On models for the estimation of the excess mortality hazard in case of insufficiently stratified life tables

FRANCISCO J. RUBIO*

Department of Mathematics, King's College London, London WC2R 2LS, UK

javier.rubio_alvarez@kcl.ac.uk

BERNARD RACHET

Cancer Survival Group, Faculty of Epidemiology and Population Health, Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

ROCH GIORGI

Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, Hop Timone, BioSTIC, Biostatistique et Technologies de l'Information et de la Communication, Marseille, France

CAMILLE MARINGE, AURÉLIEN BELOT

Cancer Survival Group, Faculty of Epidemiology and Population Health, Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
The CENSUR working survival group

SUMMARY

In cancer epidemiology using population-based data, regression models for the excess mortality hazard is a useful method to estimate cancer survival and to describe the association between prognosis factors and excess mortality. This method requires expected mortality rates from general population life tables: each cancer patient is assigned an expected (background) mortality rate obtained from the life tables, typically at least according to their age and sex, from the population they belong to. However, those life tables may be insufficiently stratified, as some characteristics such as deprivation, ethnicity, and comorbidities, are not available in the life tables for a number of countries. This may affect the background mortality rate allocated to each patient, and it has been shown that not including relevant information for assigning an expected mortality rate to each patient induces a bias in the estimation of the regression parameters of the excess hazard model. We propose two parametric corrections in excess hazard regression models, including a single-parameter or a random effect (frailty), to account for possible mismatches in the life table and thus misspecification of the background mortality rate. In an extensive simulation study, the good statistical performance of the proposed approach is demonstrated, and we illustrate their use on real population-based data of lung cancer patients. We present conditions and limitations of these methods and provide some recommendations for their use in practice.

*To whom correspondence should be addressed.

1. Net survival and excess mortality hazard model

Survival analysis after the diagnosis of cancer is an active research area in cancer epidemiology and a primary interest of many countries. The three typical frameworks adopted to model cancer survival data are: (i) the overall survival setting (where *overall or all-cause* mortality is studied), (ii) the cause-specific setting (where the cause of death is known), and (iii) the relative survival setting. The overall survival setting is not the optimal choice for cancer epidemiology when the main interest is on comparing two or more populations (e.g. two countries, two periods in the same country, two groups with different deprivation levels within the same country and at the same period, and etc.), since it is affected by other causes of mortality, which may differ for the populations of interest. In practice, the cause of death is typically unavailable or, in the absence of a standardized protocol, not reliable. Thus, the cause-specific setting may not be a reasonable choice either. The relative survival setting (Pohar-Perme *and others*, 2016) represents a useful alternative, where the mortality hazard associated to other causes is approximated by the general population hazard $h_P(\cdot; \mathbf{z})$, which is typically obtained from life tables based on the available sociodemographic characteristics $\mathbf{z}$ (e.g. sex, region, and deprivation level) in addition to age and year. The general population hazard is also referred to as the "expected hazard" and the "background mortality rate." While defined in a hypothetical world, where patients could only die from the cancer under study, net survival (the main object of interest in the relative survival setting) represents a useful way of reporting and comparing the probability of survival of cancer patients since this quantity is not affected by differences in expected mortality (due to other causes) between populations. The basic idea behind net survival consists of decomposing the hazard function associated to an individual, $h_o(t; \mathbf{x})$, as the sum of the hazard associated to the disease of interest (e.g. a specific cancer), $h_E(t; \mathbf{x})$, and the hazard associated to other causes in the population of interest, $h_{\text{Other}}(t; \mathbf{z})$. The hazard associated to other causes is approximated with the general population hazard $h_P(t; \mathbf{z})$, assuming that the contribution in the general population hazard of a specific cancer type is small compared to all other causes of death. This is:

$$h_o(t; \mathbf{x}) = h_P(A + t; y + t; \mathbf{z}) + h_E(t; \mathbf{x}), \tag{1.1}$$

where $\mathbf{x}$ and $\mathbf{z}$ are vectors of covariates, and $\mathbf{z}$ typically corresponds to a subset of covariates of $\mathbf{x}$. The variables "A" and "y" represent the age at diagnosis and the year of diagnosis, respectively, thus $A + t$ and $y + t$ represent the age and the year at time t after diagnosis. This model is also known as *excess hazard model* (Esteve *and others*, 1990). The net survival is defined as the survival function associated to the excess hazard function $h_E(\cdot; \mathbf{x})$. Estimation of the net survival function has been largely studied from parametric, semiparametric, and non-parametric perspectives (see Remontet *and others*, 2007; Pohar Perme *and others*, 2009; Perme *and others*, 2012; Rubio *and others*, 2018).

In practice, a limitation of quantities derived in the relative survival setting (Perme *and others*, 2012), such as the excess hazard, is that patients need to be matched to groups of the general population sharing the available characteristics $\mathbf{z}$ in order to obtain the background mortality rates $h_P(\cdot; \mathbf{z})$ from the life tables. The number of available characteristics $\mathbf{z}$ varies for different countries, and there exist certain characteristics that are not available at the population level that may affect the background mortality rates such as deprivation, ethnicity, drug use (tobacco, alcohol, and etc.), comorbidities, among others. For instance, it has been shown that deprivation levels are associated with life expectancy in some populations (Woods *and others*, 2005). Thus, if life tables are obtained without a proper stratification of deprivation levels, this will imply that two individuals with different deprivation levels (e.g. most affluent vs. most

deprived), but sharing other sociodemographic characteristics, will be assigned the same background mortality. We will refer to the case when the life table is not sufficiently stratified as a "mismatch in the life table." This mismatch may also apply to other characteristics (e.g. smoking status), thus potentially impacting the comparison ability of the net survival measure (Pavlič and Pohar-Perme, 2018). Moreover, Dickman *and others* (1998) and Grafféo *and others* (2012) showed that not including relevant information for matching the mortality rates of patients induces a bias in the estimation of the parameters in net survival models (even for other variables than those included in the life tables). Concerns about the implications of this kind of mismatches in the life tables have recently been discussed in Pavlič and Pohar-Perme (2018) and Bower *and others* (2018). Thus, it is desirable to produce models that can capture possible mismatching information in the background mortality rates associated to each patient, and that allow the assessment of the impact of this mismatch. In this work, we study an existing model for correcting the background mortality using a single correction parameter (Cheuvart and Ryan, 1991), and extend it to a general hazard structure for excess hazard regression models (Rubio *and others*, 2018). We also propose a correction model using a random effect (instead of a single parameter). For each of these models, we study the properties of maximum likelihood estimation of the corresponding parameters. We assess the performance of this inferential procedure in an extensive simulation study. We apply and compare these methods using a lung cancer data example. We conclude with some practical advice and summarize the conditions and limitations of these methods, as well as potential directions for further research.

2. Corrections of the background mortality in excess hazard regression

In this section, we present two parametric corrections that can account for mismatches in the life table. The first one corresponds to the single parameter correction proposed by Cheuvart and Ryan (1991). The second one corresponds to our proposal, which assumes that the correction to the background mortality is random and can be modeled with a parametric distribution. Next, we present a brief summary of these methods, as well as some of their properties and limitations.

2.1. *Single parameter correction: Cheuvart and Ryan's approach*

In the context of clinical trials, Cheuvart and Ryan (1991) proposed an extension of model (1.1) in which they allowed for a constant correction on the population hazard: the overall hazard after t years of follow-up for a patient who entered the trial at age A in year y and with covariates $\mathbf{x}$ was assumed to be

$$h_o^C(t; \mathbf{x}) = \gamma h_P(A + t; y + t; \mathbf{z}) + h_E(t; \mathbf{x}), \tag{2.2}$$

where $\gamma > 0$ is an unknown parameter differentiating the competing mortality of eligible patients from that of the general population. Cheuvart and Ryan (1991) employ the proportional excess hazard model $h_E(t; \mathbf{x}) = h_{E,0}(t; \eta) \exp(\mathbf{z}^\top \boldsymbol{\beta})$, where $h_{E,0}(\cdot; \eta)$ is the baseline excess hazard with parameter η, $\boldsymbol{\beta}$ is a vector of regression parameters, including the effects of treatment and other prognostic factors. In the context of clinical trials, the correction parameter can be reasonably assumed to be the same for all individuals since the population entering the trials is selected based on their characteristics (usually patients without comorbid conditions, and etc.).

Model (2.2) can be rewritten in terms of the cumulative hazard and the survival functions as follows:

$$H_o^C(t; \mathbf{x}, \gamma) = \gamma [H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z})] + H_E(t; \mathbf{x}), \tag{2.3}$$

$$S_o^C(t; \mathbf{x}, \gamma) = \exp\{-\gamma [H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z})]\} \exp[-H_E(t; \mathbf{x})], \tag{2.4}$$

where H_P and H_E are the cumulative hazard functions obtained from h_P and h_E, respectively.

In principle, model (2.2) can also be used to account for mismatched life tables, as the correction is made on the population hazard. The basic assumption behind this model is that the true competing mortality is proportional to the mortality of the population obtained from the life tables, and the correction is the same for all individuals. Expression (2.3) also indicates that the information used for estimating the additional correction parameter γ comes from the differences in the cumulative hazard $H_P(A+t; y+t; \mathbf{z}) - H_P(A; y; \mathbf{z})$, which are not used in the estimation of the classical model (1.1).

2.2. *Frailty correction for the population hazard*

One limitation of model (2.2) is that the correction γ, made on the population hazard, is assumed to be constant across all the individuals. This assumption may not be realistic in population studies, where the diversity of unavailable sociodemographic characteristics used to obtain the life tables may induce a non-constant mismatch. Thus, instead of assuming that γ is constant in model (2.2), we assume that γ is a positive continuous random variable. This implies that the correction factor γ is allowed to vary across the different individuals. More specifically, consider the conditional hazard model

$$\tilde{h}_o(t \mid \gamma; \mathbf{x}) = \gamma h_P(A + t; y + t; \mathbf{z}) + h_E(t; \mathbf{x}), \tag{2.5}$$

$$\gamma \sim G.$$

where G is an arbitrary absolutely continuous cumulative distribution function with support on $\mathbb{R}_+$. The conditional overall survival function is given by

$$\tilde{S}_o(t \mid \mathbf{x}, \gamma) = \exp\{-\gamma[H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z})]\} \exp[-H_E(t; \mathbf{x})],$$

$$\gamma \sim G.$$

Then, after integrating out the frailty γ with respect to the distribution G (see Appendix of the supplementary material available at *Biostatistics* online), the individual marginal overall survival function can be written as

$$\tilde{S}_o(t; \mathbf{x}) = \exp\{-H_E(t; \mathbf{x})\}\mathcal{L}_G\{H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z})\}, \tag{2.6}$$

where $\mathcal{L}_G\{s\} = \int_0^\infty e^{-sr} dG(r)$ denotes the Laplace transform of G evaluated at time s. Next, we consider a specific choice for the distribution G: a Gamma distribution. This choice allows for obtaining a closed-form expression of the marginal survival function, in addition to its appealing flexibility and interpretability of parameters.

Gamma frailty

Consider the conditional hazard model (2.5) and suppose that $\gamma \sim \text{Ga}(\mu, b)$, where $\text{Ga}(\mu, b)$ denotes a Gamma distribution with mean parameter $\mu > 0$, scale parameter $b > 0$, and probability density function $g(r; \mu, b) = \dfrac{r^{\frac{\mu}{b} - 1}}{\Gamma\left(\frac{\mu}{b}\right) b^{\frac{\mu}{b}}} \exp\left(-\dfrac{r}{b}\right)$. Then, it follows that

1. The marginal individual survival function is given by

$$\tilde{S}_o(t; \mathbf{x}) = \frac{\exp\{-H_E(t; \mathbf{x})\}}{\{1 + b[H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z})]\}^{\frac{\mu}{b}}}. \tag{2.7}$$

2. The marginal individual hazard function is given by

$$\tilde{h}_o(t; \mathbf{x}) = \frac{\mu \, h_P(A + t; y + t; \mathbf{z})}{1 + b \left[H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z}) \right]} + h_E(t; \mathbf{x}). \tag{2.8}$$

Expression (2.8) provides a nice interpretation of our approach since the observed hazard can be seen as a model with a correction function $\omega_1(t, \mathbf{z}; \mu, b) = \dfrac{\mu}{1 + b \left[H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z}) \right]}$, which is identifiable and provides a functional form that involves the mean correction, the scale or spread of the correction, and the differences on the population cumulative hazards. Another property of the correction function is that $\lim_{b \to 0} \omega_1(t, \mathbf{z}; \mu, b) = \mu$, which indicates that the correction model (2.2) is a limit case. This property, however, does *not* imply that the correction parameter in (2.2) represents the mean correction when $b > 0$, as shown in our simulations. We can also observe that if two patients have the same sociodemographic characteristics (or virtually the same) but different survival times, then the corresponding value of the weight ω_1 will differ since the corresponding differences in the cumulative hazards $H_P(A + t; y + t; \mathbf{z}) - H_P(A; y; \mathbf{z})$ will be different. This, intuitively, suggests that the more extreme the survival time t of a patient is (either too small or too large), compared to other patients with the same sociodemographic characteristics, the more likely this patient has been assigned the incorrect mortality rate from the life table. This approach implicitly assumes that the excess hazard model $h_E(t; \mathbf{x})$ is correctly specified and includes all important prognosis factors (see Section 6). It also indicates that the information about the parameters of the frailty distribution is provided by the variability in the observed survival times of similar individuals regarding their sociodemographic characteristics and tumor prognosis factors. This suggests the need for a certain amount of observations for groups of patients with similar characteristics.

2.3. *Hazard structure and baseline hazard*

For models (1.1), (2.2), and (2.8), we adopt the general excess hazard model (GH) proposed in Rubio *and others* (2018):

$$h_E^{GH}(t; \mathbf{x}_i) = h_0 \left(t \exp(\mathbf{x}_i^\top \beta_1) \right) \exp(\mathbf{x}_i^\top \beta_2), \tag{2.9}$$

$$H_E^{GH}(t; \mathbf{x}_i) = H_0 \left(t \exp(\mathbf{x}_i^\top \beta_1) \right) \exp(-\mathbf{x}_i^\top \beta_1 + \mathbf{x}_i^\top \beta_2),$$

where $h_0(\cdot)$ is the baseline hazard. This hazard structure contains, as particular cases, the proportional hazards (PH) model when $\beta_1 = 0$, the accelerated hazards (AH) model when $\beta_2 = 0$, the accelerated failure time (AFT) model when $\beta_1 = \beta_2$, as well as combinations of these for $\beta_1 \neq \beta_2 \neq 0$. Thus, this structure covers the most popular hazard structures used in the literature, allowing also to capture time-dependent effects through β_1 (i.e. effects which are not assumed to be constant over the whole follow-up period, such as the effects estimated in PH models). For an extensive discussion on the properties of this GH structure, we refer the reader to Rubio *and others* (2018).

The baseline hazard in (2.9) will be modeled using the Exponentiated Weibull (EW) distribution. The EW distribution contains three positive parameters (κ, θ, α) (shape, scale, and power), and the corresponding hazard function can capture some basic shapes: increasing, decreasing, unimodal (up-then-down), bathtub (down-then-up), and constant. The EW density and cumulative distribution functions with shape, scale, and power parameters (κ, θ, α) are given, respectively, by:

$$f_{EW}(t) = \alpha \frac{\kappa}{\theta} \left(\frac{t}{\theta} \right)^{\kappa - 1} \left[1 - \exp \left\{ - \left(\frac{t}{\theta} \right)^\kappa \right\} \right]^{\alpha - 1} \exp \left\{ - \left(\frac{t}{\theta} \right)^\kappa \right\},$$

$$F_{EW}(t) = \left[1 - \exp \left\{ - \left(\frac{t}{\theta} \right)^\kappa \right\} \right]^\alpha.$$

Thus, the combination of the GH structure with the choice of the EW baseline hazards can capture a variety of hazard structures, time-dependent effects (through the parameters in β_1), and baseline hazard shapes, while allowing for a parsimonious implementation of all the proposed models (Rubio *and others*, 2018).

3. INFERENCE

3.1. *The classical model*

Let $t_i > 0$, $i = 1, \ldots, n$, be the sample of times to event from a population of cancer patients, with covariates $\mathbf{x}_i \in \mathbb{R}^p$, and vital status indicators δ_i (1-death, 0-censored).

For the classical model (1.1), we will rely on the maximum likelihood estimation method, which has been shown to have good inferential properties with the hazard structure detailed in (2.9) (Rubio *and others*, 2018). With this model, the likelihood is defined as:

M1. Classical model (1.1):

$$
\mathcal{L}(\boldsymbol{\psi}; \text{Data}) = \prod_{i=1}^{n} h_o(t_i; \mathbf{x}_i)^{\delta_i} S_o(t_i; \mathbf{x}_i)
$$

$$
\propto \prod_{i=1}^{n} \{h_P(A_i + t_i; y_i + t_i, \mathbf{z}_i) + h_E(t_i; \mathbf{x}_i)\}^{\delta_i} \exp\{-H_E(t_i; \mathbf{x}_i)\},
$$

where $\boldsymbol{\psi}$ represent the parameters of the excess hazard model. In our case, $\boldsymbol{\psi} = (\kappa, \theta, \alpha, \beta_1, \beta_2)$. Notice that here for model (1.1), we omit the expected survival from the likelihood, as this quantity does not depend on parameters.

3.2. *The models with correction of the background mortality*

For the models of interest, (2.2) and (2.8), we will estimate the parameters using the maximum likelihood method. The corresponding likelihood functions are presented below for the single-parameter correction model (2.2) and the frailty correction model (2.8)

M2. Single-parameter correction model (2.2):

$$
\mathcal{L}_C(\boldsymbol{\psi}_C; \text{Data}) = \prod_{i=1}^{n} \{\gamma h_P(A_i + t_i; y_i + t_i; \mathbf{z}_i) + h_E(t_i; \mathbf{x}_i)\}^{\delta_i} \exp\{-H_E(t_i; \mathbf{x}_i)\}
$$

$$
\times \exp\{-[H_P(A_i + t_i; y_i + t_i; \mathbf{z}_i) - H_P(A_i; y_i; \mathbf{z}_i)]\}^{\gamma},
$$

where $\boldsymbol{\psi}_C = (\gamma, \boldsymbol{\psi})$.

M3. Frailty correction model (2.8):

$$
\mathcal{L}_F(\boldsymbol{\psi}_F; \text{Data}) = \prod_{i=1}^{n} \left\{ \frac{\mu h_P(A_i + t_i; y_i + t_i; \mathbf{z}_i)}{1 + b\left[H_P(A_i + t_i; y_i + t_i; \mathbf{z}_i) - H_P(A_i; y_i; \mathbf{z}_i)\right]} + h_E(t_i; \mathbf{x}_i) \right\}^{\delta_i}
$$

$$
\times \frac{\exp\{-H_E(t_i; \mathbf{x}_i)\}}{\{1 + b\left[H_P(A_i + t_i; y_i + t_i; \mathbf{z}_i) - H_P(A_i; y_i; \mathbf{z}_i)\right]\}^{\frac{\mu}{b}}},
$$

where $\boldsymbol{\psi}_F = (\mu, b, \boldsymbol{\psi})$.

These likelihood functions can be maximized using standard optimization routines from the R software (e.g. "nlminb" or "optim"). In order to select between models M1–M3, we use the Akaike Information Criterion (AIC), as these models are estimated using the maximum likelihood method.

3.3. *Choice of initial points for the optimization process*

For the optimization process, we consider a two-step algorithm. In the first step, we initialize the search at the initial values: $\boldsymbol{\psi}^{(0)} = (\kappa^{(0)}, \theta^{(0)}, \alpha^{(0)}, \beta_1^{(0)}, \beta_2^{(0)})$, where $\kappa^{(0)} = 1$, $\theta^{(0)} = 1$, $\alpha^{(0)} = 2$, $\beta_1^{(0)} = \beta_2^{(0)} = \mathbf{0}$, and then move from these initial values using one cycle of a coordinate descend algorithm (CDA), coupled with the R command "nlminb" on each step of the CDA (see Wright, 2015 for more details on the CDA). The CDA is an algorithm which successively maximizes an objective function along coordinate directions. In our case, we first obtain a new value of the first parameter, $\kappa^{(1)}$, after maximizing the objective function with respect to κ, while setting the other parameter values at their initial values. Then, we maximize again the objective function, obtained by setting this time the initial values to $(\kappa^{(1)}, \theta^{(0)}, \alpha^{(0)}, \beta_1^{(0)}, \beta_2^{(0)})$ to get $\theta^{(1)}$. We repeat the same process for all other parameters to finally get a vector $(\kappa^{(1)}, \theta^{(1)}, \alpha^{(1)}, \beta_1^{(1)}, \beta_2^{(1)})$. In the second step, we utilize the values obtained with the CDA as new initial values in a general purpose optimization algorithm (e.g. "nlminb" or "optim" from the R software), in order to obtain the MLE $\widehat{\boldsymbol{\psi}}$.

For model M2, we use the initial values $(\gamma, \widehat{\boldsymbol{\psi}})$, with $\gamma = 1.2$, while for model M3, we employ the initial values $(\mu, b, \widehat{\boldsymbol{\psi}})$, with $\mu = 1.2$ and $b = 0.1$. In practice, we recommend running the optimization process at several initial points in order to ensure that the global maximum of the likelihood is reached.

4. Simulation studies

In this section, we present a simulation study where we assess the impact of mismatches in the population hazard on the estimation of the excess hazard, as well as the performance of the proposed correction models (M2–M3). The true values of the parameters are chosen in order to produce scenarios that resemble cancer population studies concerning an aggressive type of cancer (relatively low 5-year net survival, see Appendix Table 1 of the supplementary material available at *Biostatistics* online), such as lung cancer, similar to the simulations presented in Rubio *and others* (2018). An additional extensive simulation scenario resembling a less aggressive type of cancer (such as colon cancer, see Appendix Table 2 of the supplementary material available at *Biostatistics* online) is presented in the Appendix (Simulation design II) of the supplementary material available at *Biostatistics* online.

4.1. *Data generation*

Briefly, in the Simulation Design I of the supplementary material available at *Biostatistics* online, we simulated $N = 1000$ data sets of size $n = 5000, 10\,000$, assuming the additive hazard decomposition given in (1.1). The variable "age" was simulated as a continuous variable using a mixture of uniform distributions with 0.25 probability on $(30, 65)$, 0.35 probability on $(65, 75)$, and 0.40 probability on $(75, 85)$ years old. The binary variables "sex" and "W" were both simulated from a binomial distribution with probability 0.5 (the binary variable "W" could be viewed as "treatment" or "comorbidity" or "stage" (early and late)). In all scenarios, we simulated the "other-causes" time to event using the UK life tables based on "age" and "sex" (assuming that all patients were diagnosed on the same year). The time to event from the excess hazard (cancer death time) was generated using the inverse transform method, assuming effects of the three variables "age," "sex," and "W" and an EW distribution. We assumed either (i) only administrative censoring at $T_C = 5$ years, which induced approximately 25% censoring in all cases, or (ii) an additional independent random censoring (drop-out) using an exponential distribution with rate parameter r, inducing approximately 30% censoring in these cases. Given the GH structure (2.9) adopted
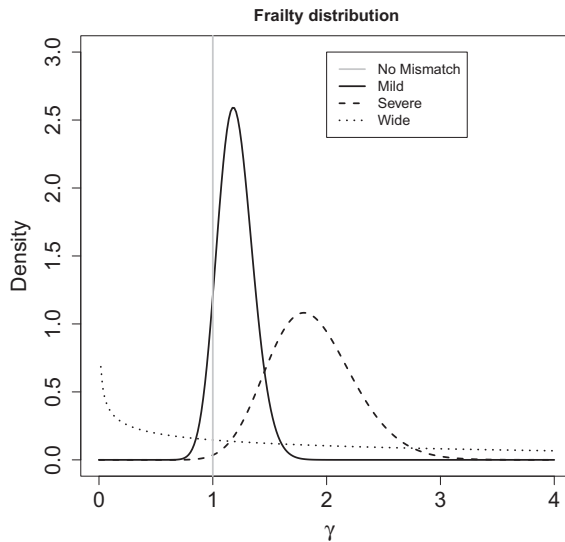
Fig. 1. Frailty distributions used in the simulation.

for the simulation, all variables affect the time scale (i.e. time-dependent effects) as well as the hazard scale (i.e. changing the level of the hazard). We refer the reader to Rubio *and others* (2018) for a more detailed discussion on the interpretation of the GH structure.

We consider four scenarios that represent mismatches in the life tables: (i) No mismatch, where the data are generated from model (1.1); (ii) Moderate mismatch, where the data are generated from the hazard model (2.5) with $\gamma \sim Ga(1.2, 0.02)$; (iii) Severe mismatch, where the data are generated from the hazard model (2.5) with $\gamma \sim Ga(1.875, 0.075)$; and (iv) Wide mismatch, where the data are generated from the hazard model (2.5) with $\gamma \sim Ga(6.5, 10)$ (see Figure 1). In all of these scenarios, we fit models M1–M3. We also consider selecting between models M1–M3 using AIC, in order to identify the model favored by the data. The estimates of the parameters of the excess hazard model selected using AIC are reported, and this model is referred to as model M4. For model M4, we report the estimated correction parameter c associated to the selected model. This is, $\widehat{c} = 1$ if AIC selects model M1, $\widehat{c} = \widehat{\gamma}$ if AIC selects model M2, and $\widehat{c} = \widehat{\mu}$ if AIC selects model M3. We report coverage of the asymptotic normal confidence intervals in all the scenarios.

4.2. *Simulation results*

For illustration, Table 1 and Figure 2 show the results for the scenario with $n = 5000$, 30% of censoring and a Wide mismatch (iv). The results of the remaining scenarios are shown in the Appendix of the supplementary material available at *Biostatistics* online, as well as the results with sample size of $n = 10\,000$ and the simulation design II with a higher net survival. As expected, we observe that when there is Severe and Wide mismatch (iii)–(iv), the model without correction (M1) will lead to biased parameters estimates, as well as poor coverage (Table 1, Figure 2; Appendix Tables 5, 8, and 9 and Appendix Figures 4, 7, and 8 of the supplementary material available at *Biostatistics* online). This is reflected in the MLEs of the model parameters as well as on the fitted excess hazards. The bias of the parameter estimates and the poor coverages of M1 are more pronounced in the simulation design II with high net survival (Appendix Tables 12, 13, 16, and 17 and Appendix Figures 11, 12, 15, and 16 of the supplementary material available at *Biostatistics* online). In scenarios (i) and (ii) with No or Moderate mismatch, the

fitted correction models M2 and M3 are centered around the true generating model, although they exhibit (as expected) a slightly larger variability compared to model M1 (Appendix Tables 3, 4, 6, 7, 10, 11, 14, and 15, and Appendix Figures 2, 3, 5, 6, 9, 10, 13, and 14 of the supplementary material available at *Biostatistics* online). In scenario (iii) with Severe mismatch, the fitted correction models M2 and M3 properly correct the mismatch as these models are centered around the true generating model, with the cost of higher variability. The parameters estimated with model M1 are biased with a very low coverage (Appendix Tables 5, 8, 12, and 16 and Appendix Figures 4, 7, 11, and 15 of the supplementary material available at *Biostatistics* online). In scenario (iv) with Wide mismatch, the fitted models M1 and M2 are biased and far from the true generating model. On the other hand, model M3 can capture this mismatch and reduce the bias properly (Table 1, Figure 2, Appendix Tables 9, 13, and 17 and Appendix Figures 8, 12, and 16 of the supplementary material available at *Biostatistics* online). The models selected with AIC (M4) are also centered around the true generating model in all scenarios. In cases with no mismatch or moderate mismatch, we observe that the bias on the estimates is small, the coverage is close to 95%, and all models tend to be relatively close to the true generating model. In those cases, the AIC tends to favor model M1. Table 18 in the Appendix of the supplementary material available at *Biostatistics* online shows the proportion of selected models using AIC. Overall, we can see that M3 is favored in scenario (iv) with Wide mismatch. In the scenario (iii) with Severe mismatch, the proportion of selected models depends on the contribution of the excess hazard compared to the population hazard. In general, selecting the models using AIC or another model selection tool is advised in order to identify the need for correcting the population hazard. The simulation study results obtained with $N = 10\,000$ per sample were very similar to the ones obtained with $N = 5000$.

5. APPLICATION: LUNG CANCER DATA

We now analyze a dataset obtained from population-based national cancer registry of Non-Small Cell Lung Cancer (NSCLC) patients diagnosed in 2012 in England. For deriving information on stage at diagnosis and presence of comorbidities at the time of diagnosis, we linked these data to administrative data (Hospital Episode Statistics [HES] and the Lung Cancer Audit data [LUCADA]) and then applied specific algorithms (Benitez-Majano *and others*, 2016; Maringe *and others*, 2017). We used a 6-year period up to 6 months before diagnosis to retrieve information on comorbidity. We checked for the presence of cardiovascular comorbidity (at least one of: myocardial infarction, congestive heart failure, peripheral vascular disease, and cerebrovascular disease) and chronic obstructive pulmonary disease (COPD). We measured deprivation using the Income Domain from the 2010 England Indices of Multiple Deprivation, defined at the Lower Super Output Area level (mean population 1500). The Income Domain measures the proportion of the population in an area experiencing deprivation related to low income, and ranges from 1% to 75% in our data (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010). Follow-up was assessed on the 31st of December 2015, at which time patients alive were censored (so the maximum follow-up was 4 years). We restricted our analysis to men with no missing data. We observed $n = 15\,688$ patients with complete cases among which $n_o = 13\,603$ died before the 31st of December 2015, and 17 patients were lost to follow-up (censored before the 31st of December 2015). The median follow-up among patients censored was 3.45 years, mainly because of administrative censoring. The 25%, 50%, and 75% quantiles of the patients' age at diagnosis was 65.8, 73.0, 80.0 while the mean was 72.5. Among the patients, 2210 were diagnosed at Stage I, 1502 at Stage II, 3679 at Stage III, and 8297 at Stage IV. Finally, 3224 patients were classified with a cardiovascular comorbidity and 3154 with a chronic obstructive pulmonary disease.

We applied models M1–M3 to estimate the excess mortality hazard using deprivation-specific life tables (detailed by sex, age, year, and Government Office Region in addition to the deprivation quintile).

Table 1. *Simulation results for the scenario GH with* $(\sigma, \kappa, \alpha) = (1.75, 0.6, 2.5)$, $\beta_1 = (0.1, 0.1, 0.1)$, $\beta_2 = (0.05, 0.2, 0.25)$, $n = 5000$, *and wide mismatch* $\gamma \sim Ga(6.5, 10)$. *Mean of the MLEs (MMLE), median of the MLEs (mMLE), empirical standard deviation (ESD), mean (estimated) standard error, root-mean-square error (RMSE), and coverage proportions (Coverage)*

Design I: $\gamma \sim Ga(6.5, 10)$

Model	Parameter	MMLE	mMLE	ESD	Mean Std error	RMSE	Coverage
	σ (1.75)	1.212	1.214	0.228	0.218	0.584	0.464
	κ (0.6)	0.592	0.593	0.046	0.045	0.047	0.934
	α (2.5)	2.365	2.315	0.328	0.308	0.355	0.887
	β_{11} (0.1)	0.119	0.118	0.012	0.012	0.023	0.700
M1	β_{12} (0.1)	0.409	0.405	0.214	0.220	0.376	0.711
	β_{13} (0.1)	−0.034	−0.030	0.233	0.221	0.268	0.902
	β_{21} (0.05)	0.065	0.065	0.002	0.002	0.015	0.000
	β_{22} (0.2)	0.296	0.296	0.047	0.045	0.107	0.449
	β_{23} (0.25)	0.161	0.159	0.047	0.046	0.101	0.484
	σ (1.75)	1.284	1.308	0.232	0.243	0.520	0.741
	κ (0.6)	0.625	0.635	0.060	0.065	0.065	0.845
	α (2.5)	2.199	2.106	0.412	0.402	0.510	0.689
	β_{11} (0.1)	0.119	0.118	0.012	0.013	0.023	0.715
M2	β_{12} (0.1)	0.411	0.406	0.215	0.223	0.378	0.717
	β_{13} (0.1)	−0.011	−0.001	0.225	0.225	0.251	0.928
	β_{21} (0.05)	0.066	0.067	0.003	0.003	0.017	0.004
	β_{22} (0.2)	0.308	0.309	0.046	0.047	0.117	0.352
	β_{23} (0.25)	0.155	0.152	0.044	0.045	0.105	0.412
	γ (6.5)	0.453	0.001	0.652	0.549	6.082	0.846
	σ (1.75)	1.406	1.323	0.745	0.721	0.820	0.978
	κ (0.6)	0.543	0.546	0.127	0.122	0.139	0.966
	α (2.5)	4.082	2.992	7.019	3.070	7.192	0.970
	β_{11} (0.1)	0.099	0.099	0.028	0.024	0.028	0.913
	β_{12} (0.1)	0.121	0.134	0.340	0.339	0.340	0.959
M3	β_{13} (0.1)	0.089	0.103	0.354	0.326	0.354	0.950
	β_{21} (0.05)	0.048	0.049	0.009	0.009	0.010	0.948
	β_{22} (0.2)	0.187	0.198	0.089	0.089	0.090	0.954
	β_{23} (0.25)	0.262	0.251	0.087	0.085	0.087	0.949
	b (10)	13.060	9.072	12.247	10.379	12.618	0.847
	μ (6.5)	7.010	7.134	1.839	1.727	1.908	0.853
	σ (1.75)	1.367	1.274	0.733	0.686	0.826	0.920
	κ (0.6)	0.540	0.546	0.122	0.117	0.136	0.968
	α (2.5)	3.992	2.978	6.479	2.694	6.645	0.960
	β_{11} (0.1)	0.100	0.100	0.028	0.023	0.028	0.895
M4	β_{12} (0.1)	0.139	0.151	0.341	0.328	0.343	0.938
	β_{13} (0.1)	0.082	0.092	0.350	0.317	0.350	0.945
	β_{21} (0.05)	0.049	0.049	0.010	0.008	0.010	0.859
	β_{22} (0.2)	0.194	0.203	0.092	0.086	0.093	0.904
	β_{23} (0.25)	0.257	0.248	0.089	0.082	0.089	0.907
	c (6.5)	6.709	7.137	2.470	–	2.477	–

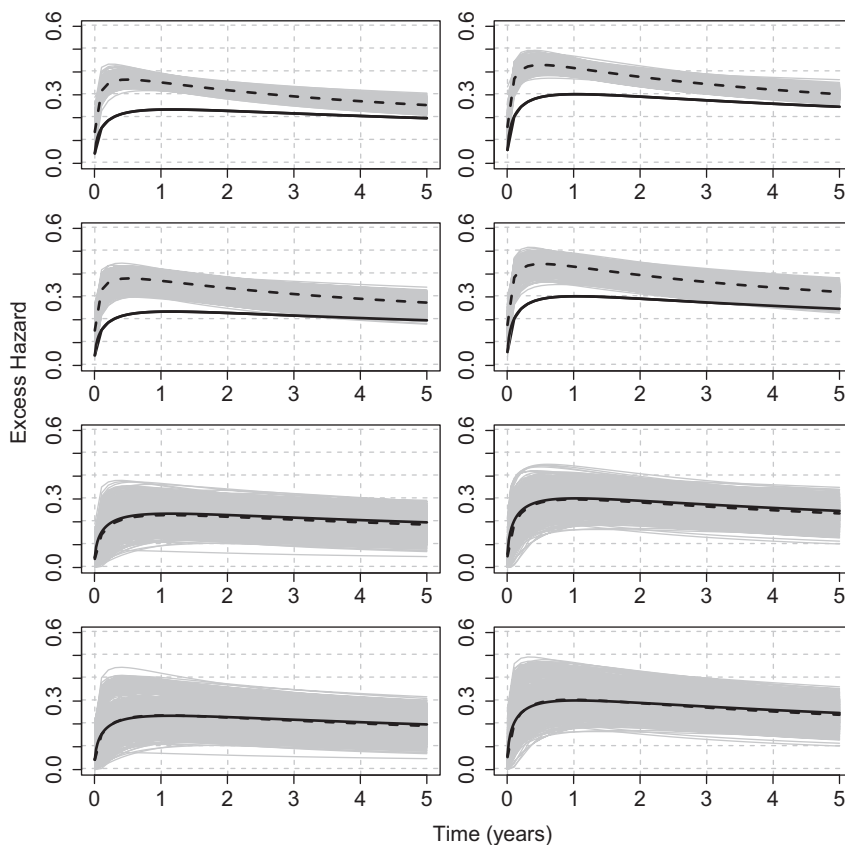Fig. 2. Scenario with wide mismatch: $\gamma \sim Ga(6.5, 10)$. Models M1–M4 from top to bottom. Mean of the fitted excess hazards (dashed lines), compared to the true generating excess hazard (continuous lines), and 1000 sample-specific fitted excess hazards (grey lines) for $n = 5000$ and 30% censoring. Panels from left to right correspond to two sets of values for the covariates (age, sex, comorbidity) = $(70, 0, 0)$, $(70, 0, 1)$, respectively. **Design I**: $\gamma \sim Ga(6.5, 10)$.

Consequently, we are implicitly assuming that the variables age, sex, deprivation, tumor stage, and comorbidity accurately explain the excess hazard in NSCLC patients. The regression parameter estimates for the excess hazard models M1–M3, as well as the correction parameters (for models M2–M3), are reported in Table 2. For illustrating the results, the excess mortality hazard and the corresponding Net Survival for two pre-defined subgroups of patients are depicted in Figure 3.

In Table 2, between the three models (M1, M2, and M3), the AIC favors model M3 (i.e. the frailty correction model). Differences on β estimates (regression coefficients) between M1 and M3 are substantial. For example, the protective effect of Stage I cancer (compared to being diagnosed with a Stage IV cancer) is even higher when accounting for mismatched life tables. This interpretation follows by noticing that the two parameters associated with Stage I are negative (see Rubio *and others*, 2018 for details on those hazard-structure models and their interpretation). The impact of the presence of a comorbidity is higher in M3 compared to M1 and M2. Thus, correcting the population life table for unobserved predicting variables of background mortality seems to be quite relevant in this example. An unobserved variable which certainly affects the population mortality hazard here is smoking status. We observe that the frailty distribution, used for correcting the population mortality in M3, cumulates 23% of the probability mass

below 1, and 77% above 1. That is, the value 1 represents the 23% quantile of the fitted Gamma frailty distribution with scale parameter 9.83 and mean 6.54. These values are in fact related to the proportion of smokers (roughly 80%, which would, in principle, require a correction higher than 1) for England lung cancer patients (Ellis *and others*, 2014), which provides an intuitive interpretation of the frailty distribution parameters. This interpretation, of course, has to be taken only at an intuitive level since the correction induced with the frailty model M3 is not interpretable in terms of a single missing characteristic, but it represents a combination of missing characteristics such as drug use (most likely tobacco in this case), the presence of comorbidities among other lifestyle related diseases and its impact on the general population mortality, and etcetera.

In order to evaluate the effect of not having deprivation-specific life tables, we have also fitted excess hazard models using life tables without the deprivation variable (i.e. national life tables). The results obtained with the deprivation-specific life tables and those obtained with the national life tables (excluding deprivation) are very similar (see Section 5, Table 19, in the Appendix of the supplementary material available at *Biostatistics* online), which is due to the high lethality of lung cancer (inducing negligible differences of population mortality between deprivation groups). Figure 1 in the Appendix of the supplementary material available at *Biostatistics* online shows the net survival curves obtained for the whole population with models M1–M3 as well as the non-parametric Pohar-Perme estimator (Perme *and others*, 2012). We observe that model M1 and the Pohar-Perme estimator are virtually the same, which indicates that M1 can properly capture time-dependent effects (Rubio *and others*, 2018), an assumption made for the correction models. Model M2 produces a net survival curve which is consistently above that obtained with M1. The net survival curve obtained with M3 is above all others, which is explained by the fact that most of the probability mass (77%) of the frailty correction is above 1. We compared estimates of the excess mortality hazard and the corresponding net survival for two subgroups of patients (Figure 3). For stage IV patients, the differences between each model is almost not visible (upper panels), while the difference between models M1–M3 could be more clearly seen in stage II patients subgroup (lower panels).

6. DISCUSSION

6.1. *Summary of findings*

Using the general hazard structure in Rubio *and others* (2018), we have proposed excess hazard regression models that can account for mismatches in the life tables induced by the unavailability of information on relevant population characteristics. The correction models based on a frailty distribution account for non-specific mismatches in the life table, in the sense that the correction is not associated to the lack of known specific variables for constructing the life tables, but to the effect of potentially several unavailable characteristics, which is allowed to be different for each patient. This is the main difference with Cheuvart's model (2.2), which we used here for comparison purpose even though this model was mainly developed in the context of a randomized clinical trial when a selection bias of the patient population is expected. Thus, Cheuvart's model assumes the same constant correction parameter for all patients, which is the main difference with our proposed frailty correction models used in population-based cancer registry data. We have shown that the proposed frailty correction models are able to properly identify and correct these mismatches in several simulation scenarios, provided that the sample size is large enough (5000 or more). Not accounting for mismatched life tables in the relative survival setting may lead to inappropriate net survival comparisons between populations. The need for relatively large samples in order to identify mismatches in the life tables is unsurprising, as there may be several reasons why general population life tables are not fit for our cancer patient population (drug use, lifestyle related diseases, deprivation, ethnicity, and etc.), and only a large enough sample would guarantee that the data contain enough information

Table 2. *Regression parameter estimates (standard errors) using models M1–M3, with their corresponding AIC on the men lung cancer dataset. Note: The time- dependent effects are indicated with -t. For model M2, γ is estimated, while μ is estimated for model M3. Age, age at diagnosis (centered at 70, and divided by 10), Dep, Income Deprivation Score (centered at 0.1, and divided by 10); CV, CardioVascular comorbidity; COPD, chronic obstructive pulmonary disease; AIC, Akaike information criteria (best model indicated in bold font)*

	M1	M2	M3
b	—	—	9.83 (3.03)
$\gamma \mid \mu$	—	2.7 (0.21)	6.54 (0.91)
θ	0.05 (0.01)	0.03 (0.01)	0.03 (0.01)
κ	0.38 (0.01)	0.35 (0.01)	0.34 (0.01)
α	4.64 (0.34)	5.64 (0.48)	5.92 (0.58)
Age-t	0.29 (0.04)	0.29 (0.04)	0.16 (0.05)
Dep-t	0.11 (0.04)	0.12 (0.04)	0.09 (0.04)
Stage 1-t	−2.66 (0.25)	−2.17 (0.32)	−5.4 (1.4)
Stage 2-t	−2.2 (0.2)	−2 (0.22)	−2.69 (0.35)
Stage 3-t	−1.66 (0.11)	−1.57 (0.11)	−1.75 (0.13)
CV-t	0.31 (0.11)	0.31 (0.11)	0.42 (0.11)
COPD-t	0.13 (0.11)	0.08 (0.12)	0.37 (0.14)
Age	0.27 (0.01)	0.23 (0.02)	0.16 (0.02)
Dep	0.06 (0.01)	0.06 (0.01)	0.04 (0.01)
Stage 1	−2.84 (0.06)	−3.13 (0.1)	−3.53 (0.36)
Stage 2	−2.16 (0.06)	−2.32 (0.07)	−2.65 (0.1)
Stage 3	−1.23 (0.03)	−1.27 (0.04)	−1.36 (0.04)
CV	0.24 (0.04)	0.26 (0.04)	0.3 (0.04)
COPD	0.19 (0.04)	0.17 (0.04)	0.25 (0.05)
AIC	20 304.69	20 241.27	**20 213.41**

to adjust for the unavailability of those variables. Intuitively, the information about these correction parameters is provided by the sample of differences of the population cumulative hazards $H_P(A_i + t_i; \mathbf{z}_i) - H_P(A_i; \mathbf{z}_i)$. The implicit assumptions behind the proposed correction model M3 (2.8) are:

(i) The set of covariates $\mathbf{x}$ includes the relevant cancer-specific variables. Thus, all missing information (if any) is produced from a mismatch of the population mortality rate.

(ii) The model $h_E(t; \mathbf{x})$ is properly specified. This is, the fitted excess hazard model is flexible enough to approximate the excess hazard.

Assumption (i) reflects the fact that the model was constructed to only capture mismatches in the life tables, since the correction parameter only affects the population mortality hazard. For instance, in our lung cancer data example, we assume that the excess hazard is accurately explained by age, sex, deprivation, tumor stage, and comorbidity. However, a potential risk factor that is not included there is the smoking status, which is not available at the population level in England. This risk factor might affect both the background mortality hazard (because of diseases or complications associated with smoking) and the cancer-related mortality hazard, in the case that patients continue smoking after the diagnosis of cancer. Indeed, smoking is a driver of lung cancer incidence, but its impact on cancer-related mortality hazard may not be that clear because comorbidity conditions already accounts (at least partially) to smoking-related
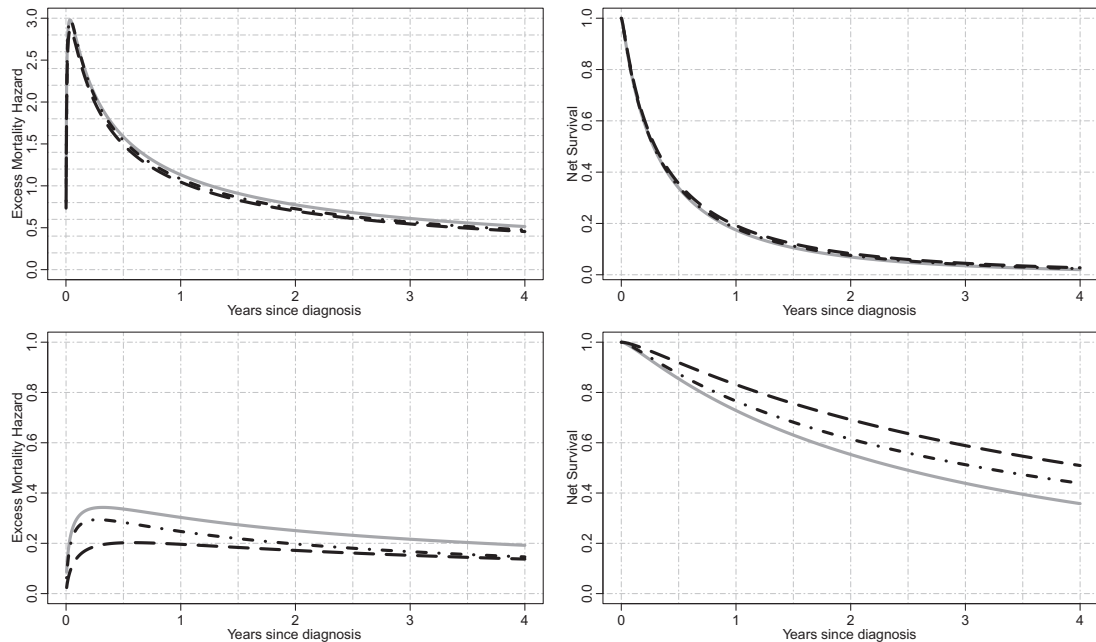
Fig. 3. Illustration for lung cancer patients: Excess mortality hazard (left panels) and the corresponding net survival (right panels) for men aged 70 years at diagnosis, with Income Deprivation score equals to 0.1 (i.e. least deprived), without Cardiovascular comorbidity nor COPD, and with stage IV cancer at diagnosis (upper panels) or stage II cancer at diagnosis (lower panels). M1=solid grey lines, M2=dot-dashed black lines, M3=long-dashed black lines.

complications. Thus, the main interest of including smoking status in the predictor variables of the cancer-related mortality hazard would be for patients who continue smoking after the diagnosis of cancer, and it would certainly be interesting to explore the effect of including this variable in the model once it becomes available. Assumption (ii) is important since model misspecification can also affect the correction made on the population hazard. If a covariate which appears in $\mathbf{z}$ and $\mathbf{x}$ is wrongly modeled in $h_E(t; \mathbf{x})$ (e.g. not accounting for time-dependent effects), this may also affect the correction. In principle, this is not an onerous condition since one would usually aim at properly modeling the excess hazard, which is typically the main function of interest. Moreover, recent developments in the use of splines and parametric models (Royston and Parmar, 2002; Giorgi *and others*, 2003; Nelson *and others*, 2007; Remontet *and others*, 2007, 2018; Charvat *and others*, 2016; Rubio *and others*, 2018) allow for a tractable inclusion of non-linear and time-dependent effects in excess hazard models. We also assume that there is enough heterogeneity about the unobserved variables of interest. For instance, if we want to assess the impact of mismatched life tables in terms of deprivation, we assume that the sample contains large enough numbers of individuals with different deprivation levels. The amount of data required to accurately estimate the parameters of the correction models has been explored through a simulation study. Certainly, we would not recommend trying to correct for mismatches in the life tables in samples containing substantially fewer than 5000 observations, or with high censoring rates (e.g. higher than 50%). Overall, we have found that model M3 is a good option for accounting for mismatches in the life tables, provided a large enough sample. Its use however is not automatic and should be analyzed on a case by case basis. Comparing the results between corrected (M3) and uncorrected (M1) models, as well as the non-parametric Pohar-Perme estimator, is advisable in practice, in addition to the use of expert knowledge from clinicians or epidemiologists in order to understand and explain the source of the mismatches.

6.2. *Other models in the literature: shared and correlated frailty models*

Zahl (1997) describes two extensions of long-term excess hazards models, where the main goal is to account for an increased risk of dying of other diseases in patients with certain cancers. The first extension consists of a shared frailty model in which a random effect (frailty) is multiplied by the population hazard and the excess hazard as follows:

$$h_o^S(t; \mathbf{x} \mid \gamma) = [h_P(A + t; y + t, \mathbf{z}) + h_E(t; \mathbf{x})]\gamma, \tag{6.10}$$

where $\gamma \sim G$, and G is a distribution with positive support, typically chosen to be a Gamma distribution with unknown shape and scale parameters. Perhaps unsurprisingly, the induced model is non-identifiable unless the random effect has unit mean. Given that the Gamma distribution is asymmetric, the assumption of unit mean implies that $\mathbb{P}(\gamma \leq 1) > \mathbb{P}(\gamma > 1)$, which may not be a reasonable assumption in some scenarios since this implies that there is a higher probability of requiring a shrinking correction to the population hazard ($\gamma \leq 1$) than an increasing one ($\gamma > 1$). For a general framework of frailty hazard models we refer the reader to Aalen *and others* (2008).

Zahl (1997) proposed a correlated frailty model, by using frailties on both the population hazard and the excess hazard:

$$h_o^Z(t; \mathbf{x} \mid \gamma_1, \gamma_2) = h_P(A + t; y + t, \mathbf{z})\gamma_1 + h_E(t; \mathbf{x})\gamma_2, \tag{6.11}$$

where $(\gamma_1, \gamma_2) \sim G_2$, and G_2 is a bivariate distribution with support on the positive quadrant. Intuitively, it is difficult (if at all feasible) to obtain information about the factors affecting the population hazard and the excess hazard (which is typically a flexible parametric model), and the dependencies between them, simultaneously. In fact, Zahl (1997) found that the maximum likelihood estimators of the parameters of model (6.11) do not exist, suggestive of identifiability issues of this model. Zahl (1997) proposed a number of restrictions of the parameter space (to a compact set) in order to alleviate these estimation issues. However, even after those restrictions, the MLE was on the boundary of the restricted parameter space, which suggests remaining lack of identifiability.

6.3. *Further research*

From the results of our simulation study, we have observed a larger variability in the estimators of additional parameters corresponding to the correction of the background mortality hazard. In order to reduce this variability, penalized maximum likelihood estimation methods could be used to shrink the correction parameter (i.e. for M2 or for M3) towards the value 1. This will be explored in future research.

From the simulation study, the coverage proportions of the additional parameter correcting the life table were lower than the nominal value. Using a robust estimator of the variance for this additional parameter may be an option to reach a better coverage, as may be calculating profile likelihood intervals.

Another extension of model (2.2) consists of modeling the correction parameters γ and μ in terms of a set of covariates, say $\mathbf{w}$. A related approach has recently been studied in Touraine *and others* (2019). Possible limitations include the inferential challenges in estimating $q \geq 2$ (the dimension of $\mathbf{w}$) when the sample size is not large enough. In addition, the assumption of proportional population hazards is often too restrictive in the cancer survival field. In practice, one natural question is whether the Gamma frailty distribution is flexible enough to model the random correction. Using maximum likelihood estimation implies that the estimators of the parameters of the frailty distribution will converge to the values that minimize the distance (in fact, the Kullback–Leibler divergence) to the true generating model. Section 6, Appendix Tables 20–21 and Appendix Figures 17 and 18 of the supplementary material available at *Biostatistics* online, shows a simulated example where the random correction is simulated from a lognormal

distribution (instead of Gamma). This example indicates that model M3 has a good performance even if the random corrections are not generated from a Gamma distribution, but as long as the Gamma distribution can approximate the shape of the true generating distribution. A possible extension consists of using a more flexible frailty distribution with a tractable Laplace transform, in order to obtain tractable expressions for the hazard and cumulative hazard functions. An attractive option is the power variance function (PVF) family of distributions (Aalen *and others*, 2008), which contains three parameters instead of two. This, of course, complicates the estimation process.

Software

Software in the form of R code, together with a sample input data set and complete documentation is available under request, and an R Markdown document entitled "Simulation design I: Excess hazard models for insufficiently stratified life tables" is available on the website http://www.rpubs.com/FJRubio/FGH and the GitHub repository https://github.com/FJRubio67/ExcessHazardModels.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

Funding

References

Aalen, O., Borgan, O. and Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View*. New York: Springer.

Benitez-Majano, S., Fowler, H., Maringe, C., Di Girolamo, C. and Rachet, B. (2016). Deriving stage at diagnosis from multiple population-based sources: colorectal and lung cancer in England. *British Journal of Cancer* **115**, 391.

Bower, H., Andersson, T. M. L., Crowther, M. J., Dickman, P. W., Lambe, M. and Lambert, P. C. (2018). Adjusting expected mortality rates using information from a control population: an example using socioeconomic status. *American Journal of Epidemiology* **187**, 828–836.

Charvat, H., Remontet, L., Bossard, N., Roche, L., Dejardin, O., Rachet, B., Launoy, G. and Belot, A. (2016). A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates. *Statistics in Medicine* **35**, 3066–3084.

Cheuvart, B. and Ryan, L. (1991). Adjusting for age-related competing mortality in long-term cancer clinical trials. *Statistics in Medicine* **10**, 65–77.

Dickman, P. W., Auvinen, A., Voutilainen, E. T. and Hakulinen, T. (1998). Measuring social class differences in cancer patient survival: is it necessary to control for social class differences in general population mortality? A Finnish population-based study. *Journal of Epidemiology and Community Health* **52**, 727–734.

ELLIS, L., COLEMAN, M. P. AND RACHET, B. (2014). The impact of life tables adjusted for smoking on the socio-economic difference in net survival for laryngeal and lung cancer. *British Journal of Cancer* **111**, 195.

ESTEVE, J., BENHAMOU, E., CROASDALE, M. AND RAYMOND, L. (1990). Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* **9**, 529–538.

GIORGI, R., ABRAHAMOWICZ, M., QUANTIN, C., BOLARD, P., ESTÈVE, J., GOUVERNET, J. AND FAIVRE, J. (2003). A relative survival regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine* **22**, 2767–2784.

GRAFFÉO, N., JOOSTE, V. AND GIORGI, R. (2012). The impact of additional life-table variables on excess mortality estimates. *Statistics in Medicine* **31**, 4219–4230.

MARINGE, C., FOWLER, H., RACHET, B. AND LUQUE-FERNANDEZ, M. A. (2017). Reproducibility, reliability and validity of population-based administrative health data for the assessment of cancer non-related comorbidities. *PLoS One* **12**, e0172814.

NELSON, C. P., LAMBERT, P. C., SQUIRE, I. B. AND JONES, D.R. (2007). Flexible parametric models for relative survival with application in coronary heart disease. *Statistics in Medicine* **26**, 5486–5498.

PAVLIČ, K. AND POHAR-PERME, M. (2018). Using pseudo-observations for estimation in relative survival. *Biostatistics*. https://doi.org/10.1093/biostatistics/kxy008.

PERME, M. P., STARE, J. AND ESTÈVE, J. (2012). On estimation in relative survival. *Biometrics* **68**, 113–120.

POHAR-PERME, M., ESTÈVE, J. AND RACHET, B. (2016). Analysing population-based cancer survival–settling the controversies. *BMC Cancer* **16**, 933.

POHAR PERME, M., HENDERSON, R. AND STARE, J. (2009). An approach to estimation in relative survival regression. *Biostatistics* **10**, 136–146.

REMONTET, L., BOSSARD, N., BELOT, A. AND ESTEVE, J. (2007). An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. *Statistics in Medicine* **26**, 2214–2228.

REMONTET, L., UHRY, Z., BOSSARD, N., IWAZ, J., BELOT, A., DANIELI, C., CHARVAT, H. AND ROCHE, L. (2018). Flexible and structured survival model for a simultaneous estimation of non-linear and non-proportional effects and complex interactions between continuous variables: performance of this multidimensional penalized spline approach in net survival trend analysis. *Statistical Methods in Medical Research*. https://doi.org/10.1177/0962280218779408.

ROYSTON, P. AND PARMAR, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* **21**, 2175–2197.

RUBIO, F. J., REMONTET, L., JEWELL, N. P. AND BELOT, A. (2018). On a general structure for hazard-based regression models: an application to population-based cancer research. *Statistical Methods in Medical Research*. https://doi.org/10.1177%2F0962280218782293.

TOURAINE, C., GRAFFÉO, N., GIORGI, R. and the CENSUR working survival group. (2019). More accurate cancer-related excess mortality through correcting background mortality for extra variables. *Statistical Methods in Medical Research*. https://doi.org/10.1177%2F0962280218823234.

WOODS, L. M., RACHET, B., RIGA, M., STONE, N., SHAH, A. AND COLEMAN, M. P. (2005). Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation. *Journal of Epidemiology and Community Health* **59**, 115–120.

WRIGHT, S. J. (2015). Coordinate descent algorithms. *Mathematical Programming* **151**, 3–34.

ZAHL, P. H. (1997). Frailty modelling for the excess hazard. *Statistics in Medicine* **16**, 1573–1585.