Published in final edited form as:

J Stat Theory Pract. 2013 April 5; 7(2): 421–441. doi:10.1080/15598608.2013.772033.

Correcting the Results of the Wrong Model: Treatment Effects under Early Detection of Cancer

Shih-Yuan Lee¹ and Alex Tsodikov²

¹Millennium Pharmaceuticals, 35 Landsdowne Street, Cambridge, MA 02139.
Connie.Lee@MPI.com

Abstract

Early detection of cancer leads to variability of the point of diagnosis advanced by the amount of the so-called lead time, a random variable. Estimated treatment effects by the proportional hazards (PH) model may be biased if this variability is ignored. We study how true and PH-estimated treatment effects differ in screened vs. unscreened populations and offer an approximate correction for the reported PH-based estimate that does not require raw data, targeting a meta-analysis-type application. We rely on a joint cancer incidence and survival model of prostate cancer to furnish key information for the correction. The procedure is applied to a series of prostate cancer data analyses using the PH models reported in the literature. Simulations are used for assessing the quality of the method and sensitivity analyses.

Keywords

Bias; Early detection; Lead time; Misspecified model; Proportional hazards

1 Introduction

Combined use of the early detection and treatment has become a major trend of cancer interventions. The treatment effect is usually assessed by the proportional hazards (PH) model fitted to cancer-specific survival time measured from the point of diagnosis. The model is usually used indiscriminately with respect to the screening tests performed prior to diagnosis. Assuming that the PH model is valid in the unscreened population, patient heterogeneity induced by early detection makes the model invalid under screening. Early detection advances cancer diagnosis by the amount of the lead time, a random variable. Moreover, survival during the lead time is virtually guaranteed since patients cannot die of cancer before it becomes symptomatic. Therefore, survival under screening is generally decomposed as a convolution of the lead time and the post-lead-time survival. We consider the PH model as misspecified under screening as a result of the additive random lead time effect (an artifact of advanced diagnosis) and an interaction between the lead-time and treatment expressing the (real) benefit of screening. Since the model under screening is non-proportional, the true hazard ratio characterizing the treatment effect in the screening era is time-dependent. When the misspecified PH model is fitted to data under screening it produces a biased hazard ratio estimate that represents the difference between treatments being studied, contaminated by the two effects associated with the lead-time.

²Corresponding author, University of Michigan, School of Public Health, Department of Biostatistics, 1415 Washington Heights, Ann Arbor, MI 48109, U.S.A. tsodikov@umich.edu.

Model misspecification induced by random effects received considerable attention in the literature. Kempthorne (1977) and Fienberg (1980) showed that collapsing a contingency table results in a biased estimator for a binary response outcome. Gail et al. (1984) further summarized asymptotic biases for common models in the exponential family GLMs when important covariates were omitted. They showed that the asymptotic bias is not zero unless the model is linear or exponential. Similar effects were observed with the PH models for failure time data. Lagakos and Schoenfeld (1984) demonstrated how ignored covariates result in a reduced power of the log-rank test; Schumacher et al. (1987) and Gail et al. (1984) showed a conservative bias in the proportional hazards setting. The general results concerning the behavior of the partial likelihood estimator and its asymptotic variance under a misspecified model were obtained in Struthers and Kalbfleisch (1986); Lin and Wei (1989). Struthers and Kalbfleisch (1986) studied misspecification induced by an omitted covariate in the PH model. Lin and Wei (1989) used the result to provide examples of robust hypotheses testing under some simplifying assumptions. Anderson and Fleming (1995) studied a setting where covariate effects were nonproportional and found that the direction of the bias depended on the nature of the covariate effect.

However, routine use of the PH model still dominates medical literature and clinical trials reports. A meta analysis of such studies would require a correction of the reported treatment differences without using raw data.

With this paper we are taking the general robust asymptotic results a step further and ask the question of whether the correct point estimator and its standard error for the parameter of interest can be restored from the reported results based on the wrong model without access to raw data in the spirit of meta analysis. Without raw data, the correction can only be done approximately under some assumptions and using additional modeling and prediction techniques. In this paper we use a small treatment effect approximation and the results from a model of prostate cancer incidence and mortality in the US population Tsodikov et al. (2006); Etzioni et al. (2008) to predict the lead time and assess the benefit of screening. To accomplish this we provide explicit expressions for the general robust asymptotics in terms of (conditional) moments of the covariate of interest, random variable (r.v.) Z , and using some natural simplifying assumptions. Using these forms, we then proceed to derive the small treatment effect approximations for use with the proposed meta-analytic correction. Facilitating the derivations is a concept of relative expectation and its properties developed in the Appendix A.2 and used as a back end of the presented results.

2 The model and assumptions

Let g , G , $\bar{F}$, $\bar{H}$ denote the density (pdf), the survival function (sf), the hazard function (hf), and the cumulative hazard function (chf) of the true model, respectively. We assume that the true model is a departure from a PH model explained by unobserved heterogeneity summarized by a mixing random variable (r.v.) W , possibly a vector. Random variables and their fixed values will be denoted by capital and script letters, respectively. Given w , the baseline survival distribution is represented by the pdf f , sf F , hf h , and chf H . The difference between treatments of interest is coded by a covariate, z , and the effect is modeled via an exponential predictor $\lambda(z) = \exp(\beta z)$, where β is a regression coefficient, possibly a vector. For example, z may represent a binary treatment assignment in a clinical trial, dose of the treatment agent, or generally any set of variables (a vector) characterizing the specific treatment of the disease. At the complete data level (given W) the model is

$$\lambda(t|z, w) = \theta(z) h(t|w), \quad G(t|z, w) = F(t|w)^{\theta(z)}, \quad g = \lambda G, \quad \Lambda = -\log G. \quad (2.1)$$

The mixed model is an average of the complete data model (2.1)

$$\bar{\lambda}(t|z) = E\{\lambda(t|z, W) | z, T > t\}, \quad \bar{G}(t|z) = E\{G(t|z, W) | z\}, \quad (2.2)$$

where expectations are taken over the conditional distribution of W , given $Z = z$, and $T > t$, where T is the survival time being modeled. In the sequel, for brevity, we will suppress the arguments t, z, W of the functions such as $\lambda(t|z, W)$ or $h(t|W)$, and assume that they are evaluated at the true θ_0 unless noted otherwise and explicitly.

The PH model that the above mixed model is based on is thought of as being valid in a population without screening, and the distributional characteristics of this basis model will be indexed by 0

$$\lambda_0(t|z) = \theta_0(z) h_0(t), \quad G_0(t|z) = F_0(t)^{\theta(z)}, \quad \theta_0(z) = e^{\beta_0 z}, \quad (2.3)$$

where the baseline distributions entering (2.1) are linked by some (at this point general) transformation

$$h(t|w) = \gamma(h_0, w, t), \quad (2.4)$$

where γ is some transformation (a functional). Of primary interest in the example will be a mixed model induced by the lead time. Lead-time measures how much a clinical diagnosis due to symptoms is advanced as a result of early detection by a screening test. With w playing the role of the lead time, we have the following specific form of the transformation

$$h(t|w) = \gamma(h_0, w, t) = h_0(t - w) 1_{\{t > w\}} e^{-\alpha w I_{CTx}}, \quad (2.5)$$

where $t - w$ expresses an artifactual benefit of screening (lead-time bias), I_{CTx} is an indicator of curative treatment strategy, and γ is the real benefit of screening associated with early diagnosis and a curative treatment strategy (a mixed interaction effect between screening and curative treatment). When patient's diagnosis is advanced by screening (SDx) by the amount of the lead time w , survival during the lead time is guaranteed as patients do not die before they develop symptoms. Hence their hf is zero during the lead time as specified by the $1_{\{t > w\}}$ term in (2.5). In the sequel we will omit $1_{\{t > w\}}$ for brevity assuming that $h(t|w) = 0, t \leq w$. The support of survival time in the presence of screening is affected by the missing variable w . In a population under screening not every patient is diagnosed by a screening test. Some patients are detected clinically when symptoms develop between tests (so-called interval cases). Others may not yet get screened at the point when the population is considered. Patients diagnosed clinically through symptoms rather than by the screening test by definition have $W = 0$ with probability 1. Hence the distribution of W has a mass at 0 (see Figure 1, top right). The distribution of the lead-time W and the treatment benefit γ are treated as known. The lead time distribution is estimated in Tsodikov et al. (2006) from cancer incidence data amassed in the Surveillance, Epidemiology, and End Results (SEER) cancer registry (www.seer.cancer.gov). Details of the model for the lead-time and its estimation from large population databases is reviewed in the Appendix A.1. Experiments with the mortality model of Etzioni et al. (2008) showed that a benefit $e^{-\alpha w}$ of up to 0.8 is needed to explain the mortality decline observed in prostate cancer since 1992. Hence 0.8 is used in key analyses, and a sensitivity analysis with respect to this parameter within the 0.8-1 range has shown that it has little effect on the correction for the relative risk (see Appendix A.3).

The meta-analytic correction for the reported PH hazard ratios will be based on a number of further assumptions. Random variables Z and W are considered independent since treatment

is assigned without knowledge of the lead-time W . Censoring is independent of W and the survival time T , given z . True treatment difference θ_0 is assumed to be small so that a first order Taylor approximation with respect to θ_0 can be used.

3 The true hazard ratio under the mixed model

Assuming that in the reference treatment group $\theta_0 = 1$, let $\theta_0(z)$ represent the hazard ratio in the correctly specified PH model (2.3) (the one without screening). In the mis-specified model (2.2) the hazard ratio is time dependent

$$\theta_t = \frac{\bar{\lambda}_{|\theta=\theta_0(z)}}{\bar{\lambda}_{|\theta=1}} = \theta_0 m(t|\theta_0), \quad \text{where} \quad m(t|\theta_0) = \frac{E\{hF^{\theta_0}|z\} E\{F\}}{E\{hF\} E\{F^{\theta_0}|z\}}, \quad (3.1)$$

and h, F are dependent on W and obtained by a transformation (2.4) of h_0 and W . The correction factor $m(t)$ that we call *true multiplier* characterizes the effect of the mixed model transformation on the hazard ratio.

Clearly, $m(t|1) = 1$ uniformly in t when there is no treatment effect ($\theta_0 = 1$). Also, $m(t|1) = 1$ as $t \rightarrow 0$, indicating that there is no bias at the start of follow up. When the baseline hazard h_0 , is an increasing function, $u(x) = h_0(t-x)1_{t>x}e^{-x}$ is decreasing in x while $v(x) = -H_0(t-x)e^{-x}$ is increasing. By Lemmas 5, 6 in the Appendix A.2 the logarithmic derivative is negative so $m(t)$ is a decreasing function of t . This implies that t is biased conservatively toward the null hypothesis $\theta_0 = 1$. Indeed since $m(t|1) = 1$, in the left neighborhood of $\theta_0 = 1$ we have $m(t) > 1$, and consequently $1 > t > \theta_0$ at least when θ_0 is close to 1. Alternatively, when $\theta_0 > 1$ (treatment is harmful) we still have attenuation towards the null as $\theta_0 > t > 1$. With small treatment effects when θ_0 is close to 1, logarithmic derivative of the multiplier at $\theta = 1$ determines the departure of the multiplier from 1. Indeed, expanding $m(t)$ around $\theta = 1$ we have $m(t|\theta) - 1 \approx \left[\frac{\partial \log m(t|\theta)}{\partial \theta} \right]_{\theta=1} (\theta - 1)$. This logarithmic derivative can be interpreted as proportional to a conditional covariance by Lemma 5 in the Appendix A.2. It is important to note that $u(x)$ cannot be made an increasing function of x , because $u(x) = 0$ for $x = t$ by definition, and h_0 is nonnegative. If, for the sake of argument, this were possible, then by Lemma 6 we would have optimistic (anti-conservative) bias with decreasing hazard. Generally, when h_0 is non-monotonic, the direction of the bias term defined by the multiplier, $m(t|1)$, depends on the shape of the hazard function h_0 , and the distribution of the lead time W . Averaging over W will weigh increasing and decreasing areas of h_0 against each other as they contribute to the opposite behavior patterns of m .

The treatment in a screened population may appear more efficient than it actually is, dependent on the time t when the effect is evaluated. Shown in Figure 1 is the multiplier behavior for a unimodal baseline hazard function typical of cure models for various values of θ_0 of the treatment effect and under the lead time distribution shown in the upper right corner of the figure.

4 Treatment effect estimate under a misspecified PH model

4.1 Assessing the bias of point estimates

When the effect is estimated by the misspecified PH model, the PH estimator is generally consistent for some hazard ratio that is neither the underlying hazard ratio θ_0 , nor the true hazard ratio θ_t . The estimators will depend on the duration of the study τ . We assume that τ is non-random and marks the right extreme of the time to censoring. Denote by $\hat{\theta}(\tau)$, $\hat{\beta}(\tau)$ the large sample limits of the estimators $\hat{\theta}, \hat{\beta}$ based on fitting the misspecified PH model.

Let $X = \min(T, C)$, where C is time to independent censoring measured from observed diagnosis (clinical by symptoms, CDx, or screening, SDx, whichever comes first), and $\delta = 1(X = T)$ is an indicator of failure. The data is represented as a sample of (X_i, δ_i, z_i) , $i = 1, \dots, n$, where z is a vector of treatment covariates. For an individual i define the counting processes $Y_i(t) = \mathbb{I}(X_i \geq t)$, $N_i(t) = \mathbb{I}(X_i \leq t, \delta_i = 1)$. Define a random variable $Z = Z(t)$ that is obtained by a length-bias transformation [notdef] of the r.v. Z as follows.

$$f_{\bar{Z}}(z|t, \beta) = \frac{f_Z(z) \mu(z)}{E\{\mu(Z)\}}, \quad \mu(z) = E\{\theta(\beta) GS|z\}, \quad (4.1)$$

where f_A is the pdf of r.v. A for any A , θ is the regression coefficient (an arbitrary argument), and G, S are the true survival functions for T and C , respectively (parameterized by β_0), and the expectation in μ is taken over W . The empirical version of the expectation $E\{Z\}$ is

$$\widehat{E}\{\bar{Z}|t, \beta\} = \frac{\sum_{j=1}^n z_j Y_j(t) \exp(\beta z_j)}{\sum_{j=1}^n Y_j(t) \exp(\beta z_j)}. \quad (4.2)$$

The score function for the partial likelihood PL (Cox, 1972) can be written as

$$U_n(\tau, \beta) = \frac{d \log PL(\tau, \beta)}{d\beta} = \sum_{i=1}^n \int_0^\tau \left\{ z_i - \widehat{E}\{\bar{Z}|t, \beta\} \right\} dN_i(t). \quad (4.3)$$

Struthers and Kalbfleisch (1986) showed that the large sample limit of the normalized score function (4.3) is

$$U^*(\tau, \beta|\beta_0) = \int_0^\tau E\{gS\} \left[E\{Z|T=t, C>t\} - E\{\bar{Z}|t, \beta\} \right] dt. \quad (4.4)$$

Note that obtaining explicit expressions for the conditional expectations such as $E\{Z|T=t, C>t\} = E\{ZgS\}/E\{gS\}$ using the length-bias transformation (cf. (4.1)) is an example of relative expectation reviewed in the Appendix A.2. The rule here is that gS corresponds to the probability in the condition. We will use this understanding below repeatedly in conditional expectations, variances, covariances and moments. With this understanding, under the PH model the true β_0 is the solution of (4.4)=0, because $\beta = h$, $g = G$, and h is non-random, which makes the difference in (4.4) zero. Under a misspecified PH model, the maximum partial likelihood estimator (MPLE), $\widehat{\beta}(\tau)$ solving (4.3)=0 will be consistent for $\beta^*(\tau, \beta_0)$, the solution of the score equation (4.4)=0. Generally, $\widehat{\beta}(\tau)$ is biased $\beta^*(\tau, \beta_0) - \beta_0$. It is easy to verify that $\widehat{\beta}$ is unbiased under the null hypothesis $\beta_0 = 0$. Indeed, in this case $\beta_0 = 1$, and $gS = hFS$ and $GS = FS$ are both independent of Z , which again makes the difference in (4.4) 0 (at $\beta_0 = 0$).

4.2 Small treatment effect approximation for point estimates

Similar to (3.1) we can represent the relationship between the large-sample limit of the PH-model-based estimate $\beta^*(\beta_0)$ and the true β_0 as

$$\beta^*(\beta_0) = \beta_0 \times m^*(\tau) + o(\beta_0), \quad (4.5)$$

where the *PH multiplier*, $m^*(\tau)$, describes the departure of β^* from β_0 ,

$$m^*(\tau) = 1 - \int_0^\tau E\{fS\} \text{Var}\{Z|C>t\} \left[E\{H|\tilde{T}=t, C>t\} - E\{H|\tilde{X}>t\} \right] dt \times \left[\int_0^\tau E\{fS\} \text{Var}\{Z|C>t\} dt \right]^{-1}, \quad (4.6)$$

where r.v. T be the survival time following the sf F , and $X = \min(T, C)$. The proof of the above result is given in Lemma 7 of the Appendix A.2. Here $\text{Var}(Z)$ is thought of as the covariance matrix if Z is a vector. Using Lemma 2 of the Appendix A.2 and (A.16) one can show that $E\{H|T=t, C>t\} - E\{H|X>t\} = \text{Cov}\{h, H|X>t\}$. This indicates that the behavior of m^* is also governed by the covariance of instantaneous and cumulative hazards relative to survival up to t (cf. Section 3). Note that if censoring does not depend on the treatment covariates Z , the condition $C>t$ and S will cancel from (4.6), and the bias will not depend on the censoring distribution. In terms of hazard ratio, we expect the limit treatment effect estimated by the PH model to be $\theta^*(\tau) \approx \theta_0^{m^*(\tau)}$. Given θ_0 , the relationship between true and large sample quantities $h_0(t)$, $\theta(t)$, θ_0 and $\theta^*(\tau)$ is displayed in Figure 2. The large-sample limit of PH estimated hazard ratio $\theta^*(\tau)$ depends on the distributions of the underlying survival from the clinical diagnosis, and the lead time (4.6).

4.3 Measuring the bias of variance estimated by the PH model

In the general setting of the M-estimation (Van der Vaart (2000), Chapter 5) the covariance matrix of the estimator $\hat{\beta}(\tau)$ based on a misspecified PH model is a sandwich

$$\Sigma^*(\tau, \beta_0) = \left[\frac{\partial^2 E\{\ell\}}{\partial \beta \partial \beta^T} \right]^{-1} E \left[\frac{\partial \ell_1}{\partial \beta} \frac{\partial \ell_1}{\partial \beta^T} \right] \left[\frac{\partial^2 E\{\ell\}}{\partial \beta \partial \beta^T} \right]^{-1}, \quad (4.7)$$

where $\ell = \log PL$, ℓ_1 is a contribution of one observation to the likelihood, and the expression is evaluated at $\beta = \theta^*(\tau, \beta_0)$, the limit in probability of PH-model based solution. The result for the misspecified PH model is derived in Lin and Wei (1989) in terms of expectations over empirical processes. We present an explicit expression in terms of moments of r.v. Z that lends itself naturally to small treatment effect approximations using the relative expectation techniques. Using the corollary to Lemma 3 of the Appendix A.2 we obtain the naïve information matrix of the misspecified PH model from (4.4) as

$$I = \frac{\partial U^*}{\partial \beta} = \int_0^\tau E\{gS\} \text{Var}\{\bar{Z}|t, \beta\} dt. \quad (4.8)$$

Proceeding similar to Section 4.1, we get the variance, Σ_1^* , of the individual score in the middle of the sandwich (4.7) as

$$\int_0^\tau E\{gS\} \left[\text{Var}\{Z|T=t, C>t\} + \left(E\{Z|T=t, C>t\} - \bar{Z}(\beta) \right)^2 \right] dt. \quad (4.9)$$

If the model is correctly specified, $\theta^*(\tau, \beta_0) = I^{-1}(\beta_0)$ because $I = \Sigma_1^*$. Indeed, similar to the discussion after the score equation (4.3), under the PH model or under the null hypothesis, the difference term in the right part of (4.9) is zero, and the conditional variance in (4.9) is equal to the variance term in (4.8). However, generally the correct variance of the misspecified model is given by $I^{-1} \Sigma_1^* I^{-1} \neq I^{-1}$.

4.4 Small treatment effect approximation for the variance

Now, consider the variance (4.7) under the small treatment effect assumption. Define

$$\begin{aligned}
I_0 &= \int_0^\tau E\{fS\} \text{Var}\{Z|C>t\} dt, \\
I_1 &= \int_0^\tau M_3\{Z|C>t\} E\{fS\} \left[1 - E\{H|\tilde{T}=t\}\right] dt, \\
I_2 &= \int_0^\tau E\{ZfS\} \text{Var}\{Z|C>t\} \left[1 - E\{H|\tilde{T}=t\}\right] dt, \\
I_3 &= \int_0^\tau M_3\{Z|C>t\} E\{fS\} \left[E\{H|\tilde{T}=t\} - E\{T|\tilde{T}>t\}\right] dt, \\
I_4 &= \int_0^\tau E\{fS\} \text{Var}\{Z|C>t\} \left[E\{H|\tilde{T}=t\} - E\{H|\tilde{T}>t\}\right] dt, \\
I_5 &= \int_0^\tau M_3\{Z|C>t\} E\{fS\} dt,
\end{aligned} \tag{4.10}$$

where M_k is a central moment of order k . By Lemma 8 in the Appendix A.2, the expansion for the variance takes the form

$$\Sigma^* = I_0^{-1} - I_0^{-1} [I_1 + I_2 + 2I_3 - 2I_5 I_4 I_0^{-1}] I_0^{-1} \beta_0 + o(\beta_0). \tag{4.11}$$

Note that for the PH model, a similar but simpler exercise gives the following variance approximation

$$\Sigma^* = I_0^{-1} - I_0^{-1} [I_1 + I_2] I_0^{-1} \beta_0 + o(\beta_0). \tag{4.12}$$

There are at least three scenarios when the PH variance is unbiased.

1. Naturally, if the true model is PH (no misspecification), then $I_3 = I_4 = 0$, and (4.11) and (4.12) are equivalent.
2. When Z is not skewed, i.e. when $M_3\{Z|C>t\} = 0$, then $I_1 = I_3 = I_5 = 0$, (4.11) and (4.12) are also equivalent and equal to $I_0^{-1} - I_0^{-1} I_2 I_0^{-1} \beta_0 + o(\beta_0)$.
3. When censoring is independent of the covariates,
 $I_5 I_4 I_0^{-1} = I_3 = M_3\{Z\} \int_0^\tau E\{fS\} \left[E\{H|\tilde{T}=t\} - E\{H|\tilde{T}>t\}\right] dt$, and the difference (4.11)-(4.12) = $o(0)$.

5 The meta-analytic correction

We have shown that estimators in the PH model are generally biased under misspecification. To correct the bias, we consider a meta-analytic framework to restore the adjusted hazard ratio β_0 from the results of reported studies without access to raw patient-level data. Consider a set of studies $i = 1, 2, \dots, k$ with sample sizes n_i that used a (misspecified) Cox model to obtain estimates of the treatment effect $\hat{\beta}_i$ and their standard errors $\hat{\sigma}_i / \sqrt{n_i}$. As we discussed in Section 4.1, we have the weak convergence result

$$\sqrt{n_i}(\hat{\beta}_i - \beta_i^*(\tau_i, \beta_0)) \xrightarrow{d} N(0, \sigma_i^2).$$

The fact that the large sample limit β_i^* varies with the study i is a consequence of model misspecification resulting in β_i^* being a functional of the distribution of the data rather than just a true parameter. In the approximate setting of this paper, this dependence on the data is absorbed by the multiplier $m_i^*(\tau_i)$ that is computed based on generally study-specific distribution of W , Z , and duration τ_i . Using the correction based on (4.5), we have, approximately,

$$\sqrt{n_i}(\hat{\beta}_{0i} - \beta_0) \xrightarrow{d} N(0, \sigma_i^{*2}),$$

where the corrected regression coefficient is obtained using the multiplier (4.6)

$$\widehat{\beta}_{0i} = \frac{\widehat{\beta}_i}{m_i^*(\tau_i)},$$

and its corrected σ_i^{*2} and its consistent estimator $\widehat{\sigma}_i^{*2}$ is generally obtained through (4.11). However, if the distribution of the treatment covariate is symmetric, the reported sample variance need not be corrected as discussed in the section 4.4.

Since the MPLE estimator from the PH model, $\widehat{\beta}_i$, is a consistent estimator of $\beta_i^*(\tau_i, \beta_0)$, by the Continuous Mapping Theorem, the corresponding $\widehat{\theta}_{0i}(\tau_i) = \exp(\widehat{\beta}_i z / m_i^*(\tau_i))$ is also a consistent estimator for θ_0 (subject to the validity of the approximation). By the Delta method, we will have

$$\sqrt{n_i}(\widehat{\theta}_{0i}(\tau) - \theta_0) \xrightarrow{d} N\left(0, \left[\frac{1}{m_i^*(\tau_i)} e^{\frac{\beta_i^*(\tau_i)}{m_i^*(\tau_i)}}\right]^2 \sigma_i^{*2}\right).$$

The meta-analytic estimate $\widehat{\beta}_0$ and $\widehat{\theta}_0$ is obtained as a weighted average of study-specific $\widehat{\beta}_{0i}$ with weights ω_i inversely proportional to the sample variance of the estimates, $\omega_i = n_i / \widehat{\sigma}_i^{*2}$. The combined point estimator $\widehat{\beta}_0$ and variance $\widehat{\sigma}^{*2}/n$ is

$$\widehat{\beta}_0 = \frac{\sum \omega_i \widehat{\beta}_{0i}}{\sum \omega_i}, \quad \frac{\widehat{\sigma}^2}{n} = \frac{1}{\sum \omega_i}. \quad (5.1)$$

6 Radical prostatectomy vs. active surveillance for prostate cancer

Radical prostatectomy is an invasive surgical procedure to remove the prostate gland. It is one of the most common treatments for patients with localized prostate cancer. Three studies, one from Europe (Bill-Axelson et al., 2005) and two from the United States (Tewari et al., 2006; Albertsen et al., 2007), have looked at large groups of localized prostate cancer patients. All patients in these three studies were newly diagnosed and many of them must have been subjected to the Prostate-Specific Antigen (PSA) tests, especially those diagnosed in the United State after 1990. Relative effects of radical prostatectomy vs. active surveillance on the disease-specific survival were estimated using PH models. Active surveillance is a deferred treatment strategy based on active monitoring that is used to trigger curative treatment at first signs of progression. Typically if the patient does progress under active surveillance, treatment is initiated much earlier than it otherwise would happen in the absence of screening. Assuming that the study cohorts are similar to patients in the Surveillance, Epidemiology, and End Results (SEER) registry data, and that the recruitment is following the pattern of cancer incidence in the SEER population, we recovered the underlying treatment effects θ_0 for radical prostatectomy vs. watchful waiting using the methods described in Section 5. Distribution of the lead time was obtained from the marginal incidence model (Tsodikov et al., 2006) fitted to SEER data, and assumed known. The baseline hazard was obtained from a PH analysis of SEER data before the year of 1988 when PSA was introduced, and assumed known for prediction of the multiplier. The lead time and survival distributions depend on age, A , and year of diagnosis, Y . Therefore we incorporated age and calendar time as covariates into the lead time and into the baseline

survival characteristics essentially by tabulating their distributions for all combinations of A and Y . We then exploit the fact that we kept the development general with respect to how $h(t|W)$ depends on W . Therefore, we can redefine the vector W to include A, Y in addition to the lead time. Essentially this means that all expectations over W turn into expectations over W, A, Y , and the key expressions of the paper are valid with this understanding. These expectations were taken with respect to the lead time conditional on A, Y , and then over the empirical distribution of A, Y specified using SEER data and the calendar period of the respective study. The results of the predicted true underlying treatment effects from all the three studies as well as the combined estimate using meta-analysis are summarized in Table 1. This analysis indicates under-estimated treatment effects by major clinical studies. However, the correction appears to be relatively modest, 6-26% relatively to the corrected one. Note that the corrected hazard ratio refers to the survival time from clinical diagnosis (the point of diagnosis by symptoms), for which there is no direct data if the patient is screen detected while asymptomatic.

A simulation study in the Appendix A.3 looked at the accuracy of the approximate correction underlying the above meta-analysis. We found that accuracy improves when the treatment effect gets smaller, when the study duration gets shorter, or when the hazard function is unimodal that is the case with cure models. The fact that prostate cancer is characterized by a large cure fraction lends credibility to the corrected estimates presented in our prostate cancer example.

Using the mortality model with the main effect of treatment Etzioni et al. (2008) we found that the decline in mortality remained not fully explained. Experimenting with the model we found that the benefit of screening of $e^- > 0.8$ is needed to explain the decline. In the Appendix A.3 we present results for $e^- = 0.8$ (realistic benefit of screening). Other screening by treatment interaction effects were studied in a sensitivity analysis (not shown). We found that while there was a substantial effect on the absolute risk (differences between hazards with and without screening), relative risk correction was virtually insensitive to the benefit within the specified range. This would certainly not be the case if treatment assignment were somehow differential by screening, in which case part of the screening benefit will be absorbed by the treatment effect.

7 Discussion

When screening is operating in the population targeted by clinical trials, the PH model is misspecified due to the fact that the time of diagnosis is advanced by a random and unobserved amount, the lead time, as patients are detected before they develop symptoms of the disease. This type of misspecification is special because patients do not die before they develop symptoms (during the lead time) meaning that the support of the survival time is random.

We have shown that even with proper randomization, evaluating the treatment effect in clinical trials recruiting from a population under screening could result in biased estimates. The bias is a function of the duration of the study, the shape of baseline hazard, the lead time distribution in the population under screening, the distribution of time to censoring if it varies by treatment group, and the size of the true treatment effect.

To correct the treatment effect estimated by the misspecified PH model, we propose a meta-analytic framework based on a study of bias and variance of a general misspecified PH model. To be able to provide a correction in the absence of raw data we entertained a small treatment effect approximation that simplified the formulas and reduced the complexity of

information needed to launch the correction. Simulation results suggest the correction we proposed is robust and accurate in realistic scenarios of modest treatment effects.

We also found that the PH model is correct at estimating the variance if the distribution of treatment covariates is not skewed, or if censoring does not depend on the treatment covariate. At the same time under no circumstances other than absence of screening can the PH model point estimate of the treatment effect be unbiased.

This study also shows that a correction of the treatment effects observed in the screening era is possible using population models of cancer that provide an estimate of the distribution of the lead time. This distribution was assumed to be known throughout the paper. This is a natural assumption since population models are fitted to big populations while the survival study of this paper concerns much smaller groups of cancer patients only.

A number of simplifying assumptions were unavoidable. We disregard stage progression between screening and clinical diagnosis in a small fraction of patients. While we incorporated the interaction between treatment and early detection resulting in the perceived benefit of screening to the population, we assumed that this interaction effect does not vary by the type of curative treatment. Last but not least, as we focused on the lead time bias, the length bias for screen-detected cancers and anti-length bias for cancers missed by screening were ignored. The latter can be recognized from (2.5) that implies, for example, that $h(t|W=0)$ applies equally to a patient from an unscreened population or to a patient from a population under screening who is detected due to symptoms with zero lead time. Length bias exhibits itself in that screening is more likely to catch slower growing tumors that spend longer times without symptoms. As a result tumors detected by screening have better prognosis than tumors missed by screening. For this reason a patient with $W=0$ from an unscreened population generally has better prognosis than a screened patient with $W=0$ whose cancer was missed by screening. This effect can be incorporated by allowing more complicated forms of $h(t|W)$ such as a transformation of $h_0(t-W)$ incorporating an additional predictor (W) modeling the dependence on the lead time. The same recipe could be used to generally model baseline survival post clinical diagnosis as being correlated with the lead time.

Because of the approximate nature of the correction proposed in this paper, the method does not always work. A detailed simulation study presented in the Appendix A.3 showed that the method tends to over-correct leading to over-optimistic assessment of the treatment effect when the treatment effects are not small. This phenomenon tends to be less pronounced in models with unimodal hazard functions such as cure models. At the same time, we found in our example that the misspecified Cox model generally leads to conservative estimates of the treatment effect. Therefore an average between naïve and the corrected estimate could be reported when the treatment effect is large. We also found that the accuracy tends to decrease with increasing duration of the study. The proposed method becomes inapplicable if the studies being combined are believed to be non-representative of the reference population used to obtain the distribution of the mixing variable W .

Software in R is available on request from the authors. More technical details and simulations are available in the Shih-Yuan Lee's dissertation thesis at the Department of Biostatistics, University of Michigan.

Acknowledgments

This work was supported by NIH/NCI grants U01 CA157224 and T32 CA083654.

A Appendix

A.1 Lead-time

Let $\lambda(a|x)$ be the hazard function (incidence rate) of cancer diagnosis at age, $A = a$, for a person born in year b . Cancer development passes through the disease-free state and the pre-clinical state before being detected or censored without a diagnosis. The subject's age at tumor onset A_O represents the duration of the disease-free stage. The duration of latent cancer growth in the pre-clinical stage is given by the delay time $D = A - A_O$, a backward recurrence time, represented by the period between cancer onset at the age of A_O and its diagnosis at the age of A . Given onset time, cancer diagnosis is a result of two competing risks, the time to detection by screening, SD_X , and the time to clinical diagnosis due to symptoms of the disease, CD_X , so that the delay time is $D = \min(SD_X, CD_X)$. The time CD_X is referred to as the sojourn time. Conditional CD_X given screen diagnosis, is stochastically larger than the unconditional one, due to the length bias. If the unconditional CD_X is exponential then the conditional will be larger by a factor of 2 by the lack of memory property and symmetry. Suppose $\mathcal{A} = \{a_1, a_2, \dots\}$ are random ages when the subject is screened (a screening schedule). We have used a two-stage model for the screening schedule point process $\mathcal{A}$. Let $\lambda_1(a, t)$ be the hazard of the first screening test for a man of age a in year t . Then the probability that a man born in year b will not be tested by the age of a is

$$G_{1S}(a|b) = \exp\left\{-\int_0^a \lambda_{1S}(\xi, b+\xi) d\xi\right\}. \quad (\text{A.1})$$

We assume that in men who already had their first screening test, secondary tests $\{a_2, a_3, \dots\}$ form a non-homogeneous Poisson process with intensity $\lambda_2(a, t)$. Both intensities of screening testing λ_{1S} and λ_{2S} are treated as known bivariate functions estimated by approximating the output of a random schedule generator developed by the National Cancer Institute (NCI) Mariotto et al. (2007).

The survival function $G_{SDX}(b, A_O)$ representing the probability of no screening diagnosis for a subject born in year b , with tumor onset at the age of A_O , and delay time since onset is derived in Tsodikov et al. (2006) as

$$G_{SDX}(\xi|b, A_O) = G_{1S}(A_O + \xi|b) + [1 - G_{1S}(A_O|b)] G_{2SDX}(\xi|b, A_O, A_O) + \int_0^\xi [1 - p(\zeta)] f_{1S}(A_O + \zeta|b) G_{2SDX}(\xi - \zeta|b, A_O + \zeta, A_O) d\zeta, \quad (\text{A.2})$$

where p is age-dependent screening sensitivity, f_{1S} is the pdf corresponding to λ_{1S} , and

$$G_{2SDX}(\xi|b, a, A_O) = \exp\left\{-\int_{\max(A_O - a, 0)}^\xi \lambda_{2S}(a + \zeta, b + a + \zeta) p(\zeta + a - A_O) d\zeta\right\}, \quad (\text{A.3})$$

and $\int_a^b = 0$ for any $b < a$. The above expressions are a result of averaging over and the Bernoulli outcomes of screening tests with probability of success p within the subject, given A_O . Using a conditional independence of the competing risks of cancer diagnosis by screening and clinically, given age at tumor onset A_O , we have

$$\lambda_l(a|b) = -\frac{d}{da} \log E\{G_{CDX}(a - A_O|b, A_O) G_{SDX}(a - A_O|b, A_O)\}, \quad (\text{A.4})$$

where G_{CDX} is the survival function of the sojourn time.

This model was fitted by maximizing a parametric likelihood for incidence rates.

The joint distribution (pdf) of the lead time $W = CD_x - SD_x$, a forward recurrence time, and age at diagnosis A is given by

$$f_{LT}(w, a|b) = \int_0^a f_O(y|b) f_{CD_x}(a - y + w|b, y) \begin{cases} G_{SD_x}(a - y|b, y), & w=0 \\ f_{SD_x}(a - y|b, y), & >0 \end{cases} dy,$$

where f_O is a pdf of the age at onset, and f_{CD_x} , f_{SD_x} are pdfs corresponding to sf G_{CD_x} , G_{SD_x} , respectively. Expression under the integral is a joint pdf of age at onset y , screen diagnosis at a , and potential clinical diagnosis at $a + w$.

Finally, the conditional lead-time distribution used in this paper is obtained from the joint one

$$f_{LT}(w|a, b) = \frac{f_{LT}(w, ab)}{f_I(a|b)}, \quad (\text{A.5})$$

where f_I is the pdf corresponding to the cancer incidence rate I . Shown in Figure 1 is a representative lead time distribution for a 65 year old patient diagnosed in 1995 predicted by the incidence model Tsodikov et al. (2006).

A.2 Proofs

For any random vector X with the joint pdf $f_X(x)$ and a function $\mu(X)$ define the random vector X_μ as a “length-biased” version of X following the joint pdf

$$f_\mu(x) = \frac{f_X(x) \mu(x)}{\int f_X(x) \mu(x) dx}. \quad (\text{A.6})$$

Many expressions in this paper can be concisely represented as characteristics of such length-biased variables. For example, for any function $v(X)$ we can write

$$E\{v(X_\mu)\} = \int v(x) \frac{f_X(x) \mu(x)}{\int f_X(\tilde{x}) \mu(\tilde{x}) d\tilde{x}} dx = \frac{E\{v\mu\}}{E\{\mu\}}. \quad (\text{A.7})$$

We call the expectation over a length-biased variable a relative expectation.

Suppose now that $\mu(X)$ is a probability of some event A , conditional on X , $\mu(X) = \Pr\{A|X\}$. Then, clearly (A.6) is a pdf of $\mu(X)$ conditional on A , and (A.7) is a conditional expectation

$$E\{v(X_\mu)\} = E\{v(X)|A\}. \quad (\text{A.8})$$

For arbitrary random variables U and V , define the “relative expectation” of U with respect to V (cf. Tsodikov (2003)) as

$$E(U \parallel V) = \frac{E(UV)}{E(V)}. \quad (\text{A.9})$$

When conditioning on a third random variable Z , we write

$$E(U \parallel V|Z) = \frac{E(UV|Z)}{E(V|Z)}. \quad (\text{A.10})$$

We will routinely use the connection

$$E\{\nu(X) \parallel \mu(X)\} = E\{\nu(X) | A\}, \quad \mu(X) = \Pr\{A|X\} \quad (\text{A.11})$$

to write relative expectations as conditional expectations and the other way around when convenient. For example, consider a derivation of the hazard function for a mixed model as a ratio of average pdf to average survival function

$$\lambda(t|z) = \frac{E\{\lambda(t|z, W) G(t|z, W)\}}{E\{G(t|z, W)\}}. \quad (\text{A.12})$$

Now note that the right part of (A.12) is $E\{G\}$, and that $G(t|z, W) = \Pr\{T > t|z, W\}$. Then according to (A.11) we may write

$$\lambda(t|z) = E\{\lambda \parallel G|z\} = E\{\lambda | T > t, z\}, \quad (\text{A.13})$$

as in the well known expression (2.2) in the Introduction. The meaning of the above expression is that subjects surviving up to $t > 0$ are generally healthier than at time zero (selection effect of survival of the fittest), and this biases the mixing variable $W|T > t > 0$ towards risk reduction with time t .

Relative variances, covariances and higher-order relative moments are defined straightforwardly by replacing usual expectations in the definition by the relative one, all under the same relative conditioning, ex. $\text{Var}(U \parallel V) = E(U^2 \parallel V) - E^2(U \parallel V)$.

For any non-random V , or when $U \parallel V$, we have $E(U \parallel V) = E(U)$. Under the PH model h and F in (A.13) are non-random, so $E\{h \parallel F|Z\} = h$ leading to the natural PH expression $\lambda = h$. Also, for a non-random a , $E(U \parallel aV) = E(U \parallel V)$.

Lemma 1 Let μ, ν be some functions of random variables independent of a parameter θ . Then

$$\frac{\partial}{\partial \theta} \log E\{\mu \parallel \nu^\theta\} = \frac{\text{Cov}\{\nu, \log \nu | \nu^\theta\}}{E\{\mu \parallel \nu^\theta\}}. \quad (\text{A.14})$$

Proof proceeds straightforwardly by differentiation and using the definition (A.9).

Lemma 2 Let μ, ν , and ξ be some functions of random variables. Define

$$E\{\mu \parallel_{\xi}^{\nu}\} = E\{\mu \parallel \nu\} - E\{\mu \parallel \xi\}. \quad (\text{A.15})$$

Then

$$E\{\mu \parallel_{\xi}^{\nu}\} = - \frac{\text{Cov}\{\mu, \xi \parallel \nu\}}{E\{\xi \parallel \nu\}}. \quad (\text{A.16})$$

Using the definition (A.9) we get the left part of (A.16) as a difference of two fractions. Bringing them to a common denominator and dividing numerator and denominator by $E(\nu)$, and again using (A.9), we arrive at the right part of (A.16).

Lemma 3 Let μ be some function of random variables, Z be some random variable, and $\nu = e^{Z}$. Define the central relative moment of Z of k th order as

$$M_k(Z \parallel \mu) = E \left\{ [Z - E\{Z \parallel \mu\}]^k \parallel \mu \right\}. \quad (\text{A.17})$$

Then if μ does not depend on

$$\frac{\partial^k}{\partial \beta^k} E\{Z \parallel \theta_\mu\} = M_{k+1}(Z \parallel \theta_\mu). \quad (\text{A.18})$$

The proof proceeds by induction, uses straightforward differentiation and the fact that $M_1(Z \parallel \mu) = 0$. As a corollary, we have

$$\frac{\partial}{\partial \beta} E\{Z \parallel \theta_\mu\} = \text{Var}(Z \parallel \theta_\mu).$$

Lemma 4 Let $\mu = \mu(\beta_0, Z, W)$, where Z, W are r.v.s, and β_0 is a parameter. Then

$$\frac{\partial}{\partial \beta_0} E\{Z \parallel \mu\} = \text{Cov} \left\{ Z, \frac{\partial \log(\mu)}{\partial \beta_0} \parallel \mu \right\}. \quad (\text{A.19})$$

Proof by straightforward differentiation.

Lemma 5 The logarithmic derivative of the multiplier is proportional to the relative covariance between the negative cumulative hazard $-H = \log F$ and the hf h .

$$\frac{\partial \log m(t|\theta)}{\partial \theta} = \frac{\text{Cov}\{-H, h|Z, T>t\}}{E\{h|Z, T>t\}}. \quad (\text{A.20})$$

The result follows from (A.14).

Lemma 6 A version of Chebyshev's inequality (See Shea (1979)). Let $u(x)$ and $v(x)$ be functions and W be a random variable such that $E\{u(W)\}$, $E\{v(W)\}$, and $E\{u(W)v(W)\}$ exist. Then

1. if u and v are both nonincreasing or both nondecreasing then $\text{Cov}(u(W)v(W)) \geq 0$;
2. if one of u and v is nonincreasing and the other nondecreasing then $\text{Cov}(u(W)v(W)) \leq 0$;
3. $\text{Cov}(u(W)v(W)) = 0$ if and only if at least one of u and v is a constant.

Lemma 7 Let $m^*(\beta_0)$ be the large-sample limit of the PH-model-based estimate of the regression coefficient β . Let r.v. T be the survival time following the sf F , and $X = \min(T, C)$. Then

$$\beta^*(\beta_0) = \beta_0 \times m^*(\tau) + o(\beta_0), \quad (\text{A.21})$$

where

$$m^*(\tau) = 1 - \int_0^\tau E\{fS\} \text{Var}\{Z \parallel S\} \{H\}_{fS}^{-1} dt \times \left[\int_0^\tau E\{fS\} \text{Var}\{Z \parallel S\} dt \right]^{-1}. \quad (\text{A.22})$$

Proof. By definition the function $m^*(\beta_0)$ is obtained by solving $U^*(\beta_0) = 0$ with respect to β_0 (dependence on τ is suppressed for brevity). Hence, taking derivative with respect to β_0 we have

$$\frac{\partial U^*(\beta|\beta_0)}{\partial \beta} \frac{d\beta^*(\beta_0)}{d\beta_0} + \frac{\partial U^*(\beta|\beta_0)}{\partial \beta_0} \Big|_{\beta=\beta^*(\beta_0)} = 0, \quad (\text{A.23})$$

uniformly in β_0 . Expanding $\beta^*(\beta_0)$ around $\beta_0 = 0$ and keeping in mind that $\beta^*(0) = 0$ we get

$$\beta^*(\beta_0) = \beta^{*'}(0) \beta_0 + o(\beta_0), \quad (\text{A.24})$$

where the prime stands for a partial derivative with respect to β_0 . The first order approximation is

$$\beta^*(\beta_0) \approx \beta_0 \times \beta^{*'}(0) \stackrel{\text{def}}{=} \beta_0 \times m^*(\tau). \quad (\text{A.25})$$

Combining (A.23) and (A.24) we obtain

$$m^* = - \frac{\partial U^*(\beta|\beta_0) / \partial \beta_0}{\partial U^*(\beta|\beta_0) / \partial \beta} \Big|_{\beta=\beta_0=0}. \quad (\text{A.26})$$

Taking derivatives in (A.26) using (4.4), Lemma 3, and under assumptions of Section 2, we arrive at (A.22).

Lemma 8 *Let*

$$\begin{aligned} I_0 &= \int_0^\tau E\{fS\} \text{Var}\{Z \| S\} dt, \\ I_1 &= \int_0^\tau M_3\{Z \| S\} E\{fS\} [1 - E\{H \| f\}] dt, \\ I_2 &= \int_0^\tau E\{ZfS\} \text{Var}\{Z \| S\} [1 - E\{H \| f\}] dt, \\ I_3 &= \int_0^\tau M_3\{Z \| S\} E\{fS\} E\{H \|_F^f\} dt, \\ I_4 &= \int_0^\tau E\{fS\} \text{Var}\{Z \| S\} E\{H \|_F^f\} dt, \\ I_5 &= \int_0^\tau M_3\{Z \| S\} E\{fS\} dt, \end{aligned} \quad (\text{A.27})$$

Then

$$\Sigma^* = I_0^{-1} - I_0^{-1} [I_1 + I_2 + 2I_3 - 2I_5 I_4^{-1}] I_0^{-1} \beta_0 + o(\beta_0). \quad (\text{A.28})$$

Proof. We have the expansions

$$\begin{aligned} E\{gS\} &= E\{fS\} + E\{ZfS\} [1 - E\{H \| f\}] \beta_0 + o(\beta_0) \\ \text{Var}\{Z \| gS\} &= \text{Var}\{Z \| S\} + [1 - E\{H \| f\}] M_3\{Z \| S\} \beta_0 + o(\beta_0) \\ E^2\{Z \|_{\theta(\beta)GS}^{gS}\} &= o(\beta_0) \\ \text{Var}\{Z \| \theta SG\} &= \text{Var}\{Z \| S\} + [m^* - E\{H \| f\}] M_3\{ZS\} \beta_0 + o(\beta_0), \end{aligned} \quad (\text{A.29})$$

where the multiplier m^* is given by (4.6), and M_k is a central moment of order k . Then using the expansions after some algebra we get (A.28).

A.3 Simulation study

We used simulations to study the accuracy of the PH multiplier under small treatment effect assumption. Sensitivity analyses were also conducted to evaluate how the shape of the baseline hazard and the magnitude of the treatment effect affect the estimated hazard ratios.

We adopted a 2×2 design to estimate the hazard ratio of the treatment group versus the control group using the PH model for both the unscreened and the screened populations. For the unscreened population, the survival time was calculated from the time of clinical diagnosis to time of death. For the screened population, a random lead time W was generated using the distribution displayed in Figure 1 and was added to the survival time in agreement with the specific convolution model (2.5). Under the PH assumption, a pre-specified treatment effect $\theta_0 = \exp(\alpha_0)$ was applied to treatment group characterized by the $z = 1$ value of the dummy variable. Survival times from the clinical diagnosis were drawn from Weibull distributions for increasing and decreasing baseline hazards, or from a cure model distribution for a changing baseline hazard. Survival functions of Weibull and cure model distributions were parameterized using median, m , and shape, s , parameters, and the treatment effect θ_0 , as defined in the following equation

$$\begin{aligned} G_{\text{weibull}}(t|\theta_0) &= \exp\left\{-\log(2)\theta_0\left(\frac{t}{m}\right)^s\right\} \\ G_{\text{cure}}(t|\theta_0) &= \exp\left\{-\theta_0\left(1 - \exp\left\{-\log(2)\left(\frac{t}{m}\right)^2\right\}\right)\right\}. \end{aligned} \quad (\text{A.30})$$

Hazard ratios were estimated using the PH model. Subjects with survival times longer than were censored. Adjusted hazard ratios were calculated using the PH multiplier defined in (4.6) and were compared with the true result. Two hypothetical treatment groups with the sample size of 500 each were used. Survival data in the treatment and control group was generated under no screening and under screening. Hazard ratios θ_0 (no screening) and θ_0^* (under screening) were estimated by the PH model applied to both sets of the data. To obtain an approximately unbiased estimate of θ_0 under a misspecified model, the log hazard ratio estimate under screening $\hat{\theta}(\tau)$ was adjusted by dividing it by the small-sample approximated multiplier $m^*(\tau)$ given by (4.6). The r.v. Z was taken to represent treatment assignment in a 1:1 ratio by simple randomization, a Bernoulli(0.5) r.v. Censoring was assumed to be independent of Z so S cancels from (4.6). Baseline distribution characteristics H and f were assumed to be known externally. The experiment was repeated for various study durations τ . Shown in Figure 3 are the results of a study to assess the quality of the proposed adjustment based on the small treatment effect approximation, and to assess its sensitivity to the screening benefit parameter α_0 . The left part of the figures shows three scenarios of varying shape of the baseline hazard function without screening vs. under screening (averaged over W). The right part of the figure shows hazard ratios estimated or predicted under different study durations represented by the x-axis. The wiggly polygon curves in the right part of the figure give estimated hazard ratios without screening (the bottom curve), and the one under screening showing a departure upwards. The smooth curves (a line at $y = \theta_0$ in the case of no screening) going through the polygons represent an average or a large sample limit of the respective PH estimate. In the case of screening, the latter is given by the exponentiated solution of (4.4)=0 with respect to θ_0 under the true θ_0, H, S . The top dashed curve shows hazard ratio under screening predicted as $\theta_0^{m^*(\tau)}$ with the multiplier $m^*(\tau)$ obtained from the approximation (4.6). The bottom dashed curve shows the corrected estimate. It is clear from Figure 3 that the bias correction using the approximation to the multiplier is reasonably accurate except in the case of large treatment effects (small α_0) for non-cure models with a monotonic hazard pattern. The simulation study just described was designed to maximize the departure of the approximation from the true values. A similar simulation study (not shown) close to the setting of the real data we used in the example showed better quality of approximation. Also we observe that the multiplier-corrected $\hat{\theta}_0$ is virtually unaltered under the two screening benefit scenarios.

We also conducted a simulation study to compare the difference between the corrected variance, $\hat{\theta}_0^{*2}$, and the variance reported by the PH model using the inverse of information

matrix. Having generated a large sample of 10000 we found that when the treatment is effective (*i.e.* $\theta_0 < 1$), the difference between the adjusted and naive variance is quite small ($< 1\%$) regardless of the treatment allocation ratio (table 2). This result is in agreement with the fact that the difference between the two was found to be $\alpha(\theta_0)$ when censoring does not depend on covariates that is the case in this simulation study. Empirical variances computed as sample variance estimator from a set of estimated θ s over 5000 replicates are in excellent agreement with the likelihood-based ones.

References

- Albertsen P, Hanley J, Penson D, Barrows G, Fine J. 13-year outcomes following treatment for clinically localized prostate cancer in a population based cohort. *The Journal of Urology*. 2007; 177(3):932–936. [PubMed: 17296379]
- Anderson G, Fleming T. Model misspecification in proportional hazards regression. *Biometrika*. 1995; 82:527–541.
- Bill-Axelsson A, Holmberg L, Ruutu M, Haggman M, Andersson S, Bratell S, Spangberg A, Busch C, Nordling S, Garmo H, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine*. 2005; 352(19):1977–1984. [PubMed: 15888698]
- Cox D. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972:187–220.
- Etzioni R, Tsodikov A, Mariotto A, Szabo A, Falcon S, Wegelin J, diTommaso D, Karnofski K, Gulati R, Penson D, Feuer E. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes and Control*. 2008; 19:175–181. [PubMed: 18027095]
- Fienberg, S. *The Analysis of Cross-Classified Categorical Data*. MIT Press; Cambridge, MA: 1980.
- Gail M, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984; 71(3):431–444.
- Kempthorne O. Why randomize. *Journal of Statistical Planning and Inference*. 1977; 1:1–25.
- Lagakos S, Schoenfeld D. Properties of proportional-hazards score tests under misspecified regression models. *Biometrics*. 1984; 40(4):1037–48. [PubMed: 6534407]
- Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*. 1989; 84:1074–1078.
- Mariotto A, Etzioni R, Krapcho M, Feuer E. Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer*. 2007; 109:1877–1886. [PubMed: 17372918]
- Schumacher M, Olschewski M, Schmoor C. The impact of heterogeneity on the comparison of survival times. *Statistics in Medicine*. 1987; 6(7):773–784. [PubMed: 3423500]
- Shea G. Monotone regression and covariance structure. *The Annals of Statistics*. 1979; 7:1121–1126.
- Struthers C, Kalbfleisch J. Misspecified proportional hazard models. *Biometrika*. 1986; 73(2):363–369.
- Tewari A, Raman J, Chang P, Rao S, Divine G, Menon M. Long-term survival probability in men with clinically localized prostate cancer treated either conservatively or with definitive treatment (radiotherapy or radical prostatectomy). *Urology*. 2006; 68(6):1268–1274. [PubMed: 17169648]
- Tsodikov A. Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*. 2003:759–774.
- Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. *Statistics in Medicine*. 2006; 25(16):2846–2866. [PubMed: 16397859]
- Van der Vaart, A. *Asymptotic Statistics*. Cambridge University Press; 2000.

Figure 1.

A comparison of the true treatment effect without screening $\theta_0 = e^0$ under the PH model (2.3) vs. the time-dependent true treatment effect θ_t under screening (3.1). Top left: Baseline hazard function h_0 . Top right: The distribution (pdf with mass at zero) of the lead time W . Bottom: The true multiplier, a ratio (3.1) of the true effect under screening averaged over the lead time to the effect without screening under various values of θ_0 .

Figure 2.

True and large-sample characteristics in screened and unscreened populations under various shapes of the baseline hazard. Left: The baseline hazard in the unscreened population $h_0(t)$ and its counterpart under screening $h_1(t) = E\{h(t|W)|T > t\}$ averaged over the lead time W as in (2.2). Right: True hazard ratio in the unscreened population θ_0 , its time-dependent counterpart under screening θ_t , and $\theta_0^{m^*(\tau)} \approx \theta_0^{m^*(\tau)}$ representing the large-sample limit of the PH-estimated hazard ratio under screening. τ is the duration of the study; t is a point in follow-up time. Note that the misspecified PH model produces a biased estimate $\hat{\theta}(\tau) \xrightarrow{P} \theta_0^{m^*(\tau)}$ of the average true hazard ratio θ_t under screening. Computation of m^* is done under the approximation.

Figure 3.

Quality of the small treatment effect approximation for the bias under screening benefit of $\epsilon^- = 0.8$. First two rows were selected to show when the approximation does not work well (monotonic hazard, large duration of the study, large treatment effect). Left: Shapes of the baseline hazard function without screening h_0 vs. under screening $h_1 = E\{h(t|W)|T > t\}$. Middle and right column: Variable polygons and smooth curves going through them correspond to PH model estimated and average hazard ratios at various study durations. The lines at $y = 1$ correspond to a simulation without screening ($W = 0$ with probability 1). The top dashed curve showing a departure upwards particularly with small θ_0 (large beneficial treatment effect) is the hazard ratio under screening predicted using the approximation (4.6) for the multiplier. The bottom dashed curve shows the PH-based estimate corrected using the approximate method of the paper.

Table 1

Estimated and predicted relative risks of radical prostatectomy vs. watchful waiting

Study	Reported RR(95%C.I.) from PH model	RR(95%C.I.) after correction
Bill-Axelsson et al. (2005, NEJM)	0.56 (0.36-0.88)	0.51 (0.24-0.77)
Tewari et al. (2006, J. Urology)	0.37 (0.25-0.55)	0.35 (0.21-0.50)
Albertsen et al. (2007, J. Urology)	0.29 (0.16-0.52)	0.23 (0.07-0.39)
Combined result from meta-analysis	0.41 (0.31-0.53)	0.36 (0.29-0.46)

Table 2

Variances ^{*} of θ_0 PH MLE based on a simulation study with duration =15

Proportion of the treatment group	0.3	0.5
	Simulation result ($\times 10^{-4}$)	Simulation result ($\times 10^{-4}$)
Increasing $h(t)$, Weibull $m=5$, $s=1.5$		
Empirical variance	7.40 $\times 10$	6.08 $\times 10$
Naïve variance	7.45	6.04
Sandwich variance	7.42	6.00
Decreasing $h(t)$, Weibull $m=5$, $s=0.9$		
Empirical variance	9.59 $\times 10$	7.62 $\times 10$
Naïve variance	9.62	7.49
Sandwich variance	9.61	7.48
Non-monotonic $h(t)$, Cure $m=3$, $s=2$		
Empirical variance	12.50 $\times 10$	9.69 $\times 10$
Naïve variance	12.52	9.65
Sandwich variance	12.55	9.68

* $\theta_0 = 0.5$ represents the treatment effect in a unscreened population. Naïve and Sandwich variances are based on a single large-sample run ($n=10000$). Empirical variance is based on 5000 replicates with sample size of $n=1000$, and should be considered one decimal place apart from the large sample run. Naïve variance is based on the inverse of profile information matrix in the PH model. Sandwich variance is the counterpart of the naïve corrected for misspecification by the sandwich (4.7). Empirical is a sample variance of the MLE of θ_0 obtained by fitting PH model to each replicate.