STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:3959–3975
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2394

Modelling intervention effects after cancer relapses

Juan R. González[1,*,†], Edsel A. Peña[2] and Elizabeth H. Slate[3]

[1]*Cancer Prevention and Control Unit, Catalan Institute of Oncology, Avda. Gran via s/n, km. 2.7, Hospitalet de Llobregat 08907, Spain*
[2]*Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.*
[3]*Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina, Charleston, SC 29425, U.S.A.*

SUMMARY

This article addresses the problem of incorporating information regarding the effects of treatments or interventions into models for repeated cancer relapses. In contrast to many existing models, our approach permits the impact of interventions to differ after each relapse. We adopt the general model for recurrent events proposed by Peña and Hollander, in which the effect of interventions is represented by an effective age process acting on the baseline hazard rate function. To accommodate the situation of cancer relapse, we propose an effective age function that encodes three possible therapeutic responses: complete remission, partial remission, and null response. The proposed model also incorporates the effect of covariates, the impact of previous relapses, and heterogeneity among individuals. We use our model to analyse the times to relapse for 63 patients with a particular subtype of indolent lymphoma and compare the results to those obtained using existing methods. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: recurrent events; effective age process; intensity models; cancer recurrence model

1. INTRODUCTION

Cancer prognostic models for overall (time until death) or for disease-free survival (time until relapse or progression) are very useful for patient management. In particular, there are

*Correspondence to: Juan R. González, Genes and Disease Program, Center for Genomic Regulation, Biomedical Research Park of Barcelona, Passeig Marítim, 37-49, 08003 Barcelona, Spain.
†E-mail: juanramon.gonzalez@crg.es

Contract/grant sponsor: Carlos III Institute, Spanish Ministry of Health; contract/grant numbers: C03/10, C03/09
Contract/grant sponsor: NSF; contract/grant number: DMS 0102870
Contract/grant sponsor: NIH; contract/grant numbers: GM056182, CA077789, 17-02-1-0138
Contract/grant sponsor: NIH COBRE; contract/grant numbers: RR17698, RR17696
Contract/grant sponsor: USC/MUSC Collaborative Research Program
Contract/grant sponsor: DAMD; contract/grant number: 17-02-1-0138
Contract/grant sponsor: DOD; contract/grant number: GC-3532-03-42153CM

Received 22 August 2005
Accepted 31 August 2005

some indolent types of cancers for which patients have long survival but experience multiple relapses, and wherein the study of factors related to time until progression is of interest because most patients die from causes related to the disease [1, 2]. Thus, estimation of the risk of recurrence allows for better planning of follow-up schedules after diagnosis or first treatment, and permits clinicians to determine therapeutic approaches based on the patient's risk of relapse.

Although there are many survival models that handle recurrent event data [3–5], several prognostic studies in major cancer and epidemiologic journals estimate the risk of relapse using only information about the time until first occurrence (follicular lymphomas [6, 7], acute leukaemias [8], colorectal cancer [9], or breast cancer [10], among others). This approach ignores the information of subsequent relapses, and hence statistical inference tends to be inefficient. To improve efficiency, analyses of other cancer studies, such us mammary tumours for rats [11] or patients with superficial bladder cancer [12, 13], rely on variants of Cox's proportional hazards model that handle recurrent event data [3]. These models deal with the additional issue of intra-subject correlation. For example, Wei, Lin and Weissfeld's (WLW) [14] marginal model, Prentice, Williams and Petersen's (PWP) [15] conditional model, and Andersen and Gill's (AG) [16] intensity-based model account for unobserved heterogeneity via robust estimation of the variance–covariance matrix. On the other hand, Oakes' [17] frailty model introduces an unobservable random covariate that induces dependence among a subject's recurrent event times. Another important aspect in recurrent event settings is the event dependence produced by a learning process or by the biologic weakening (or possibly strengthening) of the body. To handle this dynamic aspect, AG used time-dependent covariates, and both WLW and PWP models used stratification. A Markovian model for tumour occurrences proposed by Gail, Santner and Brown [11] (GSB) alternatively incorporates the impact of accumulating event occurrences. This is a special case of the general model of Peña and Hollander [18], which is the basis of the specialized model considered here.

Interventions performed on the subject after each event occurrence may also modify the event occurrence intensity. For example, when a subject suffers a heart attack he/she is advised to modify his/her existing lifestyle (e.g. eating habits; reduce stress level; increase physical activity). So, it seems reasonable to assume that patients who change their lifestyle will have less probability of suffering another relapse than those who do not follow the physician's recommendations. In our case, patients with cancer are treated after observing a progression of the tumour. In the particular case of indolent lymphomas, which are non-curable, the therapy aims to extend the time until next relapse. After therapy, the patient is monitored for disappearance of cancer or related symptoms. Patients whose disease disappears are improved in the sense of having less probability of relapsing than those where little or no response is observed. In the reliability literature, this change in the probability of relapse is captured by an adjustment (a reduction in this case, presumably) to the *effective age* of the patient.

Many reliability models incorporating an effective age process have been proposed [19–22]. Unfortunately, this concept is little used in biomedical or public health contexts. Perhaps a major reason for this is that it is difficult to obtain information about the effective age process in these settings. For example, it seems difficult to obtain information about the effect of physician's interventions in a heart attack situation. However, in patients with cancer, the degree of response to treatment following each relapse provides information about the effective age process. Moreover, this information is readily available because cancer patients

are usually well monitored due to the importance of current disease status in determining appropriate therapy.

The importance of the effective age process in biomedical settings, and, in particular, for cancer data, was partly demonstrated in a technical report by Peña, Slate and González (2004) (available at http://www.stat.sc.edu/rsrch/techrep/ or upon request from the authors) using a real bladder tumour data set [13] also analysed by Wei *et al.* [14] and Therneau and Grambsch [3]. They compared their results obtained using AG, WLW, and PWP models to those obtained using a general class of models under two specifications of the effective age process. They observed that the methods produce different estimates of covariate effects, which could possibly lead to contradictory conclusions, but pointed out that the notion of an effective age could potentially explain these differences. They also conducted an extensive simulation study and found that mis-specification of the effective age process leads to biased estimates of the baseline hazard rate function and the effect of the prior number of events. In addition, regarding indolent tumours, McLaughlin [23] argues that it is well known that the impact of therapy after relapse is a significant prognostic factor for occurrence of the next relapse [24, 25] and concludes that 'it is necessary [that] a model designed specifically for relapsing patients' be utilized.

We describe a model for cancer relapse data that incorporates the effects of intervention after each recurrence and also accommodates the effect of covariates, the impact of the number of previous relapses, and heterogeneity among individuals. We use the general recurrent event model proposed by Peña and Hollander [18], which simultaneously incorporates all effects mentioned above, but define an effective age process tailored specifically for cancer relapse data. Section 2 describes our notation and the general class of models that we will adapt for the cancer relapse setting. This section also describes briefly how this class of models subsumes some existing models for recurrent event data that have been used in biomedical settings. We define our proposed effective age process for cancer relapse settings in Section 3. A simulation study is performed to study the model behaviour under effective age mis-specification. The simulation results and design are discussed in Section 4. Section 5 illustrates the use of the model with real data from low-grade lymphomas. Section 6 provides concluding remarks.

2. NOTATION AND MODEL

We consider a patient who is being monitored for cancer relapses over a study period $[0, \tau]$, where τ may represent an administrative time, time of study termination, or some other right-censoring variable such as death. The time τ could be random governed by an unknown probability distribution function $G(t) = \Pr(\tau \leqslant t)$. In general, event occurrence times can be indexed in two scales: calendar and gap times. Calendar times are defined by the sequence $S_0 \equiv 0 < S_1 < S_2 < S_3 < \ldots$, and in our case correspond to the successive calendar times of cancer relapses. Gap times will be denoted by $T_1, T_2, T_3, \ldots$ and correspond to the time between successive cancer relapses. Thus, for $i = 1, 2, 3, \ldots$,

$$T_i = S_i - S_{i-1} \quad \text{and} \quad S_i = T_1 + T_2 + \cdots + T_i$$

Over the observation period $[0, \tau]$, the number of cancer occurrences is

$$K = \max\{k \in \{0, 1, 2, \ldots\} : S_k \leqslant \tau\}$$

a random variable whose distribution depends on the distributional properties of the relapse times T_is and the distribution G of τ. As such, K is informative regarding the distributional properties of cancer occurrences.

The occurrence of a new relapse may also be affected by covariates. For a patient, we observe a, possibly time-varying, q-dimensional vector of covariates such as gender, age, race, disease status, beta-2 microglobulin level, treatment regimen, etc. We denote this covariate process by $\{\mathbf{X}(s)=(X_1(s),X_2(s),\ldots,X_q(s))' : 0 \leqslant s \leqslant \tau\}$, where ' ′ ' represents vector/matrix transpose.

In addition, in cancer settings, after interventions are administered upon relapses, information about patient status may be obtained. Examples of interventions are chemotherapy, radiotherapy, and bone-marrow transplant, among others. We denote the subsequent patient status information by a vector $\boldsymbol{\psi}=(\psi_1,\psi_2,\ldots,\psi_K)'$, where ψ_j signifies a certain type of response to the intervention after the jth relapse. This will be explained in more detail later.

Consequently, if in the study there are n patients, over the period $[0,\tau^*]$ where $\tau^* \equiv \max_{i \leqslant n} \tau_i$, we will have the following data:

$$\mathbf{D}(\tau^*) \equiv \{[(\mathbf{X}_i(s) : 0 \leqslant s \leqslant \tau_i), \boldsymbol{\psi}_i, K_i, \tau_i, S_{i1}, S_{i2}, \ldots, S_{iK_i}, \tau_i - S_{iK_i}], i = 1,2,\ldots,n\}$$

For each $i=1,2,\ldots,n$, with $I\{A\}$ denoting indicator function of event A, we define the counting process $N_i^{\dagger}(s)=\sum_{j=1}^{K_i} I\{S_{ij} \leqslant s\}$, which records the number of events experienced by patient i at or before calendar time s, and the 'at-risk' process $Y_i^{\dagger}(s)=I\{\tau_i \geqslant s\}$.

Peña and Hollander's class of models [18] specifies the hazard rate process for patient i at calendar time s conditionally on an unobserved frailty, Z_i, and observed covariates, $\mathbf{X}_i$, via

$$\lambda_i(s \mid Z_i, \mathbf{X}_i) = Z_i \, \lambda_0[\mathscr{E}_i(s)] \, \rho[N_i^{\dagger}(s-); \alpha] \, \Phi[\beta' \mathbf{X}_i(s)] \tag{1}$$

Here, $\lambda_0(\cdot)$ is an unknown baseline hazard rate function; $\rho(\cdot;\alpha)$ is a known functional form depending on unknown parameter α that encodes the effects of accumulating event occurrences; and β is a vector of parameters associated with covariates effects. This model also incorporates the effect of performed interventions through the effective age, $\mathscr{E}_i(s)$, which serves as the argument to the baseline hazard rate function. This new class of models connects the reliability perspective of modelling the effect of intervention through the effective age with the biomedical concern of incorporating the effects of concomitant covariates, as well as capturing correlation among interoccurrence times and heterogeneity among patients. In this paper, $\rho(j;\alpha)=\alpha^j$, and as in Cox's model, $\Phi(x)=\exp(x)$. The frailty, Z_i, follows a gamma distribution with mean one and unknown variance ξ^{-1}.

As demonstrated by Peña and Hollander [18] and Peña *et al.* (2004), this class of models subsumes many existing models. For example, the PWP conditional model of Prentice *et al.* [15] can be formulated to address either time since a patient entered the study or time since the last relapse, and each of these corresponds to (1) with $\alpha=1$, $Z_i=1, i=1,\ldots,n$, and a particular choice for $\mathscr{E}_i(s)$. Specifically, $\mathscr{E}_i(s)=s$ in the PWP model for time since entry, and $\mathscr{E}_i(s)=s-S_{iN_i^{\dagger}(s-)}$ for time since last relapse. The calendar time formulation assumes that all interventions produce a 'minimal repair' or no improvement in the patient. In medical terms, the disease is proceeding in a stable manner. Alternatively, the gap time formulation assumes

that all interventions lead to perfect recovery for the patient, i.e. the disease disappears, which is known as a complete remission (CR) in the cancer literature.

In the PWP models, and also in the AG and WLW models used in biomedical settings [14–16], the implicit effective age function is assumed to be the same for each relapse. However, the effect of intervention after cancer relapse varies not only from patient to patient but also for multiple relapses for the same patient. Hence, there is a need to incorporate the effect of intervention upon relapses via a more general effective age that accounts for the response to therapy.

3. EFFECTIVE AGE PROCESS FOR CANCER DATA

The idea of *effective age*, also called *virtual age*, which encodes the effect of repair appears in the study of repairable systems in reliability settings; see, for instance, Reference [26]. These models are more common in this area than in biomedical settings because it is a relatively easier task in reliability settings to assess this effective age. For instance, upon failure of a reliability system, we may decide to replace the system by a new identical system (perfect repair), or we may simply restore it to a state just before the failure (minimally repaired). However, when dealing with patients, we do not 'repair' them, rather we treat them; and, in addition, physicians always want to eliminate all manifestation of the disease, that is, obtain a CR, which is analogous to a 'perfect' repair.

Interventions in biomedical settings can take many forms. For example, as mentioned in the Introduction, when a subject suffers a heart attack he/she is advised to alter his/her existing lifestyle (e.g. eating habits, stress level, physical activity). Similarly, in the case of alcoholism, intervention in the form of psychological methods (e.g. confinement or enforced hospitalization; correction of faulty home environment), physiological methods (e.g. elevation of blood sugar level; convulsive therapy), or through family-based or institutional-based methods (e.g. closer supervision by family members; Alcoholics Anonymous) is performed. In cancer, intervention may be removal of a tumour or some prophylactic or curative treatments (e.g. consolidation chemotherapy, radiotherapy, bone-marrow transplant). All these interventions may modify the probability distribution of the next event reoccurrence.

Patients with indolent lymphomas, and in general with cancer, are monitored from the date of diagnosis to the time of death or loss to follow-up. At any evaluation, CR is defined as the disappearance of tumour masses and disease-related symptoms, as well as the normalization of the previous test and/or biopsies, lasting for at least one month. Partial remission (PR) is said to occur when measurable lesions have decreased by at least 50 per cent. Patients not included in these categories are called non-responders (NR) [27]. It is well known that the response to treatment is related to the time until the next relapse [24, 28, 29], so it is reasonable to have a model that incorporates this information. However, neither AG, PWP, WLW, GSB, nor frailty models, which are mostly used in biomedical settings, have incorporated the effect of performed interventions upon event reoccurrences, though some of these methods could accommodate such information through the use of time-dependent covariates as will be illustrated in the example presented in Section 5.

We propose using the response to therapy, defined as CR, PR, or NR, to define a model for the effective age for relapsing patients as follows. Consider a single patient and let

$\{A_j : j = 0, 1, 2, \ldots\}$ be a sequence satisfying

$$A_0 = 0, \quad A_j = A_{j-1} + \left(\prod_{k=1}^{j} [1 - \psi_k] \right) T_j, \quad j \geqslant 1 \tag{2}$$

where $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_K)'$, with $\psi_j \in \{0, 0.5, 1\}$ and with the interpretation that $\psi_j = 0$ means that an NR (non-response) has occurred after the jth relapse, $\psi_j = 1$ means that a CR (perfect intervention) has occurred, while $\psi_j = 0.5$ means that a PR (partial remission) has transpired. The values of the ψ_js can be assessed by the clinician(s) monitoring the patient. Our proposed effective age process for cancer relapse is

$$\mathscr{E}(s) = A_{N^\dagger(s-)} + (s - S_{N^\dagger(s-)}) \tag{3}$$

The effective age (3) is a particular case of Kijima's model II [30], where in his model the ψ_js are assumed to take any values in $[0, 1]$, whereas in our model we assume that they only take three possible values. Kijima's aim was to model the situation where after each failure some repair is performed and the effectiveness of this repair could be quantified by a number that is between zero and one. We note that in cancer problems, we may also assess this 'degree' of response according to a number in $[0, 1]$; however, in realistic settings clinicians only need to know if a CR, PR, or NR was achieved to make a good therapy determination, hence our restriction of the possible values of the ψ_js to the set $\{0, 0.5, 1\}$. Dorado *et al.* [31] also studied effective age functions that encompass the form (3). Note that if all responses are CR, i.e. $\psi_i = 1, i = 1, 2, \ldots$, then the effective age corresponds to gap time formulation, $\mathscr{E}(s) = s - S_{N^\dagger(s-)}$ since all A_js in (2) become 0. Similarly, if all responses are NR, the effective age corresponds to a calendar time formulation, $\mathscr{E}(s) = s$.

To demonstrate the notion of an effective age, Figure 1 shows the effective age for a patient in a cancer study. Between 0 (or S_0) and S_2, the process depicted corresponds to $\mathscr{E}(s) = s$ (calendar or elapsed time formulation). At the first event, which occurred at time S_1, treatment or intervention did not improve the disease status. In medical parlance, the patient did not respond to treatment, i.e. NR is achieved. After the second event, which occurred at S_2, the patient responds perfectly to treatment, achieving a complete resolution of all clinical manifestation of the disease. It is considered a CR. In this case, the effective age corresponds to $\mathscr{E}(s) = s - S_{N^\dagger(s-)}$ (backward recurrence time). However, after the third event at time S_3, the patient reverts to a state between a CR and a NR, that is, the patient experiences a little improvement or a PR to treatment. Finally, a progressive disease is observed for the fourth relapse at S_4, possibly due to some complications. The fifth failure which would have happened at S_5 is not observed since the end of observational period τ for this hypothetical patient is less than S_5. Consequently, the gap time for the fifth event is right-censored by $\tau - S_4$. The effective age process (2) that we adopt for the cancer relapse model does not permit the progressive disease scenario, as shown following S_4 in Figure 1.

We will refer to the general recurrent event model (1) with effective age in (3) as a dynamic *cancer* model. Because the cancer model is a special case of the general recurrent event model considered by Peña and Hollander [18], the procedures for estimating the parameters of this general class of models apply to the cancer model. We refer the reader to Peña *et al.* (2004) for details of the semi-parametric maximum likelihood estimation procedure. The authors have developed an `R` package called *gcmrec* [32] that implements these estimation procedures. The *gcmrec* manual describes how to fit the models used in Section 5.

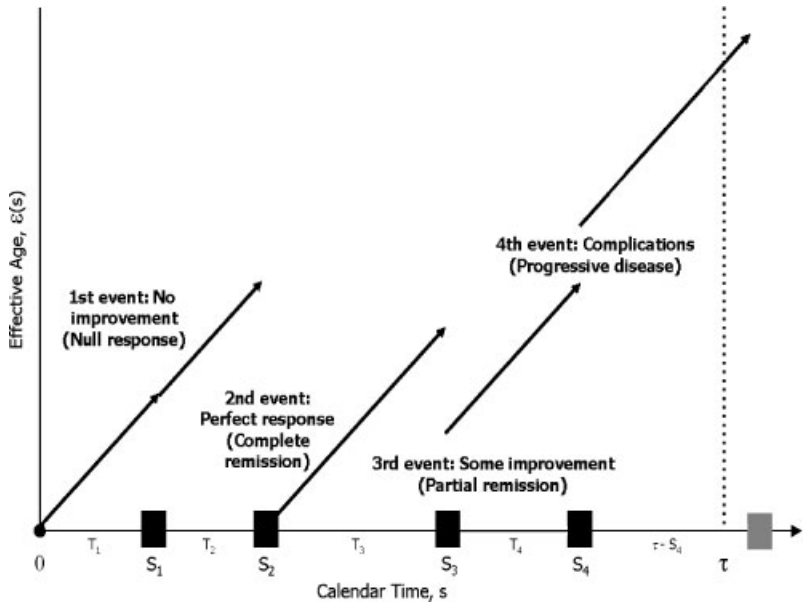

Figure 1. Pictorial representation of effective age *versus* calendar time for a hypothetical unit.

4. SIMULATION STUDY

4.1. Simulation design

We have carried out simulation studies to examine properties of estimators of parameters in the cancer relapse model with the primary goal of ascertaining the sensitivity or robustness of estimators to effective age function mis-specification. The data have been generated by the model allowing for different responses after each intervention, but with the resulting data analysed by assuming that the patients always achieve the same response (that is, either always CR or always NR) after each tumour reoccurrence.

We mimicked the simulation study performed by Peña *et al.* (2004). We point out that because their simulation was meant to cover both reliability and biomedical settings, some of the parameter values they considered were not realistic for medical settings. For our simulation, we let n take values in $\{30, 50\}$. The censoring variables τ_i, $i = 1, \ldots, n$, are generated to have, on average, approximately 5 events per patient (in contrast, Peña *et al.*'s simulation has 10 events per unit, on average). For the baseline hazard function λ_0, we choose the flexible Weibull distribution, with unit scale parameter and shape parameter, γ, taking values in $\{0.9, 2\}$. The impact of the accumulating number of relapses is assumed to be of form $\rho(k, \alpha) = \alpha^k$ with α taking values in $\{0.9, 1, 1.05\}$. We have simulated a two-dimensional covariate vector (X_1, X_2). To have both categorical and continuous covariates, X_1 has been simulated to have a Bernoulli distribution with success probability 0.5 and X_2 was set to have a standard normal distribution. These covariates are stochastically independent. The regression coefficient vector (β_1, β_2) was set to $(1, -1)$. Finally, the frailty component was generated

under a gamma distribution with unit mean and variance $1/\xi$ with ξ taking values in $\{2, 6, \infty\}$, the value of ∞ corresponding to the absence of frailties (further details are given by Peña *et al.* (2004)).

For the simulation, we considered an effective age function corresponding to the cancer model. That is, a patient can achieve a complete, a partial, or a null response depending on the vector ψ. We have assumed three different scenarios according to the following probability functions for ψ which takes values in the set $\{1 = \text{CR}, 0.5 = \text{PR}, 0 = \text{NR}\}$: $\{(0.8, 0.1, 0.1), (0.3, 0.5, 0.2), (0.1, 0.2, 0.7)\}$. Thus, in the first case, we assume that patients achieve CR with a probability of 80 per cent, and PR or NR 10 per cent of the time, respectively. These three sets of distributions allow us to cover three different scenarios: the first assumes that in a large majority of cases, perfect response is achieved after each relapse; the third has minimal response predominating; and the second distribution is an in-between scenario. For notation in the sequel, when we write $p(\psi) = (p_1, p_2, p_3)$, this indicates that the ψ values are chosen such that $1(= \text{CR})$ occurs with probability $p_1, 0.5$ $(= \text{PR})$ occurs with probability p_2, and 0 $(= \text{NR})$ occurs with probability p_3.

For each of the 324 combinations of parameter values, we performed 1000 replications. To study mis-specification of the effective age function, we have also estimated the parameters involved in (1) assuming a model with an effective age function that considers that patients always achieve a perfect response (CR) and another where patients always achieve a minimal response (NR). Through these simulations we are able to highlight the importance of the effective age function in relation to either under- or over-estimation of the baseline survivor function.

4.2. Simulation results

In the discussion of the simulation results, we focus on the consequences of analysing data from the cancer model using models that always assume the same response, either always perfect or always minimal. Regarding distributional properties of the estimators of α, β, and $v \equiv \xi/(1 + \xi)$, and the estimator of the baseline survivor function when the correct model is utilized, the simulation results reveal the same conclusions as those obtained in the simulation studies by Peña *et al.* (2004), hence we do not dwell on these but instead refer the reader to this technical report.

Figure 2 shows the estimated baseline survivor function, bias, and root-mean-squared error (rmse) curves calculated under effective age mis-specification, showing the effect of different effective age functions chosen (always minimal or perfect response), the impact of different ξ values, and for the three different cancer models analysed. The results are for $\alpha = 0.9$, $\gamma = 0.9$ and $n = 30$.

As one may expect, and as Peña *et al.* (2004) also observed, when there is no mis-specification and when the sample size increases, the performance of the estimators of the finite-dimensional parameters and the baseline survivor function improved, as can be seen by noting that both bias and standard error decrease. We also notice that when the sample size is small, there is considerable over estimation of v.

On the other hand, under effective age mis-specification, the estimators of the finite-dimensional model parameters and the baseline survivor function are highly biased (see Figure 2). The parameter that controls event dependence, v, as well as the parameters associated with the covariates, β, are more biased than when the correct model is used. The

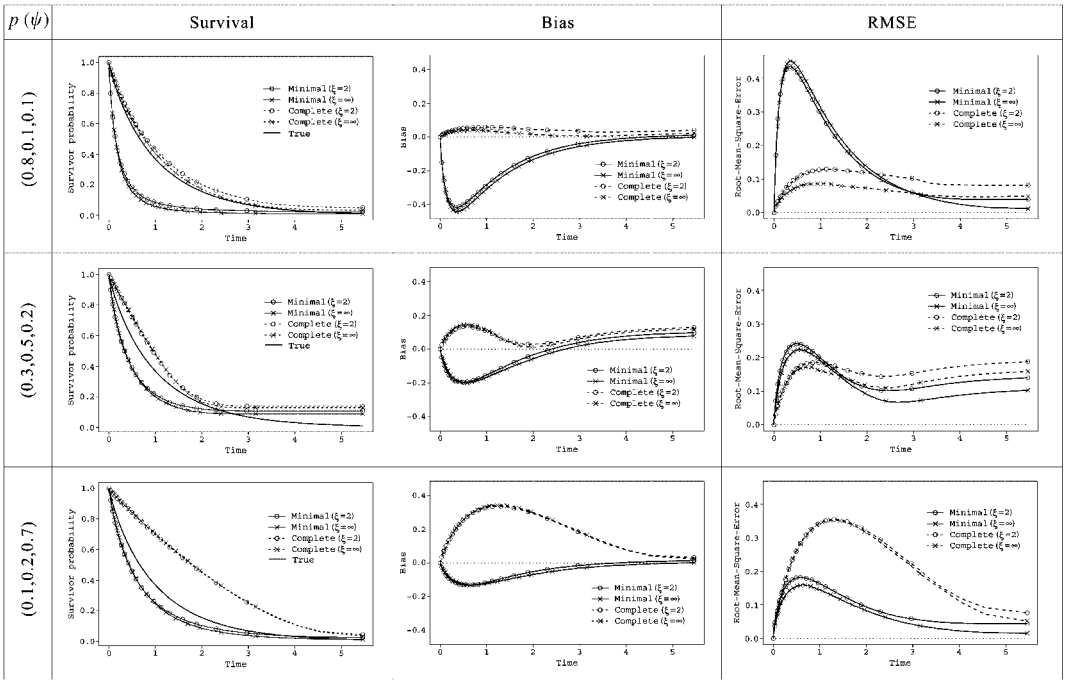

Figure 2. Estimated baseline survivor function, bias and root-mean-squared error curves for the estimator of the baseline survivor function as both the effective age $\mathscr{E}_i(s)$ (always minimal response, $\mathscr{E}_i(s)=s$, and always perfect response, $\mathscr{E}_i(s)=s-S_{N_i^\dagger(s-)}$) and the frailty parameter ξ varies ($\xi=2$ and $\xi=\infty$). This case corresponds to $\alpha=0.9$, $\gamma=0.9$ and $n=30$.

bias observed in the baseline survivor function is highly evident when the baseline hazard rate function is increasing, that is, when $\gamma=2$. Regarding the survivor curves in Figure 2, we observe that under a cancer model with probabilities $p(\psi)=(0.8,0.1,0.1)$, the model where always minimal response is assumed leads to an extremely negatively biased estimator of the survival function. In contrast, the model assuming always a perfect response after each relapse produces a less biased estimator. Clearly, this result is expected since the data were generated by a model where most of the interventions after relapses lead to CRs ($p_1=0.8$). In the case where $p(\psi)=(0.1,0.2,0.7)$, an analogous phenomenon is observed, i.e. the minimal model yields less bias because the data are generated with most interventions leading to null response ($p_3=0.7$). However, under a non-extreme case, e.g. $p(\psi)=(0.3,0.5,0.2)$, assuming a minimal response or a perfect response model lead to unacceptable results as the estimators become highly biased.

The biases of the estimator of the baseline survivor function under mis-specification of the effective age, as depicted in Figure 2, are not unexpected. When the model assumes the wrong effective age function, the estimation procedure compensates in the sense that parameter estimates, particularly of the baseline survivor function and α, are adjusted to best fit the observed data. Thus, for example, if the model assumes NR always but the true effective age

has CR always, and if $\lambda_0(s)$ is a decreasing failure rate hazard function, then the estimator of $\lambda_0(s)$ will tend to overestimate $\lambda_0(s)$ (especially for small s) to accommodate the unexpectedly short inter-event times exhibited by the data. This results in underestimation of the baseline survivor function, as shown in the first row of Figure 2 (see Peña *et al.* (2004) for more discussion of this issue).

As alluded to by a reviewer, the interplay among parameter estimates that can occur under mis-specification of the effective age suggests that bias might be better assessed (and would be expected to be less) for some global quantity computed from the full intensity function in (1), rather than for functions of λ_0 alone. Figure 2 depicts the 'bias' or 'error' in the estimation of the baseline survivor function when the observed effective age is perturbed, and hence this figure may be better interpreted as illustrating the sensitivity of the estimators to incorrect effective age data. Whereas compensatory adjustments among the parameter estimates may diminish the effects of incorrect effective age data on estimates of global quantities, our focus on the baseline survivor function in Figure 2 highlights the critical importance of correctly monitoring the effective age process to be able to properly interpret the individual model parameter estimates.

5. APPLICATION TO NON-HODGKIN'S LYMPHOMA

The indolent non-Hodgkin's lymphomas (NHL) constitute a heterogeneous group of lymphoproliferative disorders. They encompass what were called low-grade and some categories of intermediate-grade NHL in the Working Formulation [27, 33]. They are categorized based on pathologic and cytologic features. The indolent lymphomas include different subtypes of lymphomas such as follicular lymphomas, small lymphocytic lymphoma, lymphoma marginal zone, or sezary syndrome among others. Low-grade lymphomas are associated with relatively prolonged survival. Because it is considered an indolent, but not curable, type of cancer, patients tend to relapse over time. Thus, patients are treated after each recurrence with intensive therapeutic approaches in an attempt to increase the time until next relapse (that is, to increase the disease-free survival). As we have illustrated in previous sections, the treatments may produce different responses (e.g. CR, PR, NR) depending on disease status after therapy. As we have also mentioned, these responses may modify the probability of a subsequent relapse, and hence this intervention effect should be taken into account when modelling this type of data. We illustrate the application of our cancer model and compare results to those obtained from existing models using data obtained on a specific type of indolent lymphoma.

The data consist of the times to relapse, in months, for 63 patients with clinical, histopathological, and immunophenotypes of primary cutaneous marginal zone B-cell lymphoma (PCMZCL) as a particular subtype of indolent lymphoma. An analysis of a subset of these data based on 22 patients with a specific subtype of cutaneous lymphoma was presented in a recent paper [34]. We use the date of first treatment as the beginning of the study. If instead we choose the date of diagnosis, the disease status may vary among patients, invalidating the initial value $A_0 = 0$ in (3). We have also obtained information about the response achieved after treatment upon relapses, coded as CR, PR, or NR, depending on the disease status, for each relapse for each of the 63 patients. The total number of relapses among all patients is 112. The fraction of patients with no relapse is 57 per cent, and only 7 per cent have 3 or more events. The median follow-up time is 2.9 years (range 1 month to 13.5 years). Thirty

eight of 49 (77.8 per cent) responses to treatments administered after relapses are CRs, 9 (14.3 per cent) are PRs, and 2 (4.5 per cent) NRs. We include in the analyses the covariates X_1: gender of patient (0 = male, 1 = female); X_2: delay between first symptom and date of first treatment as a continuous variable (in years); and X_3: lesions involved at diagnosis (0 = single, 1 = localized, 2 = more than one nodal site, 3 = generalized), encoded as three indicator variables. The distribution of these covariates is as follows: 73 patients are males (65.2 per cent) and 39 females (34.8 per cent); the median time of the delay between first symptom and first treatment is 29.7 months (range 1–144 months); 28 patients (25.0 per cent) presented single lesions at diagnosis, 43 localized lesions (38.4 per cent), 35 more than one nodal site (31.2 per cent), and 6 (5.4 per cent) patients had generalized lesions.

We first examine the effect of assumptions concerning the effective age function. To do so, we fit some simple models that include only the lesion at diagnosis (X_3) as a covariate. We compare the results obtained using the cancer model with those obtained from Peña and Hollander's model assuming always NR or always CR for the effective age. Then, we also compare these results with the AG model including response to treatment as a time-dependent covariate. We denote by β the length-three coefficient associated with X_3 coded as a dummy variable. Figure 3 gives the estimated disease-free survival curves for three different effective age specifications (always NR, always CR, and cancer model) for patients with single and with more than one site affected. When CR is assumed at each relapse, the survival probability tends to be underestimated for short times and overestimated for longer times, relative to using the cancer model incorporating information about the intervention effect. But when NR is assumed at each relapse, the survival probability tends to be overestimated for short times and underestimated for longer times. Intuitively, the assumption of a constant intervention effect, when in fact it varies, leads to an incorrect time scale in the hazard rate function, thus inducing bias in the estimators.

Regarding the parameter estimates, the three assumed forms of the effective age give rise to differences mainly in the frailty parameter, as shown in Table I. If we use the minimal repair effective age $\mathscr{E}(s) = s$ (always NR), we obtain a small value of frailty precision, $\hat{\xi} = 2.24$, indicating the need to include a frailty component. On the other hand, if we assume $\mathscr{E}(s) = s - S_{N^\dagger(s-)}$ as effective age (always CR) we obtain $\hat{\xi} = 11\,145\,048$, a very large value that indicates that there is no need for the frailty component. Finally, if we use the cancer model formulation $\mathscr{E}(s) = A_{N^\dagger(s-)} + \left(s - S_{N^\dagger(s-)}\right)$ as effective age (different responses can be achieved) the resulting estimates again indicate the importance of the frailty with $\hat{\xi} = 1.36$. We can test the significance of the frailty to verify these statements. As Therneau and Grambsch [3] indicate, a likelihood ratio test for the frailty can be computed as twice the difference between the log-partial-likelihood with the frailty terms integrated out, and the log-likelihood of a model without frailties. These values for cancer model are -181.46 and -176.27, respectively. That is, the likelihood ratio test can be computed as $2(181.46 - 176.27)$ with one degree of freedom and p-value = 0.0013. The same procedure for the minimal and perfect repair models yields p-values of 0.999 and 0.0201, respectively. These results partly confirm that the need for the frailty term depends on the form of the effective age function.

The hazard ratios (HR) for the risk of relapse associated with X_3 vary little for the three forms of effective age. In all models, patients with localized lesions and with generalized lesions at diagnosis have a higher risk of relapse compared to those with single lesions, showing a similar HR for each model (Table I). This risk is also high for patients with more

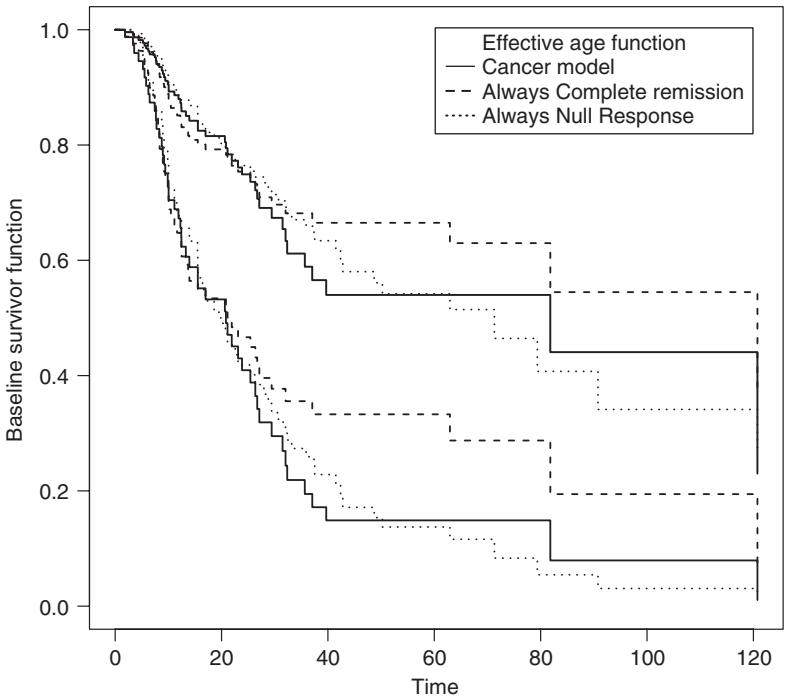

Figure 3. Estimates of survivor function (with frailties set to one) for multiple events for PCMZCL data set by lesions involved at diagnosis using three different formulation for effective age function.

Table I. Hazard ratios and confidence intervals at 95 per cent (in parenthesis) for the probability of relapse depending on lesions involved at diagnosis for the PCMZCL data set. Estimates obtained from the general model using three different effective ages processes.

	Minimal*	Perfect†	Cancer‡
α	0.72 (0.37)	0.90 (0.15)	0.68 (0.21)
Frailty	2.24	∞	1.36
Lesions			
Single	1	1	1
Localized	2.48 (0.93–6.60)	2.42 (0.90–6.50)	2.59 (0.88–7.62)
>1 nodal site	4.47 (1.64–12.15)	3.26 (1.20–8.87)	4.55 (1.22–16.93)
Generalized	3.24 (0.70–14.95)	2.69 (0.59–12.31)	3.09 (1.01–9.37)

*Effective age is $\mathscr{E}(s) = s$.
†Effective age is $\mathscr{E}(s) = s - S_{N^\dagger(s-)}$.
‡Effective age is $A_{N^\dagger(s-)} + (s - S_{N^\dagger(s-)})$.

than one nodal site being statistically significant in all models. However, we observe some differences in their magnitude. Finally, we notice that none of models provide a confidence interval for α (based on approximate normality) that excludes 1.

Table II. Hazard ratios and confidence intervals at 95 per cent (in parenthesis) for the probability of relapse for the PCMZCL data set. Estimates from the Andersen–Gill (AG), Andersen–Gill with response to treatment after relapse as time-dependent covariate (AG2), Wei, Lin and Weissfeld (WLW), and shared gamma frailty model, together with the estimates obtained from the general model using cancer model for effective ages process.

Covariate	AG	AG2	WLW	Shared frailty	Cancer*
α	—	—	—	—	0.88 (0.40)
Frailty	—	—	—	∞	8.85
Gender					
Males	1	1	1	1	1
Females	1.73	2.01	1.83	1.73	1.84
	(0.94–3.19)	(1.12–3.58)	(0.81–4.14)	(0.88–3.40)	(0.82–4.10)
Delay in years	1.01	1.01	1.02	1.04	0.99
	(0.75–1.34)	(0.90–1.11)	(0.88–1.18)	(0.77–1.40)	(0.80–1.23)
Lesions					
Single	1	1	1	1	1
Localized	3.24	3.83	5.23	3.24	3.57
	(1.27–8.23)	(1.50–9.77)	(1.71–15.96)	(1.05–9.96)	(1.17–10.89)
>1 nodal site	3.99	4.77	6.45	3.99	4.67
	(1.73–9.22)	(1.91–11.9)	(2.37–17.56)	(1.41–11.28)	(1.25–17.4)
Generalized	3.44	4.60	23.16	3.44	4.60
	(1.18–9.97)	(1.12–18.9)	(5.02–106.9)	(0.69–16.97)	(1.30–16.3)

* Effective age is $A_{N^\dagger(s-)} + (s - S_{N^\dagger(s-)})$.

Now, continuing to use only the lesions at diagnosis, X_3, as covariate, we compare the estimates resulting from the cancer model with other approaches. A simple way to incorporate the response to treatment in the AG model is by considering the disease status after relapses as a time-dependent covariate (see Reference [3] or [4] for further details). After preparing the data and including the treatment response as a dummy variable, the HR for variable lesions at diagnosis are: 2.46 (CI95 per cent 1.06–5.77), 3.25 (CI95 per cent 1.25–8.47), and 2.77 (CI95 per cent 0.83–9.22), respectively. These results are similar to those obtained using the perfect repair model. Considering that the cancer model reveals that a frailty component is important, perhaps the AG model is not adequate since this model assumes that there is no heterogeneity among patients. Finally, we compare our results to those obtained using only time to first relapse in a Cox model. In that case, the HR are: 1.40 (CI95 per cent 0.48–4.03), 2.76 (CI95 per cent 0.95–8.03), and 3.24 (CI95 per cent 0.28–37.50). Here, we ignore the information in the subsequent relapse times and we observe that this fact substantially affects the estimate of the coefficients and their statistical significance, especially in patients with localized lesions.

The heterogeneity of the risk of relapse may be explained by subject-specific factors other than lesions involved at diagnosis, such as gender or delay between first treatment and first symptom. Thus, we now include all three covariates and compare the estimates of the regression coefficients from the cancer model with those obtained using some of the currently used models (AG, AG with time-dependent covariates, WLW, and shared gamma frailty model). Table II shows these resulting HRs for the PCMZCL data. After adjusting for gender and delay, the variance of frailty decreased to 0.1130 (1/8.84) indicating that the frailty is not

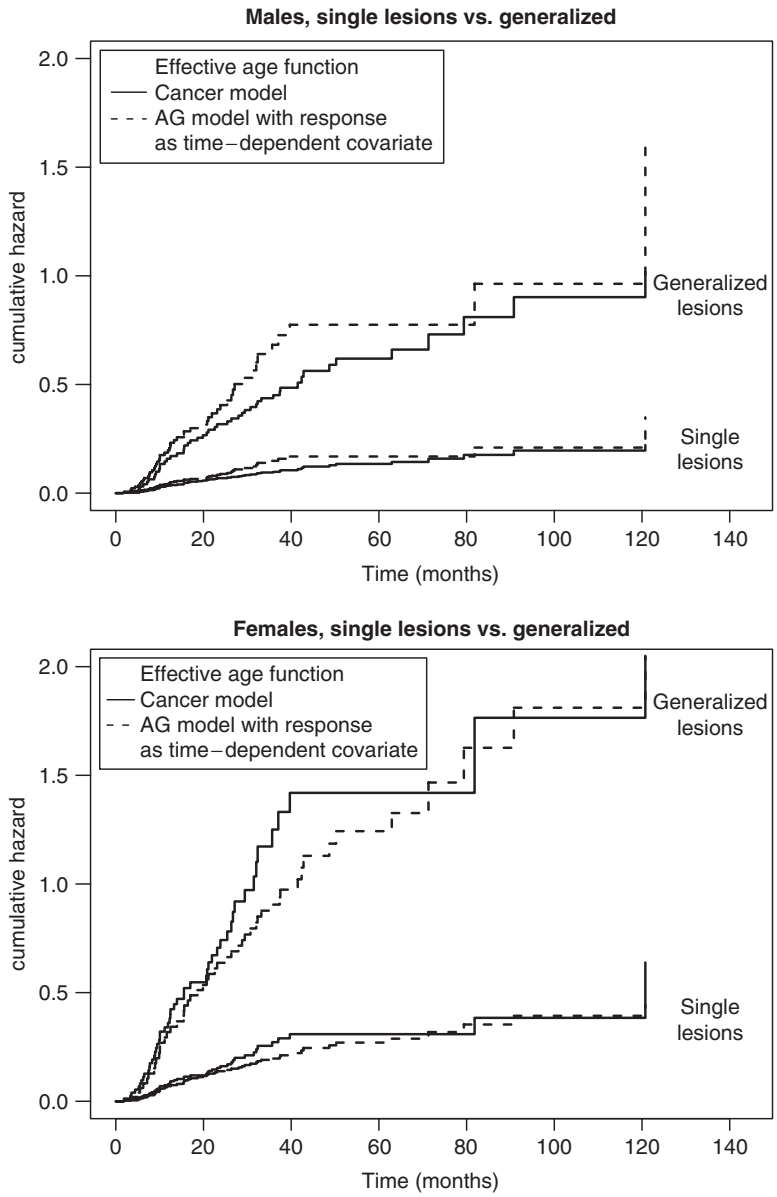

Figure 4. Estimates of cumulative hazard function for multiple events for PCMZCL data set by sex and lesions involved at diagnosis, all cases evaluated at the mean value of delay between first treatment and first symptom. Solid lines shows the hazard assuming the cancer model and dotted lines correspond to AG model which includes the response to treatment as a time-dependent covariate.

necessary (likelihood ratio test $2(180.07-179.40)=1.34$, $p=0.2476$). Similarly, although the estimate of α differs from 1, it is not statistically significant (based on assumed asymptotic normality), so it seems that the prior number of event occurrences does not have an impact. The results also indicate that the AG and frailty models underestimate the risk of having more than one nodal site relative to single lesions, while WLW overestimate it. Only AG model with time-dependent covariate shows similar results to those obtained using cancer model. In that case gender differences are statistically significant if the AG model is chosen.

The estimates of the cumulative hazard functions for the multiple event data for cancer model and for AG model with treatment response as a time-dependent covariate are shown in Figure 4. The solid lines correspond to hazard of relapse for patients with single lesions at diagnosis, obtained via

$$\hat{\Lambda}_0(s)\exp(\hat{\beta}_2\bar{X}_2) \quad \text{and} \quad \hat{\Lambda}_0(s)\exp(\hat{\beta}_1+\hat{\beta}_2\bar{X}_2)$$

for males and females, respectively, where $\hat{\Lambda}_0(\cdot)$ is the estimated cumulative baseline hazard. The dotted lines in this figure are for patients with generalized lesions which correspond to

$$\hat{\Lambda}_0(s)\exp(\hat{\beta}_2\bar{X}_2+\hat{\beta}_5) \quad \text{and} \quad \hat{\Lambda}_0(s)\exp(\hat{\beta}_1+\hat{\beta}_2\bar{X}_2+\hat{\beta}_5)$$

for males and females, respectively. The observed means are $\bar{X}_2=2.4$ for males, and $\bar{X}_2=2.7$ for females. These plots indicate that different risks are associated with the number of lesions involved at diagnosis, as is clear from their associated HRs. The AG and cancer model estimates of the hazard rate functions differ more for patients with generalized lesions than with single lesions.

6. CONCLUDING REMARKS

We have studied how to incorporate information about the effects of treatments or interventions after relapses in analyses of data from patients diagnosed with cancer. The simulation demonstrated that a mis-specification of effective age by not incorporating information about the intervention effect, but instead assuming that always a complete or null response is achieved, has undesirable consequences because the resulting estimators of the finite-dimensional parameters and the baseline survivor function are highly biased. Our application of the cancer model to an indolent lymphoma data set also highlights the need to incorporate information about the effect of intervention after each relapse.

The model can be used in a variety of applications when information about the response to intervention upon relapses can be obtained. The data used by Wei *et al.* [14] that provide the time to recurrence of bladder cancer is an example. This data set, however, does not contain information about the effective age function. Consequently, we applied two models, each using either the calendar or gap time effective age; see Peña *et al.* (2004). However, in this bladder cancer context, it may be possible to elicit from physicians or clinicians the appropriate effective age formulation.

New applications for the cancer model may require modifications of our formulation. One important consideration is the value of A_0 in the effective age (see Equation (3)). Because all our lymphoma patients achieved CR at first treatment, we were able to use this date as study origin, assured that all patients had the same initial status with $A_0=0$. In other situations,

however, CR may not be achievable at first treatment for all subjects. In this case, A_0 is not uniformly zero for all subjects, and disease status after first treatment may be assessed to assign a positive value for A_0. Another aspect of our work that requires consideration in future research is that the time between treatment and assessment of the response to treatment (the ψ_js) following each relapse may not be negligible, contrary to our context and earlier developments in reliability. The delay between application of treatment and evaluation of patient response is widely recognized. However, in cancer studies, at least in hematological diseases, this may not be a problem since all patients are routinely monitored for one month following administration of therapy.

ACKNOWLEDGEMENTS

We thank Dr Octavi Servitge and Dr Cristina Muniesa for providing the data and for their special effort in obtaining the treatment response information. We also thank the two reviewers and Prof. T. Stijnen for their helpful comments that led to improvements in the manuscript.

J. González acknowledges the research support provided by co-operative Research Networks in Cancer (Contract C03/10) and Epidemiology and Public Health (Contract C03/09), funded by the Carlos III Institute, Spanish Ministry of Health. E. Peña acknowledges the research support provided by NSF Grant DMS 0102870, NIH Grant GM056182, NIH COBRE Grant RR17698, and the USC/MUSC Collaborative Research Program. E. Slate acknowledges the research support provided by NIH Grant CA077789, NIH COBRE Grant RR17696, DAMD Grant 17-02-1-0138, NIH Grant 1 U24 DE016508, DOD Grant GC-3532-03-42153CM, and the USC/MUSC Collaborative Research Program.

REFERENCES

1. Lister T. The management of follicular lymphoma. *Annals of Oncology* 1991; **2**(suppl 2):131–135.
2. Romaguera JE, McLaughlin P, North L, Dixon D, Silvermintz KB, Garnsey LA, Velasquez WS, Hagemeister FB, Cabanillas F. Multivariate analysis of prognostic factors in stage IV follicular low-grade lymphoma: a risk model. *Journal of Clinical Oncology* 1991; **9**(5):762–769.
3. Therneau T, Grambsch P. *Modeling Survival Data*: *Extending the Cox Model*. Springer: New York, 2000.
4. Therneau T, Hamilton S. rhDNase as an example of recurrent event analysis. *Statistics in Medicine* 1997; **16**(18):2029–2047. DOI: 10.1002/(SICI)1097-0258(19970930)16:18<2029::AID-SIM637>3.0.CO;2-H
5. Kelly PK, Lim LY. Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine* 2000; **19**(1):13–33. DOI: 10.1002/(SICI)1097-0258(20000115)19:1<13::AID-SIM279>3.0.CO;2-5
6. Lombardo M, Morabito F, Merli F, Molica S, Cavanna L, Sacchi S, Broglia C, Angrilli F, Ilariucci F, Stelitano C, Luisi D, Berte R, Luminari S, Federico M, Brugiatelli M. Bleomycin, epidoxorubicin, cyclophosphamide, vincristine and prednisone (BACOP) in patients with follicular non-Hodgkin's lymphoma: results of a retrospective, multicenter study of the gruppo italiano per lo studio dei linfomi (GISL). *Leukemia and Lymphoma* 2002; **43**(9):1795–8012. DOI: 10.1080/1042819021000006457
7. Lopez-Guillermo A, Cabanillas F, McLaughlin P, Smith T, Hagemeister F, Rodriguez MA, Romaguera JE, Younes A, Sarris AH, Preti HA, Pugh W, Lee MS. Molecular response assessed by PCR is the most important factor predicting failure-free survival in indolent follicular lymphoma: update the MDACC series. *Annals of Oncology* 2000; **11**(suppl 1):137–140.
8. Godder K, Eapen M, Laver JH, Zhang MJ, Camitta BM, Wayne AS, Gale RP, Doyle JJ, Yu LC, Chen AR, Garvin Jr JH, Sandler ES, Yeager AM, Edwards JR, Horowitz MM. Autologous hematopoietic stem-cell transplantation for children with acute myeloid leukemia in first or second complete remission: a prognostic factor analysis. *Journal of Clinical Oncology* 2004; **22**(18):3798–3804.
9. Schwandner O, Schiedeck T, Bruch H, Duchrow M, Windhoevel U, Broll R. p53 and Bcl-2 as significant predictors of recurrence and survival in rectal cancer. *European Journal of Cancer* 2000; **36**(3):348–356. DOI: 10.1016/S0959-8049(99)00271-3
10. Fredriksson I, Liljegren G, Arnesson L, Emdin S, Palm-Sjovall M, Fornader T, Holmqvist M, Holmberg L. Local recurrence in the breast after conservative surgery—a study of prognosis and prognostic factors in 391 women. *European Journal of Cancer* 2002; **38**(14):1860–1870. DOI: 10.1016/S0959-8049(02)00219-8

11. Gail M, Santner T, Brown C. An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* 1980; **36**(2):255–266.
12. Cheuvart B. A nonparametric model for multiple recurrences. *Applied Statistics* 1988; **37**:157–168.
13. Byar D. The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of Placebo, Piroxidine, and Topical Thiotepa. In Pavone-Macaluso M, Smith PH, Edsmyn F (eds), *Bladder 20 Tumors and Other Topics in Urological Oncology*. Plenum: New York, 1980.
14. Wei L, Lin D, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of American Statistical Association* 1989; **84**:1065–1073.
15. Prentice R, Williams B, Petersen A. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**:373–379.
16. Andersen P, Gill R. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; **10**:1100–1120.
17. Oakes D. Frailty models for multiple event times. In Klein JP, Goel PK (eds), *Chapter in Survival Analysis*: *State of the Art*. Kluwer: Dordrecht, 1991; 371–379.
18. Peñna E, Hollander M. Models for recurrent events in reliability and survival analysis. In Mazzuchi T, Singpurwalla N, Soyer R (eds), *Chapter in Mathematical Reliability*: *An Expository Perspective*. Kluwer Academic Publishers: Dordrecht, 2004; 105–123.
19. Whitaker L, Samaniego F. Estimating the reliability of systems subject to imperfect repair. *Journal of American Statistical Association* 1989; **84**:301–309.
20. Block H, Borges W, Savits T. Age-dependent minimal repair. *Journal of Applied Probability* 1985; **22**:51–57.
21. Brown M, Proschan F. Imperfect repair. *Journal of Applied Probability* 1983; **20**:851–859.
22. Hollander M, Presnell B, Sethuraman J. Nonparametric methods for imperfect repair models. *Annals of Statistics* 1992; **20**(2):879–896.
23. MacLaughlin P. Editorial on 'Survival after progression in patients with follicular lymphoma: analysis of prognostic factors'. *Annals of Oncology* 2002; **13**(4):499–500. DOI: 10.1093/annonc/mdf149
24. Montoto S, Lopez-Guillermo A, Ferrer A, Camos M, Alvarez-Larran A, Bosch F, Blade J, Cervantes F, Esteve J, Cobo F, Colomer D, Campo E, Montserrat E. Survival after progression in patients with follicular lymphoma: analysis of prognostic factors. *Annals of Oncology* 2002; **13**(4):523–530. DOI: 10.1093/annonc/mdf11921
25. Spinolo JA, Cabanillas F, Dixon DO, Khorana SM, McLaughlin P, Velasquez WS, Hagemeister FB, Redman JR, Swan Jr F. Therapy of relapsed or refractory low-grade follicular lymphomas: factors associated with complete remission, survival and time to treatment failure. *Annals of Oncology* 1992; **3**(3):227–232. DOI: 10.1093/annonc/92265651
26. Kijima M, Morimura H, Suzuki Y. Periodical replacement problem without assuming minimal repair. *European Journal of Operational Research* 1988; **37**:194–203.
27. Cheson BD, Horning SJ, Coiffier B, Shipp MA, Fisher RI, Connors JM, Lister TA, Vose J, Grillo-Lopez A, Hagenbeek A, Cabanillas F, Klippensten D, Hiddemann W, Castellino R, Harris NL, Armitage JO, Carter W, Hoppe R, Canellos GP. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI sponsored international working group. *Journal of Clinical Oncology* 1999; **17**(4):1244–1253.
28. Weisdorf D, Andersen J, Glick J, Oken M. Survival after relapse of low-grade non-Hodgkin's lymphoma: implications for marrow transplantation. *Journal of Clinical Oncology* 1992; **10**(6):942–947.
29. Davidge-Pitts M, Dansey R, Bezwoda W. Prolonged survival in follicular non-Hodgkin's lymphoma is predicted by achievement of complete remission with initial treatment: results for a long term study with multivariate analysis of prognostic factors. *Leukemia and Lymphoma* 1996; **24**(1–2):131–140.
30. Kijima M. Some results for repairable systems with general repair. *Journal of Applied Probability* 1989; **26**:89–102.
31. Dorado C, Hollander M, Sethuraman J. Nonparametric estimation for a general repair model. *Annals of Statistics* 1997; **25**(3):1140–1160.
32. Gonzalez JR, Slate E, Peña E. gcmrec: General class of models for recurrent event data. R package version 1.0-1. http://www.r-project.org [last access May 2005].
33. The Non-Hodgkin's Lymphoma Pathologic Classification Project. National cancer institute sponsored study of classifications of non-Hodgkin's lymphomas: summary and description of a working formulation for clinical usage. *Cancer* 1982; **49**(10):2112–2135.
34. Servitje O, Gallardo F, Estrach T, Pujol RM, Blanco A, Fernandez-Sevilla A, Petriz L, Peyri J, Romagosa V. Primary cutaneous marginal zone B-cell lymphoma: a clinical, histopathological, immunophenotypic and molecular genetic study of 22 cases. *British Journal of Dermatology* 2002; **147**(6):1147–1158. DOI: 10.1046/j.1365-2133.2002.04961.x

