Time-Dependent and Non-linear Predictor Effects in Survival Analyses: A Case Study Comparing Alternative Models for Cancer Mortality

Michal Abrahamowicz[1(✉)], Marie-Eve Beauchamp[2], Richard J. Cook[3], Malka Gorfine[4], Jason Agulnik[5], Bruno Gagnon[6], and Steve Ferreira Guerra[1]

[1] McGill University, Montreal, QC H3A 1G1, Canada
michal.abrahamowicz@mcgill.ca
[2] Research Institute of the McGill University Health Centre, Montreal, QC H4A 3S5, Canada
[3] University of Waterloo, Waterloo, ON N2L 3G1, Canada
[4] Tel Aviv University, 6997801 Tel Aviv, Israel
[5] McGill University, Montreal, QC H3T 1E2, Canada
[6] Université Laval, Québec, QC G1J 1Z4, Canada

Abstract. Regression analysis with multivariable survival data requires specification of a model describing the relationship between predictors and some function of the event time distribution. Popular choices include proportional hazards (PH), accelerated failure time (AFT), and additive hazards (AH) models. Each model imposes an *a priori* assumption that, respectively, hazard ratios, relative time scales, or hazard differences, associated with a given change in a predictor value, are constant during the entire follow-up period. However, the effects of *some* of the predictors of interest may not be consistent with the underlying modeling assumption, which requires extending the model to include time-dependent effects. In addition, for each continuous covariate a suitable functional form of its relationship with the outcome has to be determined. Several flexible methods for addressing these modeling challenges were proposed in the literature but there is little evidence regarding head-to-head comparisons of flexible extensions of PH vs. AFT vs. AH models in real-world analyses. We first present a brief overview of selected flexible methods available to estimate time-dependent effects and, for continuous variables, non-linear effects. We also identify the software that allows the implementation of such computationally intensive flexible models. The practical importance of these challenges is illustrated using a case study of prognostic factors associated with cancer mortality.

Keywords: Biostatistics · Survival Analysis · Splines · Prognostic Studies

1 Introduction

The ultimate goal of statistical research is to develop validated methods to allow accurate analyses of real-world data. One of the most common goals of real-world analyses is to explore associations of different risk and prognostic factors, treatments, lifestyle

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15887, pp. 393–410, 2026.
https://doi.org/10.1007/978-3-031-97589-9_27

characteristics and environmental or occupational exposures with relevant outcomes. If the outcome is a binary (presence/absence) or a continuous variable, then its associations with multiple covariates can be directly modeled. In contrast, if the outcome is defined as a time to a specific event, such as death or cancer diagnosis, the analyses are complicated by the fact that for some study participants no outcome may be observed until the end of the study. This results in right-censored observations, and requires more specialized analytical methods, being developed by statisticians working in the field of survival analysis or time-to-event analysis. One common challenge in this field is that the outcome, for individual i, is bivariate, i.e. includes (a) a continuous value of total follow-up time T_i, and (b) a binary indicator of status at time T_i, with $\Delta_i = 1$ indicating the event and $\Delta_i = 0$ a censored observation [1]. Thus, while using regression modeling to estimate and test the associations of interest, the dependent variable has to be represented as some function of the event times rather than the directly observed individual (bivariate) outcome. Accordingly, in statistical literature, several alternative models were proposed for regression analyses of right-censored time-to-event outcomes, each of which postulates a different mathematical relationship between the predictors and some function of event times.

In real-world applications of survival analysis, by far the most popular regression model is the proportional hazards (PH) model, typically estimated using the maximum partial likelihood estimation procedure developed by Sir David Cox [2], in the original publication that has close to 64,000 citations on Google Scholar. The PH model estimates the effects of predictors on the hazard, i.e. the instantaneous risk, and the popularity of the Cox model is largely related to the fact that it avoids the need to specify the parametric form of the hazard function [2]. However, more recently the Cox model has been criticized for so-called built-in selection bias [3], i.e. systematic bias toward a null effect induced by the presence of an unmeasured or unaccounted for risk factor or "frailty" [4]. An alternative regression model for survival analysis is the additive hazards (AH) model proposed by Aalen [5, 6]. In the AH model, the predictors are also acting on the hazard function but, in contrast to the PH model, their effects are assumed to be additive, i.e. are quantified using hazard differences [5], rather than multiplicative hazard ratios (HRs). Finally, more recently, there is an increased interest in the accelerated failure time (AFT) model, which assumes that predictors affect directly the event time and, thus, their effects are estimated using multiplicative event time ratios [7]. In their classic forms, each of the three models imposes a strong assumption that, respectively, hazard ratios [2], hazard differences [8] or event time ratios [7] associated with a given change in the predictor are constant across the entire follow-up period. Yet, interestingly, only if the underlying hazard function is constant, corresponding to the simple exponential model, the AFT, AH and PH models are compatible with each other [7]. In all other situations, if the data are truly generated e.g. by the AH model, then the assumptions underlying the AFT and PH models are violated, and *vice versa.* However, even if the different models may be more suitable for different datasets, in a vast majority of real-world survival analyses, the PH, AFT or AH model is selected *a priori,* without presenting any clinical and/or statistical evidence supporting such choices. Indeed, the head-to-head comparisons of any two of the three regression models for survival analysis are very

limited [9], and we are not aware of any empirical study that compared all three models for the same dataset.

A related issue is that if the assumptions underlying a given *a priori* selected model are violated, then the resulting estimates are biased and lead to incorrect interpretation of the way predictors are associated with the outcome. For example, if the Cox model is selected even if the PH assumption is violated, the resulting estimates will incorrectly suggest that a given change in the predictor value is associated with a constant HR, while in reality the true HR varies across the follow-up period [10]. In fact, in his original AH publications, Aalen has proposed a flexible formulation of the AH model that allows for hazard differences to vary during follow-up [5]. In a similar spirit, in the 1990's several authors have proposed various flexible extensions of the PH model, each of which uses some spline modeling to estimate time-dependent (TD) hazard ratios (see, e.g., a brief review in [10]). Finally, recently, two different extensions of the AFT model have been proposed to allow flexible modeling of TD predictor effects [11, 12]. Yet, again, except for a comparison of selected models for TD effects in the PH framework, there is a paucity of empirical head-to-head comparisons of TD extensions of PH vs. AH vs. AFT models. Furthermore, additional modeling assumptions are required regarding the functional form of the, possibly non-linear, dose-response function expressing how the outcome (e.g., log hazard in the PH framework) is associated with a continuous predictor. In the PH framework, we have demonstrated that the non-linear (NL) and TD effects of the same continuous predictor have to be assessed jointly to avoid biased estimates and incorrect conclusions [13]. We have then developed and validated flexible extensions of the PH model [14] and, more recently, AFT model [12] that allow for joint estimation of TD and NL effects of multiple predictors on the hazard, and likelihood ratio tests (LRTs) of their statistical significance. However, no comparison of TD/NL estimates between PH vs. AFT models, applied to the same empirical or simulated dataset(s), has been reported in the literature to date. Furthermore, to the best of our knowledge, no publications have reported NL estimates in the AH modeling framework, and the fundamental AH papers do not discuss explicitly flexible modeling of the functional forms for continuous covariates, either under constant additive effects [8] or under time-dependent effects [5].

To fill the aforementioned knowledge gaps, the current study carries out empirical head-to-head comparisons of PH, AH and AFT models and their flexible extensions with TD and/or NL effects of predictors on the hazard. The various models are employed in a real-world prognostic study of predictors of all-cause mortality in a non-small cell lung cancer. In Sect. 2, we describe the models being compared, including their flexible TD/NL extensions. Then, Sect. 3 describes how these models are employed to assess the effects of different biomarkers for mortality in non-small cell lung cancer. A brief discussion concludes the paper.

2 Alternative Regression Models for Survival Analyses

In this section, we present the alternative regression models for multivariable survival analyses. For each of the three models – PH, AFT and AH models – we first define its classic version, in which the effect of each covariate is estimated using a single parameter and thus is constrained to be constant-over-time and, for each continuous predictor, a

linear dose-response relationship is assumed. Next, we introduce the flexible extensions of each of the three models that allow the modeling of both (i) time-dependent (TD) effects of either binary or continuous covariates, and (ii) possibly non-linear (NL) effects of continuous covariates.

2.1 Proportional Hazards Model and Its Flexible Extensions

The classic proportional hazards (PH) model assumes that predictors have multiplicative constant-over-time effects on the hazard. The model is typically written as [2]:

$$\lambda(t|\mathbf{X}) = \lambda_0(t)\exp(\mathbf{X}^T\boldsymbol{\beta}) \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard, corresponding to a hypothetical reference population with all predictor values equal to 0 ($\mathbf{X} = 0$), $\mathbf{X} = (X_1, \ldots, X_p)$ is a vector of p, possibly time-dependent, predictors, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ the vector of regression coefficients, representing the log hazard ratios (HRs). In the classic PH model (1), the adjusted HRs for all predictors are constrained to be constant over time and the continuous predictors are assumed to have linear relationships with the log hazard.

The regression parameters ($\boldsymbol{\beta}$) of the PH model (1) are commonly obtained using maximum partial likelihood estimation proposed by Cox [2], in which the baseline hazard is a nuisance parameter. However, it is possible to get the full maximum likelihood estimation (MLE) using, for example, the approach developed and validated in our earlier study [15]. More details are presented below, in the context of the flexible extension of the PH model.

In the 1990's different flexible extensions of the PH model were proposed, most involving spline modeling of either TD or NL effects of predictors (see, e.g., a brief review in [10]). However, we have demonstrated that the NL and TD effects of the same continuous predictor must be assessed jointly, i.e. adjusted for each other, to avoid biased estimates and incorrect conclusions [13]. Therefore, we proposed the following flexible extension of the PH model, which incorporates both the TD effects and, for continuous predictors, the NL effects on the log hazard [13]:

$$\lambda(t|\mathbf{X}) = \lambda_0(t)\exp\left[\boldsymbol{\beta}(t)\mathbf{g}(\mathbf{X})\right] \tag{2}$$

where the TD effect is represented by $\boldsymbol{\beta}(t)$ which models the change over time in the log(HR) for the effect of $\mathbf{X}$, and $\mathbf{g}(\mathbf{X})$ represents the NL effect of $\mathbf{X}$ which models how the log(HR) changes with increasing value of $\mathbf{X}$. In the flexible model (2), $\boldsymbol{\beta}(t)$ and $\mathbf{g}(\mathbf{X})$ can be modeled using (unpenalized) polynomial regression B-splines [13].

Direct estimation of model (2) will be impossible due to an identifiability problem that occurs because the $\boldsymbol{\beta}(t)$ and $\mathbf{g}(\mathbf{X})$ functions for the same continuous predictor X are multiplied by each other [13]. To avoid such problems, we have proposed to use the iterative alternating conditional estimation (ACE) [14] that we have validated in extensive simulations [16]. The ACE algorithm alternates between two steps, each involving conditional maximization of the partial likelihood for either $\boldsymbol{\beta}(t)$ or $\mathbf{g}(\mathbf{X})$, while holding the parameters of the other function fixed at their current values. This repeated updating process allows ACE to circumvent nonidentifiability constraints.

Because the constant-in-time $\boldsymbol{\beta}$ parameter and the linear function are both special cases of the polynomial B-spline, the classic PH model (1) is nested within the flexible model (2) [13]. Thus, statistical significance of, respectively, the TD and NL effects can be tested using model-based likelihood ratio tests (LRTs) [13].

2.2 Accelerated Failure Time Model and Its Flexible Extension

Another important class of survival models is provided by the accelerated failure time (AFT) class of models [7, 17]. In its classic form, the AFT model can be written as:

$$\log(T) = -\mathbf{X}^T\boldsymbol{\delta} + W \tag{3}$$

where T denotes the event time, $\mathbf{X}$ is again the vector of covariates, and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)$ is the corresponding vector of regression coefficients. W is an error term assumed to follow some distribution (e.g., Gaussian, logistic, or extreme value) which determines the distribution of T [7]. In contrast to the PH and AH models, where the covariates are assumed to modify the hazard function, either multiplicatively or additively, the AFT parametrization in (3) implies that changes in the predictor value are associated with an additive change in the logarithm of the survival time, implying a multiplicative change in the expected event time. In particular, a one-unit increase in a given continuous predictor X_j is associated with increasing the log survival time by the corresponding regression coefficient, δ_j in Eq. (3) [7]. Thus, the AFT model regression parameters $\boldsymbol{\delta}$ directly quantify the impact of covariates on the time scale, extending or shrinking the expected survival time by $\exp(\boldsymbol{\delta})$, which can be intuitively appealing to end-users and clinicians. On the other hand, specifying the error distribution can be challenging in practice, and its misspecification may lead to biased results, as demonstrated in simulations [18, 19]. To avoid this limitation, recently Pang *et al.* [18] have proposed a more flexible AFT model that replaces parametric assumptions about the event time distribution by cubic spline modeling of the baseline hazard and demonstrated its advantages in simulations.

An alternative way to present the AFT model, mathematically equivalent to the classic formulation in (3), describes the predictor effects on the hazard function [7]:

$$\lambda(t|\mathbf{X}) = \exp\left(\mathbf{X}^T\boldsymbol{\delta}\right)\lambda_0\left(\exp\left(\mathbf{X}^T\boldsymbol{\delta}\right)\cdot t\right) \tag{4}$$

where $\lambda_0(\cdot)$ is the baseline hazard, corresponding to a hypothetical reference population with all predictors equal to 0 ($\mathbf{X} = 0$). Notice that the regression parameter vector $\boldsymbol{\delta}$ in (4) has exactly the same interpretation as in (3). Model (4) reveals that in the AFT framework, predictors $\mathbf{X}$ affect the hazard in two different ways. Firstly, there is a "vertical" shift: at a given time t the baseline hazard is multiplied by $\mathbf{X}^T\boldsymbol{\delta}$. Secondly, there is a "horizontal" shift where the time t in $\lambda(t)$ is multiplied by the same constant $\mathbf{X}^T\boldsymbol{\delta}$.

Both AFT models (3) and (4) constrain the estimated predictor effects to be constant-over-time and linear. Although, already in 1984, Cox and Oakes [7] outline the general approach for incorporating TD predictor effects in the AFT setting, practical implementations and real-world applications emerged only recently. Crowther, Royston, and Clements [11] introduced a flexible AFT framework, which allows estimating TD acceleration factors. However, they do not discuss the estimation of NL effects of continuous

predictors on the logarithm of the event time [11]. To the best of our knowledge, the first extension of the AFT model that incorporates both TD and NL effects was proposed by Pang *et al.* [12]. These authors developed a flexible hazard-based AFT model that extends model (4) to include joint spline-based full maximum likelihood estimation of (i) the baseline hazard, (ii) TD covariate effects on the hazard, and (iii) nonlinear (NL) relationships for continuous predictors. Their flexible AFT model is written as [12]:

$$\lambda(t|\mathbf{X}) = \exp\Big(\mathbf{g}(\mathbf{X})^T\boldsymbol{\delta}(t)\Big)\lambda_0\Big(\exp\Big(\mathbf{g}(\mathbf{X})^T\boldsymbol{\delta}(t)\Big)\cdot t\Big) \tag{5}$$

where $\boldsymbol{\delta}(t)$ models the change over time in the log acceleration factor, and $\mathbf{g}(\mathbf{X})$ represents the NL effects of $\mathbf{X}$ that estimate its associations with the log event time. To avoid aforementioned identifiability problems, Pang *et al.* [12] adapted the ACE algorithm outlined in Sect. 2.1 above. Statistical significance of the NL and TD predictor effects is assessed using model-based LRTs.

2.3 Additive Hazards Model and Its Flexible Extensions

Another alternative to the proportional hazards (PH) formulation (1) is the additive hazards (AH) model [20, 21]. In contrast to the PH model (1), where covariates act multiplicatively on the baseline hazard, the AH model assumes that changing the predictor value is associated with adding or subtracting a constant to/from the baseline hazard function. Formally, the AH model is written as:

$$\lambda(t|\mathbf{X}) = \lambda_0(t) + \mathbf{X}^T\boldsymbol{\alpha} \tag{6}$$

where $\lambda_0(\cdot)$ is the baseline hazard, defined as in Eqs. (1) and (4), and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ are the regression coefficients. Here, each α_j represents the difference in the hazard associated with a one-unit increase in the jth covariate, holding other covariates fixed. Various estimation procedures were proposed for AH models [6, 21, 22]. The interpretation of coefficients in AH models may be considered more intuitive in settings where absolute differences in hazard rates are of direct interest. However, the estimation has to ensure that the resulting hazards, associated with all covariate vectors present in a given analysis, are always strictly positive.

In the classic AH model (6), the estimated hazard differences are assumed to be constant-over-time and linear for continuous predictors. Although various forms of AH models have been proposed and used [6, 8], the original publications by Aalen already included an extension of the AH model that incorporated non-parametric estimation of time-dependent effects of predictors [5, 6]. Specifically, the Aalen's AH model is written as:

$$\lambda(t|\mathbf{X}) = \lambda_0(t) + \mathbf{X}^T\boldsymbol{\alpha}(t) \tag{7}$$

where each $\alpha_j(t)$ is allowed to vary flexibly over time. Notice that in Aalen's AH model (7), the estimation and interpretation of TD effects focus on the cumulative functions:

$$A_j(t) = \int_0^t \alpha_j(s)\mathrm{d}s. \tag{8}$$

This allows for easy estimation of $A(t)$ through least squares [5, 6], although MLE is also possible [22]. At any particular time t, one can interpret the estimate of the cumulative function $\widehat{A}_j(t)$ in (8) as the accumulated effect of a one-unit increase in X_j on the hazard up to time t. The corresponding estimates $\widehat{\alpha}_j(t)$ of the TD effect of X_j on the current hazard at time t can be obtained by taking the derivative of $\widehat{A}_j(t)$ with respect to t. TD effects in Aalen's model (7) can be tested via two alternative null hypotheses, either $H_{01} : \alpha_j(t) = 0$ or $H_{02} : \alpha_j(t) = c$, for a given constant c (see [23] for details).

Aalen's non-parametric AH model (7) has been further extended by McKeague & Sasieni [8] to allow for some predictors to have TD effects and others constant-in-time effects. Their semi-parametric AH model is written as [8]:

$$\lambda(t|\mathbf{X}) = \lambda_0(t) + \mathbf{X}^T\boldsymbol{\alpha}(t) + \mathbf{Z}^T\boldsymbol{\eta} \tag{9}$$

where the $\boldsymbol{\eta}$ components represent fixed-in-time coefficients for predictors $\boldsymbol{Z}$.

Neither Aalen [5, 6] nor McKeague and Sasieni [8] discuss how the respective AH models can be extended to include the estimation of NL effects of continuous predictors on the hazard. Indeed, to the best of our knowledge, there is no publicly available software that allows such modeling. To address this gap, below we propose a pragmatic approach for flexible modeling of additive NL effects of continuous predictors on the hazard in the AH framework, and indicate how statistical significance of departures from the linear effects can be tested. Our approach involves two steps for each predictor. The first step focuses on the estimation of the NL effect of a continuous predictor X_j. To this end, in the multivariable AH model that adjusts for other variables as constant-in-time covariates, we include two separate terms related to X_j. The first term represents the untransformed X_j and models its linear association with the hazard. The second term represents the departure from linearity, modeled using unpenalized cubic B-splines, with 1 interior knot at the median of the sample distribution of X_j. To avoid identifiability problems, one of the spline coefficients has to be constrained to 0 [13], this implies a 4-degree-of-freedom (df) function. Next, the spline coefficients, estimated at the first step, are used to calculate, for each study participant i, a new variable representing the spline-transformed value of the predictor, $BS[X_{i,j}]$. Then, in the second step, to test the statistical significance of the NL effect of X_j, we refit the multivariable model by including both (i) the untransformed value of X_j and (ii) the resulting spline-transformed value $BS[X_j]$. Finally, to test whether X_j has a statistically significant NL effect, when adjusting for its linear effect, we simply test whether the coefficient for $BS[X_j]$ is significantly different from 0.

2.4 Software

The classic PH model was estimated with the `coxph` function in the survival R package [24]. The latest version of the R scripts to estimate the flexible extension of the PH model [14] is available at https://github.com/mebeauchamp/CoxFlex. R scripts for implementing the AFT model with splined-based estimation of the baseline hazard [18] can be found at https://github.com/MenglanPang/Spline-based-AFT-Model. R scripts for the flexible extension of the AFT model including TD/NL effects [12] are available at https://github.com/MenglanPang/Flexible-AFT-Model. Aalen's AH model (7) and

McKeague and Sasieni's model (9) can be fit using the `aalen` function of the timereg R package. The new extension of the AH model that allows B-spline modeling of NL effects (Sect. 2.3) can be implemented using our R code shown in Appendix.

3 Case Study of Prognostic Factors for Mortality in Lung Cancer

3.1 Data Source

In this section, to provide head-to-head comparisons between PH, AH and AFT models, and their flexible extensions, we present a case study of all-cause mortality in non-small cell lung cancer (NSCLC). Lung cancer is one of the most commonly diagnosed cancers globally, and remains the leading cause of cancer-related deaths, with about 1.8 million fatalities worldwide [25]. NSCLC represents 85% of all lung cancer cases, making it the most prevalent subtype [26]. Unfortunately, half of NSCLC cases are diagnosed at the most advanced stage 4, which is associated with a short median survival [27]. However, recent advancements in molecular testing, targeted therapies, and immunotherapy have significantly improved survival in NSCLC [28].

To unravel complex associations of different clinical and demographic variables with mortality in NSCLC, we rely on clinical data collected for a cohort of patients, newly diagnosed with NSCLC, who were treated at the Jewish General Hospital Pulmonary Oncology Clinic in Montreal, QC, Canada, from 9 April 2002 to 18 September 2008 [29]. Their follow-up was continued until 15 March 2009. A total of 269 patients had complete data on all relevant variables and were included in our analyses.

Specifically, we focus on the values, measured at cancer diagnosis, of several plasma biomarkers, which our team has previously demonstrated to improve prognostication of NSCLC outcomes [30]. Albumin, a marker of nutritional status, plays a role in DNA replication and cell proliferation, with low levels linked to poor responses to anticancer therapies [31]. C-reactive protein (CRP) plays a crucial role in the acute phase response, and its production is increased in response to inflammation under the influence of pro-inflammatory cytokines, involved in cancer development [32]. High pre-treatment levels of lactate dehydrogenase (LDH) are associated with a poor prognosis since cancer cells shift from oxidative phosphorylation to aerobic-anaerobic glycolysis, converting pyruvate to lactate [33]. Because high neutrophil-to-lymphocyte ratio is a robust marker of poor overall survival in various tumors, including NSCLC [34], we evaluated the effects of both its components: the absolute neutrophil count (ANC) and the lymphocyte levels. Finally, we assess the impact of cancer stage (high vs. lower), chemotherapy treatment, current smoking, and patient's sex and age.

3.2 Methods of Analysis

Multivariable Modeling of Mortality in NSCLC. All analyses used time-to-event methods for right-censored survival outcomes. Time 0 was defined as the date of NSCLC diagnosis. The event of interest was death of any cause. Study participants who did not have the event until the administrative end of the study on 15 March 2009 were censored at that time. There were no loses to follow-up. All analyses relied on multivariable models that always included the same set of predictors, selected *a priori* based on clinical considerations. Predictors included four binary variables: patient's sex, smoking status at diagnosis (yes vs. no), chemotherapy type (double-agent vs. single-agent), and cancer stage at diagnosis (IIIB + pleural effusion/IV vs. IIIA/IIIB). The six continuous variables, all assessed at the time of NSCLC diagnosis, included the patient's age and values of five biomarkers, briefly described in Sect. 3.1 above: CRP, albumin, LDH, ANC, and lymphocyte levels. Because of very positively skewed distributions, both CRP and LDH values were log-transformed (with base 2), consistently with the way they are typically analyzed [29]. To avoid numerically unstable estimates, all continuous variables were converted to z-scores, with mean 0 and standard deviation of 1.

We first estimated the classic versions of the multivariable PH, AH and AFT models, in which all predictor effects are constrained to remain constant across the entire follow-up time, and all continuous predictors were assumed to have linear dose-response relationships. For full maximum likelihood estimation of the AFT model, the baseline hazard was estimated using a quadratic regression B-spline [18]. Next, we estimated flexible extensions of each of the three models. To decide which of the possible TD and/or NL effects of the predictors result in statistically significant improvement of the model's fit to data and, thus, should be included in a given flexible final model, we relied on a forward selection algorithm described in the next subsection. Finally, we compared the results of flexible extensions of the PH, AH and AFT models. These comparisons focused mostly on two issues. First, we assessed if the final conclusions regarding whether particular predictors had TD and/or NL effects varied depending on whether the analyses relied on PH vs. AH vs. AFT models. Secondly, we compared graphically the corresponding TD and/or NL estimates across the three types of regression models.

Selection of TD and/or NL Effects in the Final Flexible Multivariable Models. The selection algorithm described in this subsection were independently employed for each of the flexible extensions of PH, AFT and AH models.

Earlier real-world analyses and simulation studies [12, 14] indicate that statistical significance of TD or NL effects of a given continuous predictor X_j may vary considerably depending on whether the model includes both: (i) respectively, the NL or TD of the same predictor, and/or (ii) TD and NL effects of other predictors. Therefore, it is important to use a systematic algorithm to build the final multivariable model by selecting those TD and/or NL effects of specific variables that improve the prediction of the outcome. For large datasets with a limited number of predictors, backward elimination may be recommended for this purpose [14]. However, backward elimination requires fitting, at the first step, the full flexible model with TD effects of all predictors and NL effects of all continuous predictors. Given that our dataset includes four binary and six continuous predictors, this would require a model with about 74 df (10*5 df for TD effects + 6*4

df for NL effects). With only 211 deaths, this would definitely violate the minimum requirement of 5 events for each df in the estimated model, resulting in very unstable estimates and inaccurate statistical inference.

To avoid these problems, our analyses relied on forward selection of TD and/or NL effects, which does *not* involve fitting the full flexible model. Forward selection involves starting with a simple model with no TD or NL effects, and gradually building up the model by adding TD and/or NL effects of a single predictor at each step. At the first step, in separate models, we tested the TD and/or (for continuous variable) NL effects of a single variable, while adjusting for constant-over-time and linear effects of the nine other predictors. For binary variables, this involved simply testing statistical significance of their TD effects, at $\alpha = 0.05$. In contrast, for each continuous variable, we had to decide what combination of potential TD and/or NL functions represents best its adjusted effect. Yet, previous simulations indicated that the p-value for the NL effect of X_j may change considerably if the TD effect of X_j is adjusted for, and *vice versa* [13, 14]. Thus, for a continuous variable X_j, we first fit three alternative flexible models with combinations of presence/absence of TD and/or NL effects, while adjusting for constant-over-time linear effects of the nine other predictors. Then, if both TD and NL effects of X_j were significant at $\alpha = 0.05$, while adjusted for each other, X_j was assigned the lower of the two corresponding p-values, and both effects were candidates for the final model. If neither effect was significant when adjusted for each other, but one or both yielded a p-value < 0.05 in separate models that included only one of the two effects, the effect with a lower p-value was considered as a candidate for selection. If neither TD nor NL effect of X_j was significant in any of the three flexible models, these effects were *not* considered as candidates for selection at this step of the forward procedure. After all models for all variables were run, we compared the p-values for all effects considered as candidates for selection (according to the above rules) at the first step of forward selection, and included into the final multivariable model the TD and/or NL effects of a single predictor, corresponding to the lowest of the relevant p-values. At the second and subsequent steps of forward selection, the same models were fit for each variable without any TD/NL effects yet included in the final model, adjusting for all TD and NL effects of other predictors, selected at the earlier steps, as well as for the linear constant-in-time effects of other variables. The selection process stopped when all of the remaining TD and NL effects were statistically non-significant (p-value > 0.05).

3.3 Results: Comparison of the Alternative Estimates of Predictor Effects

During a median follow-up of 8.6 months, 211 (78.4%) patients died. Table 1 summarizes the results of multivariable analyses based on classic PH, AFT and AH models, in which all estimated effects were *a priori* constrained to be constant-over-time and linear, corresponding to, respectively, models (1), (4) and (6) in Sect. 2. (An asterisk indicates estimates statistically significant at $\alpha = 0.05$). Results are quite consistent across the three different models. All three models indicate statistically significant: (i) increases in mortality for patients with higher cancer stage and for smokers, and (ii) decreases for those who received double chemotherapy. For sex, whereas point estimates seem to vary from decreased (PH and AH models) to increased (AFT model) hazard for women, in all three models the effects are non-significant, so these apparent discrepancies reflect

mostly sampling errors. For continuous variables, all three models reveal statistically significantly higher mortality associated with higher values of $\log_2$ CRP, $\log_2$ LDH and ANC, and with lower lymphocyte counts, and a lack of a significant association with age (Table 1). The only, minor, discrepancy between the three models is that while higher albumin is always associated with lower mortality, this relationship is statistically significant only in the AFT model (Table 1).

Table 1. Estimates and confidence intervals (CIs) for classic constant/linear effects of PH, AFT and AH models.

	PH model, Hazard ratio (CI)	AFT model, Time ratios (CI)	AH model, Hazard differences (CI)
Stage (higher vs. lower)	1.87 (1.31, 2.67)*	1.67 (1.04, 2.34)*	9.56e-04 (3.07e-04, 1.60e-03)*
Smoking (ever vs. never)	2.01 (1.29, 3.12)*	1.75 (1.23, 2.59)*	1.11e-03 (1.11e-04, 1.81e-03)*
Chemotherapy (double-vs. single-agent)	0.62 (0.43, 0.89)*	0.69 (0.57, 0.95)*	–1.21e-03 (–2.21e-03, -2.12e-04)*
Sex (female vs. male)	0.99 (0.74, 1.33)	1.02 (0.90, 1.11)	–1.23e-04 (–8.46e-04, 6.00e-04)
$\log_2$ CRP	1.30 (1.10, 1.53)*	1.23 (1.00, 1.55)*	6.15e-04 (2.27e-04, 1.00e-03)*
Albumin	0.93 (0.78, 1.10)	0.94 (0.89, 0.96)*	–2.35e-04 (–6.25e-04, 1.55e-04)
$\log_2$ LDH	1.64 (1.43, 1.88)*	1.45 (1.16, 1.78)*	1.51e-03 (8.75e-04, 2.15e-03)*
ANC	1.32 (1.13, 1.54)*	1.21 (1.04, 1.46)*	7.52e-04 (1.66e-04, 1.34e-03)*
Lymphocytes	0.84 (0.72, 0.98)*	0.87 (0.76, 0.96)*	–2.76e-04 (–5.46e-04, -5.52e-06)*
Age	1.02 (0.89, 1.18)	1.03 (-0.00, 0.00)	4.37e-05 (–2.50e-04, 3.38e-04)

* Indicates a statistically significant effect with two-tailed tests with $\alpha = 0.05$.

Table 2 compares the results of the flexible extensions of the PH, AFT and AH models, built using the forward selection of TD and/or NL effects described in Sect. 3.3. Specifically, in Table 2 a symbol "NL", "TD" or "TD + NL" indicates if the final flexible multivariable model (identified in the heading of the corresponding column) included, respectively, the non-linear, time-dependent or both effects for the predictor in the corresponding row. (See Figs. 1 and 2 for graphs of the corresponding estimated TD/NL effects.) For predictors for which neither the TD nor the NL effect was selected, Table 2 shows the point estimate of the adjusted parametric effect, corresponding to a constant hazard ratio, time ratio or hazard difference, respectively. Generally, these

parametric estimates in Table 2 are very similar to the corresponding estimates in Table 1, indicating that for these variables adjusting for TD and/or NL effects of other predictors has not materially affected the results. For all binary variables, regardless of the model, no TD effects were identified (top 4 variables of Table 2), providing consistent evidence that their effects are constant across the follow-up. In contrast, for each of the six continuous variables, at least in one of the three models the TD and/or NL effect(s) were selected (Table 2). Furthermore, except for the TD effect of albumin that was significant in all three models, for the other predictors different effects were selected in different models. Generally, the results of AFT and AH analyses were similar regarding the TD effects, selected in both models for several predictors, but NL effects were not selected at all in AFT models (Table 2). Interestingly, in the AH framework for four predictors both the TD and NL effects were statistically significant.

Table 2. Estimates from the final flexible extensions of PH, AFT and AH models.

	PH model, Hazard ratio	AFT model, Time ratio	AH model, Hazard difference
Stage (higher vs. lower)	1.85	1.69	1.19e-03
Smoking (ever vs. never)	2.21	1.72	1.38e-03
Chemotherapy (double- vs. single-agent)	0.59	0.69	-1.24e-03
Sex (female vs. male)	1.04	1.06	1.23e-04
$\log_2$ CRP	TD + NL	TD	6.25e-04
Albumin	TD	TD	TD + NL
$\log_2$ LDH	1.62	TD	TD + NL
ANC	1.28	TD	TD + NL
Lymphocytes	0.83	TD	-2.51e-04
Age	1.00	1.04	TD + NL

TD/NL indicates that this effect was selected.

Figure 1 shows the TD estimates that were selected into respective final flexible models as shown in Table 2. For each predictor, the TD curve shows how the strength of its either NL or linear relationship with the hazard varies with increasing follow-up time, depending on whether the NL effect of the same predictor was selected or not in the same model (see Table 2). For flexible AH analyses, to produce the TD plots, representing $\widehat{\alpha}_j(t)$ in Eqs. (7) and (9), we first performed a piece-wise linear regression of $\widehat{A}_j(t)$ against time using the `segmented` function of the eponymous R package. Then, in Fig. 1, the slope of these linear segments was plotted to approximate $\widehat{\alpha}_j(t)$. Thus, the resulting TD estimates are unsmooth but still illustrate how the strength of the effect varies over follow-up. The decreasing curve in Fig. 1b shows that the linear effect of $\log_2$ CRP in the AFT model becomes gradually weaker as time since NSCLC diagnosis increases but at any time higher CRP is associated with higher mortality, because the values are always > 0. Figure 1a shows that the NL effect of $\log_2$ CRP in the flexible

PH model (shown in Fig. 2a) also decreases over time. Thus, both models indicate that higher CRP at diagnosis is a strong predictor of early mortality, in the next 6 months, but not for later mortality. In contrast, Fig. 1f shows an *increasing* curve for TD effect of $\log_2$ LDH in the flexible AFT model, indicating that higher LDH at diagnosis is an important predictor of later mortality, after the first year of follow-up.

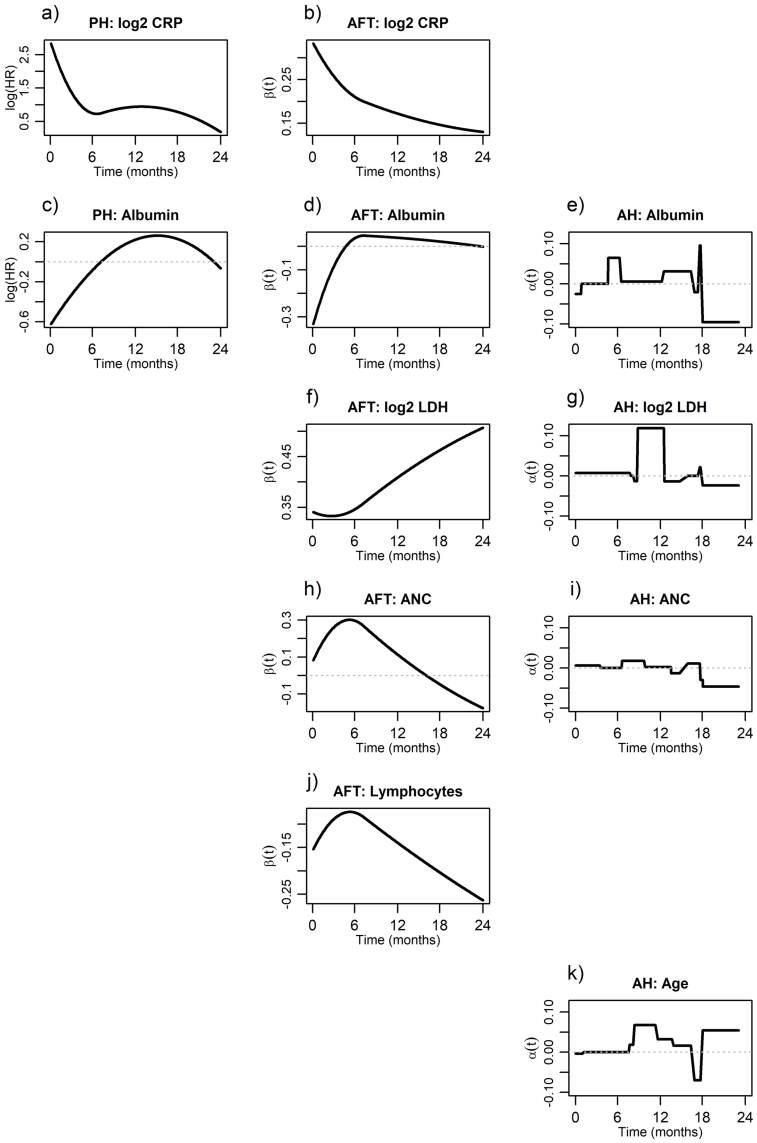

Fig. 1. Time-dependent (TD) estimates in the final PH, AFT, and AH models.

Figure 2 shows the estimated dose-response curves for those predictors whose NL effects were selected in the flexible extensions of the PH or AH models (as indicated by "NL" in the respective cell of Table 2). Figure 2a suggests a complex, possibly non-monotone effect of $\log_2$ CRP, normalized to z-scores, in the flexible extension of the PH model. However, the non-linearity is limited to both tails of the $\log_2$ CRP distribution, where there are only a few observations, whereas the relationship is close to linear for the middle range of values that do not deviate by more than one standard deviation (SD) from the mean. In flexible AH analyses, the NL effect of age indicates that mortality increases strongly as age increases from 3 to 2 SDs below the mean but not beyond (Fig. 2e), and if albumin at diagnosis is at least 2 SDs below the mean (Fig. 2b).

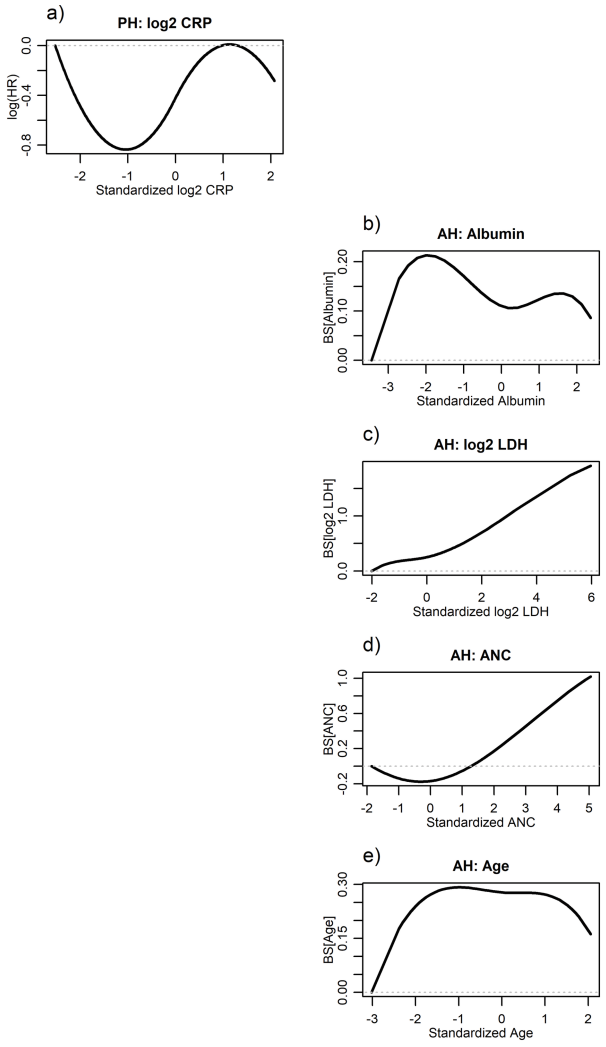

Fig. 2. Non-linear (NL) estimates in the final flexible extensions of PH and AH models

4 Discussion

Recent parallel advances in statistical methodology and in computational resources allow us nowadays to estimate complex, flexible multivariable models for survival analyses. In this paper, we outline the range of alternative models that can be estimated in real-world analyses. Then, using a case study of predictors of mortality in advanced lung cancer, we provide empirical evidence of both (a) the practical importance of accounting for time-dependent and/or non-linear effects, and (b) the possible dependence of the final estimates and conclusions on whether the analyses rely on proportional hazards (PH), accelerated failure time (AFT) or additive hazards (AH) models. However, because (i) PH model is estimated using partial MLE while AFT model requires full MLE, and (ii) publicly available software for AH does not produce the model's deviance, it is difficult to assess which model fits best in a given real-world study. Future research should develop common estimation frameworks for all three models and use extensive simulations to compare their performance across a range of plausible scenarios.

Acknowledgments. Funding by the Canadian Institutes of Health Research (grant PJT-180634).

Disclosure of Interests. The authors have no competing interests to declare. Michal Abrahamowicz is a Distinguished James McGill Professor at McGill University.

Appendix

As described in Sect. 2.2, this R code demonstrates how to test for non-linear (NL) effects in an additive hazards (AH) model using the `aalen` function from the timereg package. It first fits for a covariate X_j both a linear term and a cubic B-spline $BS[X_j]$ (omitting one basis to avoid collinearity), to create a "deviation" from linearity. The corresponding p-value for that spline-based variable, $BS[X_j]$, indicates whether there is a significant NL effect, which in this example below ($p = 0.00018$) confirms a NL effect of "age". A similar approach can be used for time-dependent effects by omitting the `const` function in the model. Note that this simple code uses a model with a single predictor, whereas the analyses in Sect. 3 used multivariable models.

```
## EXAMPLE CODE TO CONSTRUCT BS[] TO TEST FOR NL EFFECT OF AGE ##

library(survival)
library(splines)
library(timereg)
data(sTRACE)

AH.bs_AGE <- aalen(Surv(time, status==9) ~
                     const(bs(age, degree=3,
                              knots=c(median(age)))[,1:3]) +
                     const(age), data=sTRACE)
sTRACE$NL.AGE <- as.vector(bs(sTRACE$age, degree=3,
                              knots=c(median(sTRACE$age)))[,1:3] %*%
                             coef(AH.bs_AGE)[1:3,1])

set.seed(12305202)
AH.bs_AGE_test <- aalen(Surv(time, status==9) ~ const(NL.AGE) +
                          const(age), data=sTRACE)
AH.bs_AGE_test
```

References

1. Andersen, P.K., et al.: Analysis of time-to-event for observational studies: guidance to the use of intensity models. Stat. Med. **40**(1), 185–211 (2021)
2. Cox, D.R.: Regression models and life tables (with discussion). J. Roy. Stat. Soc. [Ser. B] **34**(2), 187–220 (1972)
3. Hernán, M.A.: The hazards of hazard ratios. Epidemiology **21**(1), 13–15 (2010)
4. Gail, M.H., Wiesand, S., Piantadosi, S.: Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika **71**(3), 431–444 (1984)
5. Aalen, O.O.: A linear regression model for the analysis of life times. Stat. Med. **8**(8), 907–925 (1989)
6. Aalen, O.O.: A model for nonparametric regression analysis of counting processes. In: Proceedings, Sixth International Conference: Mathematical Statistics and Probability Theory; Wisła (Poland): Springer New York, NY (1980)
7. Cox, D.R., Oakes, D.: Analysis of Survival Data. Chapman and Hall/CRC, New York (1984)
8. McKeague, I.W., Sasieni, P.D.: A partly parametric additive risk model. Biometrika **81**(3), 501–514 (1994)
9. Etezadi-Amoli, J., Ciampi, A.: Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. Biometrics **43**, 181–192 (1987)
10. Abrahamowicz, M., MacKenzie, T., Esdaile, J.M.: Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. J. Am. Stat. Assoc. **91**(436), 1432–1439 (1996)
11. Crowther, M.J., Royston, P., Clements, M.: A flexible parametric accelerated failure time model and the extension to time-dependent acceleration factors. Biostatistics (2022)
12. Pang, M., Platt, R.W., Schuster, T., Abrahamowicz, M.: Flexible extension of the accelerated failure time model to account for nonlinear and time-dependent effects of covariates on the hazard. Stat. Methods Med. Res. **30**(11), 2526–2542 (2021)

13. Abrahamowicz, M., MacKenzie, T.: Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. Stat. Med. **26**(2), 392–408 (2007)
14. Wynant, W., Abrahamowicz, M.: Impact of the model-building strategy on inference about nonlinear and time-dependent covariate effects in survival analysis. Stat. Med. **33**(19), 3318–3337 (2014)
15. Wynant, W., Abrahamowicz, M.: Flexible estimation of survival curves conditional on non-linear and time-dependent predictor effects. Stat. Med. **35**(4), 553–565 (2016)
16. Wynant, W., Abrahamowicz, M.: Validation of the alternating conditional estimation algorithm for estimation of flexible extensions of Cox's proportional hazards model with non-linear constraints on the parameters. Biom. J. **58**, 1445–1464 (2016)
17. Wei, L.J.: The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. Stat. Med. **11**(14–15), 1871–1879 (1992)
18. Pang, M., Platt, R.W., Schuster, T., Abrahamowicz, M.: Spline-based accelerated failure time model. Stat. Med. **40**(2), 481–497 (2021)
19. Komárek, A., Lesaffre, E., Hilton, J.F.: Accelerated failure time model for arbitrarily censored data with smoothed error distribution. J. Comput. Graph. Stat. **14**, 726–745 (2005)
20. Cox, D.R.: Analysis of Survival Data. Chapman and Hall/CRC (2018)
21. Lin, D.Y., Ying, Z.: Semiparametric analysis of the additive risk model. Biometrika **81**(1), 61–71 (1994)
22. Lu, C., Goeman, J., Putter, H.: Maximum likelihood estimation in the additive hazards model. Biometrics **79**(3), 1646–1656 (2023)
23. Martinussen, T., Scheike, T.: Dynamic Regression Models for Survival Data. Springer, New York (2006)
24. Therneau, T.M.: A Package for Survival Analysis in R. R package, version 3.8–3 (2024). https://CRAN.R-project.org/package=survival
25. Sung, H., et al.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J. Clin. **71**(3), 209–249 (2021)
26. Govindan, R., Page, N., Morgensztern, D., Read, W., Tierney, R., Vlahiotis, A., et al.: Changing epidemiology of small-cell lung cancer in the united states over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. J. Clin. Oncol. **24**(28), 4539–4544 (2006)
27. Wao, H., Mhaskar, R., Kumar, A., Miladinovic, B., Djulbegovic, B.: Survival of patients with non-small cell lung cancer without treatment: a systematic review and meta-analysis. Syst. Rev. **2**(1), 10 (2013)
28. Goring, S., Varol, N., Waser, N., Popoff, E., Lozano-Ortega, G., Lee, A., et al.: Correlations between objective response rate and survival-based endpoints in first-line advanced non-small cell lung Cancer: a systematic review and meta-analysis. Lung Cancer **170**, 122–132 (2022)
29. Gagnon, B., Abrahamowicz, M., Xiao, Y., Beauchamp, M.E., MacDonald, N., Kasymjanova, G., et al.: Flexible modeling improves assessment of prognostic value of C-reactive protein in advanced non-small cell lung cancer. Br. J. Cancer **102**(7), 1113–1122 (2010)
30. Gagnon, B., Agulnik, J.S., Gioulbasanis, I., Kasymjanova, G., Morris, D., MacDonald, N.: Montreal prognostic score: estimating survival of patients with non-small cell lung cancer using clinical biomarkers. Br. J. Cancer **109**(8), 2066–2071 (2013)
31. Chandra, R.K.: Nutrition and immunology: from the clinic to cellular biology and back again. Proc. Nutr. Soc. **58**(3), 681–683 (1999)
32. Balkwill, F., Mantovani, A.: Inflammation and cancer: back to Virchow? Lancet **357**(9255), 539–545 (2001)

33. Semenza, G.L.: Oxygen sensing, homeostasis, and disease. N. Engl. J. Med. **365**(6), 537–547 (2011)
34. Guthrie, G.J.K., Charles, K.A., Roxburgh, C.S.D., Horgan, P.G., McMillan, D.C., Clarke, S.J.: The systemic inflammation-based neutrophil–lymphocyte ratio: experience in patients with cancer. Crit. Rev. Oncol. Hematol. **88**(1), 218–230 (2013)