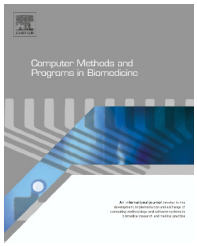
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 112 (2013) 343–355

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data

Edson Z. Martinez[a,*], *Jorge A. Achcar*[a], *Alexandre A.A. Jácome*[b], *José S. Santos*[c]

[a] *Department of Social Medicine, University of São Paulo (USP), Ribeirão Preto School of Medicine, Brazil*
[b] *Department of Medical Oncology, Barretos Cancer Hospital, Str. Antenor Duarte Villela, 1331 Barretos, Brazil*
[c] *Department of Surgery and Anatomy, University of São Paulo (USP), Ribeirão Preto School of Medicine, Brazil*

ARTICLE INFO

Article history:
Received 20 February 2013
Received in revised form
17 July 2013
Accepted 20 July 2013

Keywords:
Survival analysis
Bayesian analysis
Generalized modified Weibull distribution
Cure fraction model
Gastric cancer

ABSTRACT

The cure fraction models are usually used to model lifetime time data with long-term survivors. In the present article, we introduce a Bayesian analysis of the four-parameter generalized modified Weibull (GMW) distribution in presence of cure fraction, censored data and covariates. In order to include the proportion of "cured" patients, mixture and non-mixture formulation models are considered. To demonstrate the ability of using this model in the analysis of real data, we consider an application to data from patients with gastric adenocarcinoma. Inferences are obtained by using MCMC (Markov Chain Monte Carlo) methods.

1. Introduction

In the lifetime data analysis, researchers commonly use standard non-parametrical techniques, such as Kaplan–Meier estimators or log-rank test [1], semi-parametrical models (for example, proportional hazards model in presence of covariates [2] or standard parametrical models using some popular lifetime distributions [3]. One of the distributions widely used in cancer research is the Weibull distribution [4], mainly due to the flexibility of its hazard function and the facility to estimate its parameters. However, in medical lifetime research, we usually have data sets which require more sophisticated parametric models. To achieve this goal, new classes of parametric distributions based on extensions of the Weibull distribution have been introduced in the literature. As special cases, we have the exponentiated Weibull (EW) [5,6], the generalized modified Weibull [7] and the log-beta Weibull distributions [8]. In addition, other common situation in the analysis of time-to-event data, particularly in cancer research, occurs when it is expected that a fraction of individuals will not experience the event of interest. In this case, it is assumed that the

* Corresponding author. Tel.: +55 1636022569.
E-mail addresses: edson@fmrp.usp.br, edson.martinez@ig.com.br (E.Z. Martinez).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.07.021

studied population is a mixture of susceptible individuals who experience the event of interest and non-susceptible individuals that supposedly will never experience it. The presence of immune or cured individuals in a data set is usually suggested by a Kaplan–Meier plot of the survival function, which shows a long and stable plateau with heavy censoring at the extreme right of the plot [9]. Different parametric and non-parametric approaches have been considered to model the proportion of immunes and interested readers can refer, for example, to Boag [10], Berkson [11], Haybittle [12], Meeker [13], Gamel et al. [14], Ghitany and Maller [15], Copas and Heydary [16], Ng and McLachlan [17], De Angelis et al. [18], Peng and Dear [19], Lambert et al. [20] and Yu et al. [21]. In addition, Bayesian inference methods for survival data with a surviving fraction were introduced by some authors such as Castro et al. [22], Chen et al. [23], Ibrahim et al. [24], Kim et al. [25] and Seltman et al. [26]. As a motivation for this paper, we consider a gastric cancer lifetime data introduced by Jácome et al. [27]. For a statistical analysis of this data set, we assume the four-parameter generalized modified Weibull distribution (GMW) [7] in presence of cure fraction, censored data and covariates. We implemented the statistical model under a Bayesian framework, where the parameter estimation is based on Markov Chain Monte Carlo (MCMC) techniques. We organize the rest of the paper as follows. The gastric cancer data set is described in Section 2. In Section 3, we describe the mixture and non-mixture cure fraction models, the GMW distribution [7] and some of their special cases. In this section we also introduce the formulation of the likelihood functions considering mixture and non-mixture cure fraction models based on the GMW distribution. The Bayesian analysis for the proposed models is described in Section 4. The obtained results of the Bayesian analysis for this medical data set, considering the proposed mixture and non-mixture models, are presented in Section 5. Finally, in Section 6, we present a discussion of the obtained results.

2. The gastric cancer data

Gastric cancer is one of the leading causes of cancer-related death [28] and the mucosal resection is accepted as a treatment option for early cases of the disease. In a review of the literature [29], it was found that the 5-year survival rate following all type of resections has increased significantly from 20.7% before 1970 to 28.4% before 1990. In addition, the 5-year survival rate following curative or radical resection has risen from 37.6 to 55.4% over the same period. Thus, new technologies to optimize medical decisions and the development of new therapies are of great importance to improve survival in gastric cancer. Jácome et al. [27] conducted a retrospective study in patients with gastric adenocarcinoma who underwent curative resection with D2 lymphadenectomy in the Barretos Cancer Hospital (Hospital de Câncer de Barretos, Brazil) between January 2002 and December 2007. The effectiveness of lymphadenectomy for cure in patients with early gastric cancer and lymph node metastasis is discussed by Okamura et al. [30]. It is known that adjuvant chemoradiotherapy (CRT) is the standard treatment in Western countries for gastric cancer patients submitted to curative resection. Aiming a more precise evaluation of the treatment, Jácome et al. [27] considered 185 patients with stage II to IV gastric adenocarcinoma with no distant metastases and compared the 3-year overall survival of the two treatments, that is, adjuvant CRT versus resection alone. In the present article, as an illustration for the use of the GMW distribution, we consider the entire data set obtained from this study, considering 201 patients of different clinical stages. Table 1 shows this data set, which includes 76 patients that received adjuvant CRT and 125 that received resection alone. The data in this table refer to the times until death in months since surgery, where a plus symbol (+) indicates censored data. We observe that we have 53.2% of censored data, that is, 57.9% if we consider the patients treated with CRT and 50.4% if we consider the patients treated with resection alone.

The Kaplan–Meier estimate of the survival function for the gastric cancer data is given in Fig. 1, where the presence of a plateau near to 0.5 observed in the graph presented in panel (a) suggests that models that ignore the proportion p of long-term survivors will not be suitable for these data. The graph presented in panel (b) of Fig. 1 describes the empiric survival functions for each type of treatment, where the presence of stable plateaus at the right tail of the plot also assures the adequacy of the cure fraction model approach.

3. Models

3.1. *Mixture and non-mixture cure fraction models*

Following Maller and Zhou [31], a mixture model for lifetime data sets assumes that the probability of the time-to-event to be greater than a specified time t is given by the survival function

$$S(t) = p + (1-p)S_0(t), \tag{1}$$

where p is a parameter which represents the proportion of "long-term survivors" or "cured patients", regarding the event of interest ($0<p<1$), and $S_0(t)$ is the baseline survival function for the susceptible individuals [10]. Common choices for $S_0(t)$ are the Gompertz, exponential and Weibull distributions. The probability density function for the lifetime T is

$$f(t) = \frac{dF(t)}{dt} = (1-p)f_0(t),$$

where $F(t) = 1 - S(t)$ and $f_0(t)$ is the baseline probability density function for the susceptible individuals. Considering a random sample (t_i, δ_i) of size n, $i = 1, \ldots, n$, the contribution of the ith subject for the likelihood function is given by

$$L_i = [f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i} = [(1-p)f_0(t_i)]^{\delta_i}[p + (1-p)S_0(t_i)]^{1-\delta_i},$$

where δ_i is a censoring indicator variable, that is, $\delta_i = 1$ for an observed lifetime and $\delta_i = 0$ for a censored lifetime. Alternatively, a non-mixture formulation has been suggested by several authors [32,33]. This model defines an asymptote for