 2015
Available online 20 April 2015

Keywords:
Net survival
Breast cancer
Cause-specific
Relative survival
Informative censoring

ABSTRACT

Background: Both cause-specific and relative survival settings can be used to estimate net survival, the survival that would be observed if the only possible underlying cause of death was the disease under study. Both resulting net survival estimators are biased by informative censoring and prone to biases related to the data settings within which each is derived. We took into account informative censoring to derive theoretically unbiased estimators and examine which of the two data settings was the most robust against incorrect assumptions in the data. *Patients and methods*: We identified 2489 women in the Geneva Cancer Registry, diagnosed with breast cancer between 1981 and 1991, and estimated net survival up to 20-years using both cause-specific and relative survival settings, by tackling the informative censoring with weights. To understand the possible origins of differences between the survival estimates, we performed sensitivity analyses within each setting. We evaluated the impact of misclassification of cause of death and of using inappropriate life tables on survival estimates. *Results*: Net survival was highest using the cause-specific setting, by 1% at one year and by up to around 11% twenty years after diagnosis. Differences between both sets of net survival estimates were eliminated after recoding between 15% and 20% of the non-specific deaths as breast cancer deaths. By contrast, a dramatic increase in the general population mortality rates was needed to see the survival estimates based on relative survival setting become closer to those derived from cause-specific setting. *Conclusion*: Net survival estimates derived using the cause-specific setting are very sensitive to misclassification of cause of death. Net survival estimates derived using the relative-survival setting were robust to large changes in expected mortality. The relative survival setting is recommended for estimation of long-term net survival among patients with breast cancer.

1. Introduction

Net survival is defined as the survival that would be observed if the only possible underlying cause of death was the disease under study [1]. This definition of survival probability is of particular interest since it is not influenced by changes in mortality from other causes and therefore allows accurate evaluation of survival from the disease, essential for cancer control.

Two main approaches have been developed to estimate net survival, each requiring different data settings and assumptions. First, the cause-specific approach, which requires a data setting with reliable individual information on the underlying cause of death. Thus, only deaths from the cancer under study are defined as events whilst others are censored. Second, the relative survival approach [2] compares the overall survival of a cohort of patients to that which they would have experienced if they had had the same mortality experience of the general population from which they were drawn. This approach requires a different data setting, where mortality data about the population from which the cancer

* Corresponding author at: Geneva Cancer Registry, Global Health Institute, University of Geneva, 55 Boulevard de la Cluse, 1205 Geneva, Switzerland. Tel.: +41 22 379 49 57.

E-mail address: robin.schaffar@unige.ch (R. Schaffar).

http://dx.doi.org/10.1016/j.canep.2015.04.001

patients are drawn is available. Information about the cause of death is not required and we assume that the cancer-specific mortality included in the overall mortality is negligible compared to the overall mortality.

Both approaches are prone to a bias called *informative censoring* [3]. This is where the assumption of independence between the censoring process and the occurrence of the event (death) does not hold. For instance, an older patient is more likely to die from other causes than the disease under study than a younger patient. Thus, the older patients are more likely not to experience the death from cancer of interest simply because of their older age. The censoring process is therefore dependant on age and becomes informative. To take into account this bias, Robins [4] and Satten [5] proposed to weight the observed data by the inverse of the probability of not dropping out of the risk set, in order to find a cohort which would have been seen without the withdrawals. Pohar Perme [6] used this idea to propose an unbiased estimator of net survival within the relative survival setting.

As long as informative censoring is accounted for appropriately, both cause-specific and relative survival approaches derive theoretically unbiased estimators of net survival. However these estimators are prone to biases related to the data settings within which each is derived. These biases are independent of the method of estimation.

In the cause-specific data setting, what defines a cancer-related death versus a death from another cause is reliant upon the judgment of the person extracting the information and often prone to misclassification. Several studies have described this bias as being non-negligible [7–14]. For this reason the relative survival method has generally been preferred to estimate net survival with population-based data [15,16]. However, within the relative survival data setting non-comparability between the cohort of patients and the general population [17] life tables used can also lead to bias. If a factor is differently distributed between patients and the general population, the resulting expected mortality of the cohort will be incorrectly estimated [18]. For instance, patients with lung cancer are more often smokers compared with the general population. Their expected mortality is therefore under-estimated as they are more likely to die from other causes than the general population [19]. In the long term, this under-estimation may be balanced by the selection process over time of the more robust patients, who may die less than the general population [20]. This may impact net survival estimates [21]. Similarly, several factors can be associated with both cancer mortality and other diseases and lead to non-comparability between observed and expected mortality.

Our objective was to compare the two data settings, cause-specific and relative survival, when estimating long-term net survival. Both are subject to bias as described above; either misclassification of the cause of death or use of inappropriate life tables. We first derived theoretically unbiased estimators by using weights for both approaches, which took into account informative censoring. We then performed two sensitivity analyses in order to examine which of the two data settings was more robust against incorrect assumptions. We used each estimator as a reference for the other in order to evaluate the impact on the net survival estimates (Table 1).

We used data from the Geneva Cancer Registry which holds high quality data on cancer patients collected since 1970. This enabled us to evaluate the effect of these biases on long-term net survival. Furthermore, it afforded a privileged situation for estimating net survival within the cause-specific setting as information on cause of death had been independently verified.

2. Material and methods

2.1. Data

The data were provided by the Geneva Cancer Registry.

The Geneva Cancer Registry collects information on incident cancer cases from various sources, including hospitals, laboratories and private clinics, all requested to report new cancer cases. Trained registrars systematically extract information from the medical records and conduct further investigation in the case of missing key data. The registry regularly assesses survival, taking as the reference date the date of confirmation of diagnosis or the date of hospitalization (if it preceded the diagnosis and was related to the disease). In addition to passive follow-up (standard examination of death certificates and hospital records), active follow-up is performed yearly using the files of the Cantonal Population Office who maintain a register of the resident population. The cause of death is validated or revised from death certificates by registrars using all available clinical information. Autopsy reports, letter at death written by general practitioners and all patients' medical notes are used for the assessment of the revised cause of death. The treatment can therefore be considered as breast cancer death when information is found about it being part of the morbid events leading directly to death [22]. We included all women diagnosed with an invasive primary breast cancer between 1981 and 1991. These women have all been followed-up for a minimum of 20 years, and the last date of follow-up was 31st December 2011.

2.2. Statistical methods

2.2.1. Informative censoring

Informative censoring in a cohort of cancer patients is a differential selection process which affects the likelihood of the event of interest being observed. Different strategies have been derived for each data setting and are able to take into account informative censoring when estimating net survival (Appendix A).

Table 1
Description of the two data settings available for the estimation of net survival.

	Setting	Net survival		
			Cause-specific	Relative survival
Biases	Theoretical/Methodological		Informative censoring	
	Data		Misclassification of the cause of death	Non comparability between the cohort and the general population
Solutions	Tackle informative censoring	Concept/Idea	Weight the net survival estimator with the expected mortality	
		Application	Use the cancer data to estimate the expected mortality	Use the expected mortality derived from general population expected mortality
	Check the extent of biases related to the data	Concept/Idea	Sensitivity analyses: Modify the data to check the robustness of the net survival estimate	
		Application	Modify the number of specific death	Modify the expected mortality rates of the general population