Cancer Epidemiology 53 (2018) 72–80

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

A new approach to estimate time-to-cure from cancer registries data

Olayidé Boussari[a,b,c], Gaëlle Romain[a,b], Laurent Remontet[d,e,f,g], Nadine Bossard[d,e,f,g], Morgane Mounier[h], Anne-Marie Bouvier[a,b], Christine Binquet[b,i,j], Marc Colonna[k], Valérie Jooste[a,b,*]

[a] Dijon-Bourgogne University Hospital, Registre Bourguignon des Cancers Digestifs, Dijon F-21000, France
[b] INSERM, U1231, EPICAD team, Univ Bourgogne-Franche-Comté, UMR 1231, Dijon F-21000, France
[c] LabEX LipSTIC, ANR-11-LABX-0021, Dijon F-21000, France
[d] Service de Biostatistique-Bioinformatique, Hospices Civils de Lyon, Lyon F-69003, France
[e] Université de Lyon, Lyon F-69000, France
[f] Université Lyon 1, Villeurbanne F-69100, France
[g] CNRS UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Équipe Biostatistique Santé, Pierre-Bénite F-69310, France
[h] Dijon-Bourgogne University Hospital, Univ Bourgogne-Franche-Comté, Registre des Hémopathies Malignes de Côte d'Or, Dijon, France
[i] INSERM, CIC1432, Clinical Epidemiology Unit, Dijon F-21000, France
[j] Dijon-Bourgogne University Hospital, Clinical Investigation Centre, Clinical Epidemiology/Clinical Trials Unit, Dijon F-21000, France
[k] Registre du Cancer de l'Isère, Grenoble University Hospital, Grenoble F-38000, France

ARTICLE INFO

Keywords:
Net survival
Cure models
Probability of being cured
Time-to-cure

ABSTRACT

Background: Cure models have been adapted to net survival context to provide important indicators from population-based cancer data, such as the cure fraction and the time-to-cure. However existing methods for computing time-to-cure suffer from some limitations.

Methods: Cure models in net survival framework were briefly overviewed and a new definition of time-to-cure was introduced as the time TTC at which P(t), the estimated covariate-specific probability of being cured at a given time t after diagnosis, reaches 0.95. We applied flexible parametric cure models to data of four cancer sites provided by the French network of cancer registries (FRANCIM). Then estimates of the time-to-cure by TTC and by two existing methods were derived and compared. Cure fractions and probabilities P(t) were also computed.

Results: Depending on the age group, TTC ranged from to 8 to 10 years for colorectal and pancreatic cancer and was nearly 12 years for breast cancer. In thyroid cancer patients under 55 years at diagnosis, TTC was strikingly 0: the probability of being cured was > 0.95 just after diagnosis. This is an interesting result regarding the health insurance premiums of these patients. The estimated values of time-to-cure from the three approaches were close for colorectal cancer only.

Conclusions: We propose a new approach, based on estimated covariate-specific probability of being cured, to estimate time-to-cure. Compared to two existing methods, the new approach seems to be more intuitive and natural and less sensitive to the survival time distribution.

1. Introduction

In some survival studies, a fraction of subjects never experience the event under study whatever the length of the follow-up period. In such cases, the observed survival curve levels off at a non-zero value which corresponds to the proportion of subjects who are free of failing from the event of interest and who are considered "statistically cured". Since their first formulation by Boag [1], cure models have been widely developed to analyse survival data with cure fractions (e.g., [2–5]).

In cancer population-based (such as registry-based) studies, the cause of death is often unreliable or unknown. Specific methods use the excess mortality rate (mortality due to cancer only) to estimate net survival without needing the cause of death [6,7]. Net survival is the survival in the hypothetical world where cancer would be the only cause of death [8,9]. When the net survival curve ends by flattening to a non-zero value, it is a typical case where cure models are useful for cancer survival analyses [10,11]. Several authors have extended cure models to the net survival framework [11–15]; we propose in

Abbreviations: CNS, conditional net survival; P, cure fraction; P(t), estimated covariate-specific cure probability; $S_n(t)$, net survival; $S_u(t)$, net survival of uncured subjects; T_{95}, time at which $S_u = 0.05$; T_{CNS}, time at which 5-year CNS = 0.95; TTC, time-to-cure

* Corresponding author at: Registre Bourguignon des Cancers Digestifs, UFR des sciences de santé BP 87900, Dijon cedex 21079, France.
E-mail address: valerie.jooste@u-bourgogne.fr (V. Jooste).

https://doi.org/10.1016/j.canep.2018.01.013
Received 6 September 2017; Received in revised form 19 January 2018; Accepted 21 January 2018

Supplementary file 1 an overview of these extensions. In these models, two quantities describe the net survival: i) the cure fraction P or proportion of cured subjects ($0 \leq P \leq 1$) and ii) the net survival time distribution $S_u(t)$ of uncured subjects described by a parametric or semi-parametric survival distribution function. Cure models assume that a given patient belongs to a given group (cured or uncured) since diagnosis. The probability that this patient belongs to the cured group is P at the time of diagnosis and increases as time elapses if the patient is still alive [16]. The time-to-cure, i.e. the time elapsed between diagnosis and cure is a useful indicator. Two definitions of the time-to-cure have been already proposed by the specialized literature [17,18] and are recalled in the next section.

In this work, we proposed a new definition for the time-to-cure, based on the patient's probability of being cured over time: This is a synthetic way to indicate the time at which patients can be reasonably confident to belong to the cured group and to attach a probability to such confidence. We applied flexible parametric cure models [19] to four cancer real datasets and derived the time-to-cure estimates by the new definition and by the two existing definitions. Then the results were compared and discussed.

2. Recall on the two existing approaches to estimate the time-to-cure

Chauvenet et al. [17] proposed to estimate the time-to-cure as the time at which "almost" all uncured patients would have died. From that time, the number of deaths attributable to the cancer of interest becomes negligible. It can be estimated as the time T_{95} at which 95% of the uncured would have died: i.e., $S_u(T_{95}) = 0.05$.

Dal Maso et al. [18] used the 5-year conditional net survival (CNS) to propose a definition for the time-to-cure: because it is considered that cure is reached when the net survival becomes nearly constant, it can be assumed that cure is reached when the conditional net survival becomes close to 1. Using the 95% cut-off, these authors defined the time-to-cure as the shortest time T_{CNS} after diagnosis at which the 5-year conditional net survival reaches 0.95; i.e. $S_n(T_{CNS} + 5)/S_n(T_{CNS}) = 0.95$.

3. A new definition for the time-to-cure

A work by Sposto [16] allows estimating for a given patient i (i.e. with characteristics x_i), the probability $P_i(t)$ of being cured at a given time t after diagnosis knowing that he/she was alive up to time t.

$P_i(t)$ = (probability of being cured and alive up to time t given x_i)/(probability of being alive up to time t given x_i)

We demonstrate in Appendix, that:

$$P_i(t) = P_i/S_{n,i}(t) = P_i/[P_i + (1 - P_i)S_{u,i}(t)]$$

where P_i (resp. $S_{u,i}$) denotes the proportion of cured patients (resp. the net survival function of the uncured patients) within the group of patients sharing the same characteristics as i.

This probability is an indicator that has been rarely used so far [20] but is interesting because: 1) as cure models assume that the two subpopulations (cured and uncured) are defined right at diagnosis, $P_i(t)$ corresponds to a dynamic prediction of a patient's probability of belonging to the cured group; 2) it is intuitive since, for a group of patients with same covariates x_i, it corresponds at time t, to the proportion of patients that belong to the cured group among still alive patients. Moreover, it is easy to see that $P_i(t = 0)$ is simply the cure proportion P_i and that $P_i(t)$ increases with t from P_i at diagnosis (t = 0) to 1.

We propose here to estimate the time-to-cure by the time $t = TTC_i$ from which $P_i(t) \geq 0.95$. From this definition it can be deduced that when $P_i \geq 0.95$, $TTC_i = 0$. In the sequel, 'i' will be omitted for easier reading.

4. Illustrative application of the new approach to real cancer data

4.1. Materials and methods

The French network of cancer registries (FRANCIM) and the Service de Biostatistique-Bioinformatique (Hospices Civils de Lyon, France) are currently maintaining a common database started in 1975 that counts currently more than a million patients diagnosed with any of thirty cancer types [21]. Quality controls are performed by each registry and on the whole database with tools provided by the International Agency for Research on Cancer.

We illustrated the new approach with data from four cancer sites (colon-rectum, pancreas, breast and thyroid, of which respectively 46 371, 8169, 71 947 and 8762 cases were included) showing different settings of the dynamics of the excess hazard and for which cure assumption could be accepted. These cancer sites were chosen because they illustrate both long-time survival cancer, such as thyroid cancer, and short survival cancer, such as pancreatic cancer. We included all patients aged 15 to 74 years at diagnosis and diagnosed between 1995 and 2010. They were followed-up until June 30, 2013.

The analyses were conducted separately in men and women and age was included as a categorical covariate. Age groups were: [15–45[, [45–55[, [55–65[, and [65–75[years for colorectal, thyroid, and breast cancers. For pancreatic cancer the number of cases aged [15–45[was low (357 cases), we combined this group with the [45–55[leading to [15–55[, [55–65[, and [65–75[age groups. Breast cancer was considered in women only. Table 1 summarizes the baseline data.

For each site, the net survival was estimated by fitting a flexible parametric survival model [22]: the logarithm of the baseline cumulative excess hazard is written as a restricted cubic spline of the log time with four internal knots placed at the 25th, 50th, 75th and 95th centiles of the observed death times and the boundary knots placed at the 1st centile of the observed death times and 17 years. We introduced age at diagnosis (using the previously defined age groups) as a covariate. We evaluated time dependent effect of age at diagnosis through an interaction term with spline of time (2 internal knots located at the 33rd and 67th centiles of the observed death times), using a likelihood ratio test with 0.05 significance level. We checked graphically that the excess hazard derived from this model approached zero (i.e., the associated net survival reached a plateau) and then the net survival was modelled using a flexible parametric non-mixture cure model [19], according to the recommendations of Yu et al. [23]. The splines considered for this model used the same knots as that in the previous model with an additive knot at the 99th centile of the observed death times. Note that separate models were fitted for each sex for eligible site. Estimates of the net survival flexible parametric non-mixture cure model were validated by comparing them with the non-parametric estimates obtained from the Pohar-Perme estimator [24]. Results of the graphical checking of the cure assumption and the validation of the net survival estimates by the cure model are provided in Supplementary file 2 (Figs. S1–S4). Cure assumption was accepted for colorectal, thyroid and pancreatic cancers. For women with breast cancer, although the plateau in net survival was not obvious, the excess mortality rate was very low and survival curves with and without cure assumption were very close to each other for all age groups except 65–74 years (Fig. S3). Cure assumption was accepted for women under 65 years at diagnosis.

Based on the regression coefficients from the flexible parametric non-mixture cure model we computed P, $S_u(t)$ and $S_n(t)$, and then derived $P(t) = P/S_n(t)$. Thereafter, estimates of TTC and T_{95} were obtained using a Newton-Raphson technique. T_{CNS} estimates were derived by evaluating $CNS(t) = S_n(t + 5)/S_n(t)$ at yearly intervals (starting from t = 0). For each cancer site considered, the values of P(t), $S_u(t)$, and CNS(t) were compared at t = TTC, T_{95}, and T_{CNS} to better examine the three definitions of the time-to-cure.

In order to check the sensitivity of TTC to the cut-off changes for the four illustrative cancer sites, TTC was estimated for 3 different cut-off