Time Scales in Epidemiological Analysis: An Empirical Comparison

Prabhakar Chalise[1], Eric Chicken[2] & Daniel McGee[2]

[1] Department of Biostatistics, University of Kansas Medical Center, Kansas, USA

[2] Department of Statistics, Florida State University, Florida, USA

Correspondence: Prabhakar Chalise, Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, 66160, USA. Tel: 1-913-945-7987. E-mail: pchalise@kumc.edu

Abstract

The Cox proportional hazards model is routinely used to analyze time-to-event data. This model requires the definition of a unique well-defined time scale. Most often, observation time is used as the time scale for both clinical and observational studies. Recently after a suggestion that it may be a more appropriate scale, chronological age has begun to appear as the time scale used in some reports. There appears to be no general consensus about which time scale is appropriate for any given analysis. It has been suggested that if the baseline hazard is exponential or if the age-at-entry is independent of covariates used in the model, then the two time scales provide similar results. In this report we provide an empirical examination of the results using the two different time scales using a large collection of data sets to examine the relationship between systolic blood pressure and coronary heart disease death. We demonstrate, in this empirical example that the two time-scales can sometimes lead to differing regression coefficient estimates but time-on-study model has better predictive ability in general.

Keywords: Baseline age, bootstrap, cumulative hazard, left-truncation, AUC

1. Introduction

Survival models are extensively used in epidemiological studies. There are various survival models available to analyze time to event data but the most frequently used model is the semi-parametric Cox proportional hazards (PH) model (Cox, 1972). The PH model is widely used with both experimental (clinical trials) and observational data and provides a semi-parametric method of analyzing the association between a set of risk factors and the time to occurrence of an outcome. The PH model makes no assumptions concerning the nature or shape of the underlying survival distribution, but assumes a parametric form for the effect of the predictors on the hazard. In many situations, we are most interested in estimates of the model parameters since if the model is correct they provide measures of the strength of the association between a characteristic and time to event. Most statistical software contains procedures for deriving estimates of the parameters of the model.

Current literature addressing the association between a characteristic and time to event based on the analysis of observational studies using the PH model contains a mixture of analyses using the two different time scales: time-on-study and chronological age. For example, there are two widely used sets of models to predict cardiovascular disease. The models that are widely used in the United States were derived from the Framingham Heart Study (Wilson et al., 1998) using time-on-study as the time scale, while the models that are widely used in Europe (Conroy et al., 2003) were developed using age as the time scale. Similarly, there has been long-term interest in estimating the effects of obesity on mortality and two papers have appeared addressing this question but using different time scales. In 1999, Allison et al. published their estimates of the number of deaths in the U.S. that are attributable to obesity. For their analysis, they used time-on-study as the time scale. In 2007, Flegal et al. published a similar analysis, based on slightly different data, but used chronological age as the time scale. A natural question is whether at least part of the differences between the results of these analyses are due to different time scales being selected for developing the proportional hazards models used.

An extreme example was provided by Cheung, Gao and Khoo (2003) who demonstrated that a PH model using different time scales could result in contradictory results that the models in which the parameters have opposite sign depending on the time scale. They examined women in the Surveillance, Epidemiology, and End Results (SEER) program diagnosed with Stage I breast cancer and demonstrated that when time-on-study was used as the time scale, a younger age at diagnosis was associated with a lower mortality. If chronological age was used as the time scale, the

opposite effect was found.

In 1997, Korn, Graubard and Midthune pointed out that the majority of published analyses based on the National Health and Nutrition Examination Survey I Epidemiological Follow up Study used time-on-study as the time scale in the PH model but suggested that chronological age might be a more appropriate time scale for some observational studies. Korn et al. (1997) also suggested two conditions for which the analysis results using the time-on-study time scale and the chronological age scale are equivalent. The first condition is that the baseline hazard λ_0 as a function of age is exponential, i.e., $\lambda_0(a) = c \exp(\psi a)$, where a is age and c and ψ are constants. The second condition is that the covariate z is independent of the baseline age a_0. Thiebaut and Benichou (2004) and Pencina, Larson and D'Agostino (2007) conducted simulation studies to investigate these two conditions and the choice of time scale. Thiebaut and Benichou (2004) reported that if the cumulative hazard function is exponential, the two time scale models yielded similar regression coefficients. On the other hand they found that the models could be significantly different even when the covariate of interest was independent of the baseline age. Pencina et al. (2007) found that when the correlation between the age-at-entry and the risk factor is zero, the biases from the two models using time-on-study time and chronological age time are very close. However, these two simulation studies are inconclusive about which time scale is the best.

The two scales vary in their choice of origin of the time scale. If we use chronological age as the time scale then the origin of the time scale is the date of birth. If time-on-study scale is used the origin is the date of diagnosis or date of randomization. The mathematical formulation of the Cox model is similar for both the time-on-study and the chronological age time scales but the implicit mechanism for estimation is different (Thiebaut and Benichou, 2004). In both of the formulations time is used only to order the times to event. This ordering defines the number of individuals at risk, the risk set at a particular time. Time scales that produce equivalent risk sets will produce equivalent results. The difference between the two scales is that using observation times, the risk sets are nested. That is, at any time t, the risk set, R_t is contained in the risk set R_{t^*} for any $t^* < t$. Using age does not necessarily result in this nesting of risk sets. This is because at a given age, some subjects may have already experienced failures while others may not be under study at that age and thus subjects keep on entering and exiting the study.

A couple of previous studies (Chalise, Chicken and McGee, 2012; Chalise, Chicken and McGee, 2013) have carried out extensive simulation to address the question of robustness using one time scale when the other is actually the correct one. They simulated two sets of data specifying time-on-study as the true time scale in one and chronological age as true time scale in other. Both types of models were fitted for the two sets of data and were assessed with respect to bias, mean square error and concordance index (measure of predictive discrimination). When time-on-study was correctly specified, the time on study models were better with respect to all three measures. But, when age was the correct time scale both time-scale models performed approximately equally well suggesting that the time-on-study models are more robust to misspecification of the underlying time scale. However, the studies did not provide real data example. It is extremely valuable to use real datasets to assess the models in empirical analysis. Therefore, the purpose of this article is to provide extensive empirical examples comparing the models with respect to regression coefficient estimates and measure of predictive discrimination.

2. Data and Methods

2.1 Data

We use a large collection of data sets from the Diverse Populations Collaboration (DPC) (McGee et al., 2005). DPC investigators pooled data from 27 observational studies to examine issues of variation in results in observational studies when differing methodologies and/or sampling methods are used for implementation and analysis. The studies can often be stratified by characteristics such as gender, area of residence (urban/rural) etc. When this stratification is accomplished there are 78 strata that we refer to as cohorts. In our analyses we excluded studies that did not have follow-up time records and that did not measure blood pressure. We additionally required that all of the cohorts included in our analysis have at least 50 Coronary Heart Disease (CHD) deaths during follow-up. Our analytic data consists of information form 25 studies stratified into 54 cohorts that contained 236,623 observations among which 14,156 deaths due to CHD occurred.

2.2 Statistical Methods

We restrict our attention to the proportional hazards model (PH) in this paper since this is the most widely used method in epidemiological studies. Also, we focus on a single covariate, systolic blood pressure (sbp) so that we can examine whether results are consistent using the different times scales and whether they are consistent with published relationships between *sbp* and *CHD* deaths as an event. The purpose of our empirical study is limited to comparing the two different time scale models but not to provide substantial analysis of the data.

The PH model assumes that the underlying hazard (rather than survival time) is a function of the independent variables (covariates) and the contributions of the covariates to the hazard are multiplicative. This model specifies that the hazard function associated with the covariate Z satisfies,

$$\lambda(t \mid \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{z}) \tag{1}$$

where $\lambda(t \mid \mathbf{z})$ is the hazard conditional on the value of the characteristics $\mathbf{z} = (z_1, z_2, ..., z_k)$, $\lambda_0(t)$ is the unspecified baseline hazard (the baseline hazard when all $z_i = 0$), t is time and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)$ is a vector of unknown constants, the parameters of interest. Estimates of the parameters and inferences about them are based on maximum partial likelihood (Cox 1972) and the asymptotic properties are justified using martingale and counting process theory (Anderson and Gill, 1982). In the PH model, time is used in ordering the events to determine the risk sets of subjects still being followed when each event occurs. It is not used directly in the estimation of the coefficients of the covariates. Therefore, the different time scales in PH models lead to different estimates only if they differ in the ordering of the times to event or censoring. Chronological age and time on study will in general produce different ordering of times to event and censoring. We mainly focus on the two models in this paper: (i) time-on-study as time scale including age at entry as a covariate and (ii) chronological age as time scale with left truncation on the entry age. This is because the previous simulation studies (Thiebaut and Benichou, 2004; Pencina et al., 2007; Chalise et al., 2012; & Chalise et al., 2013) have found very close results between these two models after extensively examining several possible models using the two time scale models.

Let a_0 be the age at which an individual enters the study, t be the length of time the individual is followed until he/she experiences an event of interest or terminates participation in the study, and a be the age of the individual at the point of event or censoring. Then the models are given as follows:

M1: Time-on-study as the time scale with age at entry included as a covariate:

$$\lambda(t \mid a_0, z) = \lambda_0(t) \exp(\beta z + \gamma a_0) \tag{2}$$

where β is the coefficient of z, the *sbp* and γ is the coefficient of a_0, the baseline age.

M2: Age as time scale with left truncation on the entry age:

$$\lambda(a \mid a_0, z) = \lambda_0(a \mid a_0) \exp(\beta z) \tag{3}$$

where $\lambda_0(a \mid a_0)$ is the baseline hazard conditional on entry age.

In addition, we also examine unadjusted age scale model to learn how this compares to the **M1** and **M2.**

M3: Age as time scale without adjustment:

$$\lambda(a \mid z) = \lambda_0(a) \exp(\beta z) \tag{4}$$

Using *sbp* as a risk factor we fitted the three Cox PH models 2, 3 and 4 for each of the 54 cohorts separately. The models were compared with respect to regression coefficient estimates and predictive discrimination as measured by time dependent area under the receiver operating curve (Hung and Chiang, 2010; Blanche, Dartigues and Jacqmin-Gadda, 2013) using R package *timeROC*.

2.2.1 Comparison of regression coefficients:

There are several possible methods for comparing the coefficients estimated from the fitted models. We used bootstrap methods to determine whether the coefficients for pairs of models differed significantly. Bootstrap samples with one thousand replications were used to calculate the standard error of the difference between two β's for each pair. The significant difference in coefficients from the two methods was calculated assuming normality of the differences:

$$\text{Test Statistics} = \frac{\widehat{\beta}_1 - \widehat{\beta}_2}{\text{Bootstrap } SE(\widehat{\beta}_1 - \widehat{\beta}_2)} \tag{5}$$

For each of 54 cohorts, we estimated the cumulative baseline hazard function as a function of *age* using Breslow method (Cox, 1972, with discussion). We plotted these estimates against *age* to determine the shape of the baseline hazard and then we plotted the log of the estimates against the *age* for the test of exponentiality. Two examples of such plots (one exponential and another non-exponential) are shown in supplementary **Figure S1**. We calculated correlation coefficients between *sbp* and age at entry to examine how strongly they were associated. The empirical correlation and whether the hazard function appeared to be exponential were used to determine if, at least approximately, the conditions from Korn et al. (1997) were met.

As previous studies and our comparison of regression coefficients suggested that the unadjusted age scale model performed poorly compared to age scale with left truncation on the entry age, we do not consider **M3** further to assess predictive discrimination.

2.2.2 Comparison of Predictive Discrimination

The area under receiver operating curve (ROC) as proposed by Hung and Chiang (2010) and Blanche et al. (2013) was used to compare the predictive discrimination of the two models **M1** and **M2**. The time dependent area under ROC curves (timeROC) were calculated for three different time points: 1 year, 3 years and 5 years. In order to avoid overfitting of the models, resampling based 5-fold cross validation method was used in order to calculate predicted area under the curves (AUC). The data is randomly divided into 5-sets: 4 sets for training ($4/5^{th}$ of data) and 1 set for validation ($1/5^{th}$ of data) sets. In the first iteration, first set of data is used as validation data and remaining as training data. **M1** and **M2** were fitted using the training data and relative risks were predicted for the validation data. For the estimated relative risks, time dependent areas under the curves were computed at 1, 3 and 5 years respectively for each of the models. In next iteration, second set of data is used as validation data and remaining as training data and so on up to 5 iterations until all data are used as training and validation set. Then, average AUCs over 5 iterations are computed for each time point for both models.

3. Results

The coefficient estimates (with their standard errors) obtained from **M1**, **M2** and **M3** are shown in appendix **Table A**. Each row of the table represents the values obtained from a separate cohort. β_1 and $se(\beta_1)$ are the coefficients and their standard errors from **M1**, β_2 and $se(\beta_2)$ are from **M2** and β_3 and $se(\beta_3)$ are from **M3**. The p-values comparing the models, two at a time, are given in columns 8, 9 and 10. The estimated correlation coefficients between *sbp* and *age at entry* are listed in the second to last column of the table. The final column contains an indicator of whether the survival distribution for that cohort appeared to be consistent with the exponential hazard function (i.e. 1 means the shape of the cumulative baseline hazard function looks exponential and 0 means that the shape of the cumulative baseline hazard does not look exponential).

Table 1. Number of cohorts for which the estimated regression coefficients are significantly different between models by correlation between the *sbp* and *age at entry*. (**M1**: Time-on-study Model, **M2**: Left truncated age model, **M3**: Unadjusted age scale model)

Correlation	M1 vs M2		M1 vs M3		M2 vs M3	
	Sig	**Non-sig**	**Sig**	**Non-sig**	**Sig**	**Non-sig**
0.5 +	0	9	6	3	9	0
0.4 - 0.5	0	7	4	3	6	1
0.3 - 0.4	1	15	12	4	16	0
0.2 -0.3	2	13	12	3	15	0
0.1 - 0.2	0	6	6	0	6	0
0.0 - 0.1	0	1	0	1	1	0
Total	3	51	40	14	53	1

In 26 cases the coefficients from **M1** are bigger than those from **M2** and in 27 cases they are smaller while in 1 case they are equal. In 53 cases, the coefficients obtained from the **M3** are smaller than those from **M1** and in all 54 cases the coefficients from **M3** are smaller than that from **M2**. **Figure 1** represents the pairwise relationship between the coefficients from the three models. It is evident from the plots that **M1** and **M2** are closer but are very different from **M3**. However, the standard errors were fairly equal in all models.

M1 = Time−on−study model, M2 = Left truncated age model, and M3 = Unadjusted age model
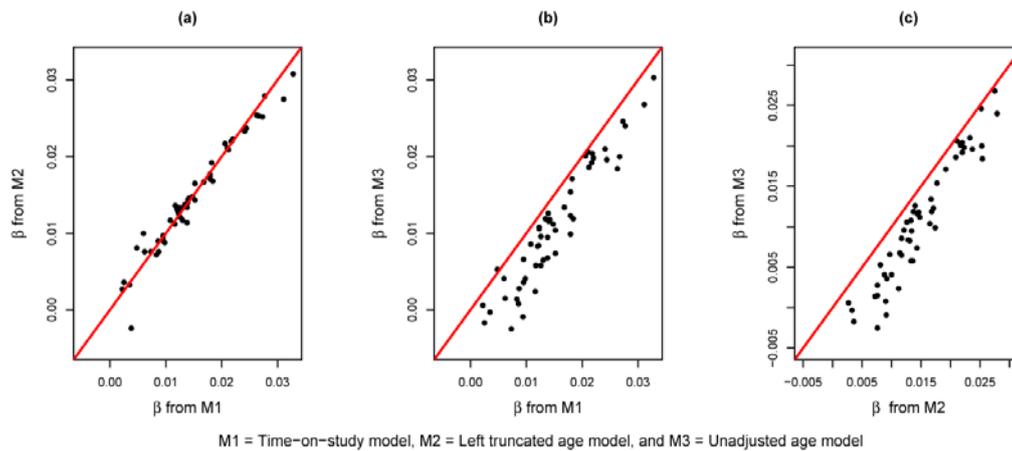
Figure 1. (a) Plot of coefficients from the time-on-study model vs left truncated age model (b) Plot of coefficients from the time-on-study model vs unadjusted age scale model (c) Plot of coefficients from the left adjusted age model vs unadjusted age scale model.

The correlation and exponentiality divide up the 54 cohorts. **Table 1** presents the number of cohorts for which the estimated regression coefficients are significantly different among the 54 cohort data sets according to the magnitude of the correlation between *sbp* and *age at entry*. All of the correlation coefficients are positive and they range from 0.0963 to 0.6280. The comparisons were made at the $\alpha = 0.05$ level of significance. The regression coefficients from **M1** and **M2** are significantly different in 3 cases out of 54 cases. When the correlation is above 0.4 and below 0.2 none of the pairs of the models differ significantly but 3 cases are significant at the moderate values of the correlation (0.2 to 0.4). 1 case is significant out of 16 when correlation is between 0.3 and 0.4 and 2 cases are significant out of 15 when the correlation is in the range of 0.2 to 0.3. This shows no obvious influence of the correlation on whether the regression coefficient estimates are significantly different between models. The coefficients estimated from **M1** and **M3** are significantly different in 40 cases out of 54 cases. Similarly, the estimated regression coefficients are significantly different between **M2** and **M3** in 53 cases out of 54.

Table 2. Number of cohorts for which the estimated regression coefficients are significantly different between models by whether the baseline hazard appeared to be exponential or not (**M1**: Time-on-study Model, **M2**: Left truncated age model, **M3**: Unadjusted age scale model)

	M1 vs M2		M1 vs M3		M2 vs M3	
Exponential	**Sig**	**Non-sig**	**Sig**	**Non-sig**	**Sig**	**Non-sig**
Yes	1	17	14	4	18	0
No	2	34	26	10	35	1
Total	3	51	40	14	53	1

Next, we estimated the cumulative baseline hazard function as a function of age using Breslow's method (1972) and plotted them against age. We further plotted the log of the cumulative hazard estimate against age to see if it appears linear. Example figures of one exponential and one non-exponential case are shown in appendix **Figure A**. The results show that 18 cases appear close to the exponential form of hazard and the remaining 36 plots looked having non-exponential form. **Table 2** represents the number of cases showing whether the results from three pairs of models are different or not according to the shape of the cumulative baseline hazard function. The regression coefficients from **M1** and **M2** are significantly different in 1 case even though the cumulative baseline hazard function appears exponential. Similarly, **M1** and **M2** are different from **M3** in 14 and 18 cases with the baseline hazard exponential.

We also noted that in one cohort data (Cohort 35, Appendix **Table A**), although not significant, **M1** and **M2** indicate the association between the risk factor and hazard of occurring disease are in opposite direction. In 5 cohort data sets the coefficients from **M1** (+ve) and **M3** (-ve) and in 4 cases the coefficients from **M2** (+ve) and **M3** (-ve) are in opposite direction.
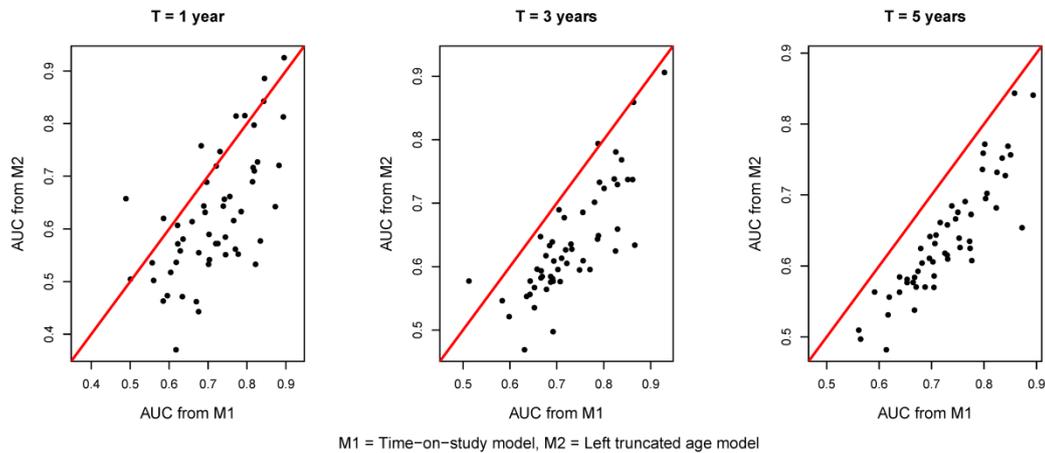
Figure 2. Plot of time dependent area under curve (values) for time-on-study vs left truncated age scale models at time points 1, 3 and 5 years respectively

The comparison of the regression coefficients shows that unadjusted age scale model (**M2**) is always inferior to time-on-study (**M1**) and left truncated age scale model (**M3**). The time-on-study and left truncated age scale models are closer. This result is consistent with previous simulation studies (Thiebaut and Benichou, 2004; Pencina et al., 2007; Chalise et al., 2012; Chalise et al., 2013). Therefore, we do not consider **M3** further. We next explore how **M1** and **M2** compare with respect to the predictive discrimination of the models.
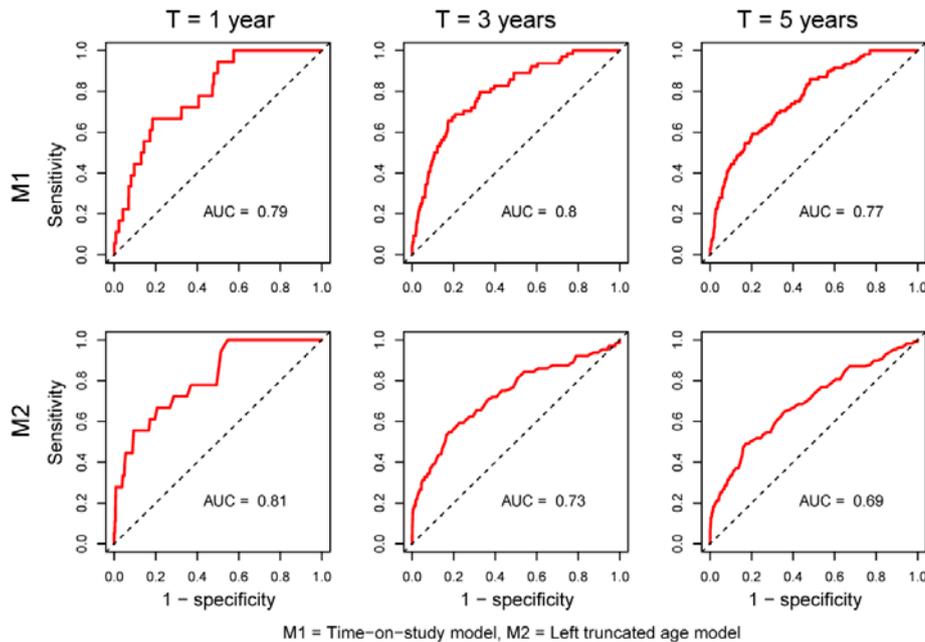
Figure 3. An example of time dependent area under receiver operating curve at time points 1, 3 and 5 years for time-on-study (row 1) and left truncated age scale model (row 2).

Table 3. Time dependent area under curves estimated at 1, 3 and 5 years for three cohort data set for which the estimated regression coefficients are significantly different between time-on-study and left truncated age scale model (**M1**: Time-on-study Model, **M2**: Left truncated age model)

| | **M1** | | | **M2** | | |
Sl no	**T = 1**	**T = 3**	**T = 5**	**T = 1**	**T = 3**	**T = 5**
1	0.5603	0.5984	0.6171	0.5023	0.5213	0.5312
2	0.6178	0.7047	0.7165	0.3705	0.6896	0.6614
3	0.5567	0.7102	0.6534	0.5358	0.6135	0.5813

Appendix **Table B** represents the time dependent area under curves computed at 1, 3 and 5 years for **M1** and **M2**. In most of the cases, the predicted area under the curve is higher for **M1** as compared to those from **M2**. At time points 1, 3 and 5 years, the AUC from **M1** were bigger in 45, 52 and 54 cases respectively, **Figure 2**. Examples of such area under the curves are shown in **Figure 3**. The AUCs in the three cohort dataset in which the regression coefficients from **M1** and **M2** were significantly different are shown in **Table 3**. In all three cases, the AUCs are higher for **M1** as compared to **M2**.   In order to better understand the prediction trajectory over several time points, we created sequence of 50 time points between 1 to 5 years and AUC were estimated at each of the time points. An example of such prediction trajectory for both models is shown in **Figure 4(a)** and the difference between the two prediction curves is shown in **Figure 4(b)**.   In many cohort data sets, the prediction curve for time-on-study model are higher than that for left truncated age model, while in others the curves are closer for some time points (other plots are not shown).
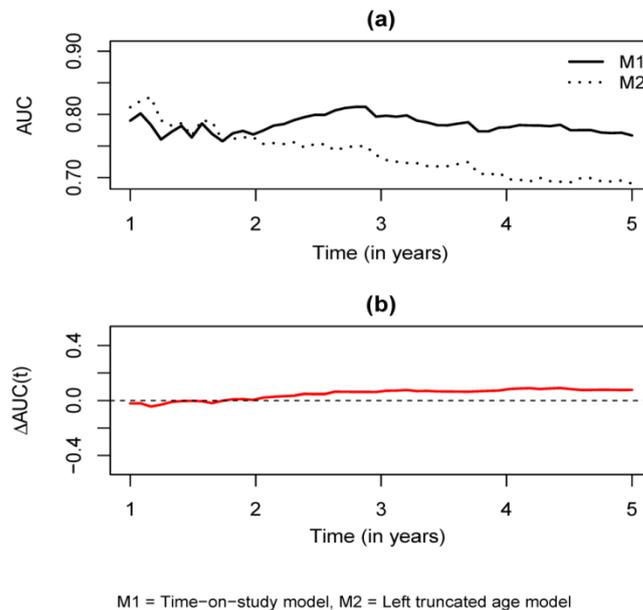

Figure 4. An example of trajectory of time dependent area under curve computed at 50 different time points within 1 to 5 years for the same example data presented in Figure 3. (a) Trajectories of time dependent area under curves for time-on-study and left truncated age scale models (b) Difference in area under the curves between the two models over the time range

These empirical results are consistent with the simulation studies by Chalise et al. (2013). Chalise et al. (2013) concluded that the time-on-study models are robust to misspecification of the underlying time scales. Here, we notice that, in most cases, the time-on-study models have at least as good predictive discrimination as the age scale models do.

4. Discussion

In this report we presented a comparison of results of PH models using two different time scales in 54 different datasets. In our results, the estimates obtained from the time-on-study and left truncated age time scale models are closer in most of the cases. However, we did not observe any close influence of (i) low degree of correlation between covariate and baseline age and (ii) exponentiality of the baseline hazard, on the significant difference of regression coefficient estimates. Our empirical example coincides with what one would expect given the method of estimation for the PH model. In order to estimate the coefficients, partial likelihood function is used in which the baseline hazard is factored out and the time scale provides the method of ordering the time to event for determining risk sets. If ages of all subjects at baseline were the same, the models would all be equivalent. The distribution of ages at baseline plays a role in determining whether the time scales will produce equivalent results (Chalise et al., 2012). The low performance of unadjusted age scale model as compared to the other two models discussed in this article shows that it is extremely important to appropriately adjust the model for the baseline age.

In some situations, there may be multiple plausible time scales. For example, automobile warranties usually use two time scales, calendar time and cumulative mileage; in studies of skin cancer among occupationally exposed workers, cumulative exposure to radiation may provide a better time scale than does the person's age or time on study. Some investigators have examined methods for deriving an optimal time scale.   Farewell and Cox (1979) and Oakes (1995) suggested choosing a time scale that combines two or more times scales in such a way that the resulting scale accounts

for as much the variation as possible. Duchesne and Lawless (2000) introduced the concept of an ideal time scale. Their work, however, focuses on usage (e.g. mileage) or exposure (e.g. asbestos exposure) variables that used as time scales (and could be adapted to epidemiological analyses in some instances) but do not solve the problems inherent in comparing time scales differ only in having different origins.

It is important to perform simulation studies to assess the models as presented by previous studies. A shortcoming of doing so is that the comparisons are based on some pre-specified data generating mechanism. However, in real datasets we have no clue about such mechanism. Therefore, it is extremely valuable to use real datasets because the ultimate purpose of application of any statistical model is to analyze real data. The strength of our empirical examination presented in this article is that we have assessed the models in large number of datasets collected in several parts of the world using varying data collection scheme. The empirical results agree with the simulation studies by Chalise et al. (2013) in that the time-on-study model is at least as good as left truncated age scale model. However, this will not establish which model is the best in general. Both models adjust for age at entry but in different ways. In time-on-study model the baseline age is adjusted as a parametric covariate while the baseline age is handled non-parametrically in left truncated age scale model. What our study shows is that if the adjustment for age at entry is made there is very little difference in the regression coefficient estimates between the two models but predictive discrimination is generally higher for time-on-study model. A unique well-defined time scale is indispensable for event history analysis. Therefore, given the lack of an agreed upon definition for an optimal or even correct time scale for the study under consideration, when working with real data, our studies show that using time-on-study model may be preferable since it performs at least as well as the age scale model.

Acknowledgments

Conflict of Interest

None declared

References

Allison, D., Fontaine, K., Manson, J., Stevens, J., & VanItallie, T. (1999). Annual deaths attributable to obesity in the Unites States. *Lifetime Data Analysis*, *16,* 1530-1538. http://dx.doi.org/10.1001/jama.282.16.1530.

Anderson, P., & Gill, R. (1982). Cox's Regression model for counting processes: A large sample study. *The Annals of Statistics, 10,* 1100-1120. http://dx.doi.org/10.1214/aos/1176345976

Blanche, P, Dartigues, J. F., & Jacqmin-Gadda, H. (2013). Estimating and comparing time dependent areas under receiver operating characteristics curves for censored event times with competing risks. *Statistics in Medicine*, 32:5381-5397. http://dx.doi.org/10.1002/sim.5958

Chalise, P., Chicken, E., & McGee, D (2012). Baseline age effect on parameter estimates in Cox models. *Journal of Statistical Computation and Simulation*, *82,* 1767-1774. http://dx.doi.org/10.1080/00949655.2011.595010

Chalise, P., Chicken, E., & McGee, D. (2013). Performance and prediction for varying survival time scales. *Communications in Statistics - Simulation and Computation*, *42,* 636-649. http://dx.doi.org/10.1080/03610918.2011.650259

Cheung, Y., Gao, F., & Khoo, K. (2003). Age at diagnosis and the choice of survival analysis methods in cancer epidemiology. *Journal of Clinical Epidemiology*, *56,* 38-43. http://dx.doi.org/10.1016/S0895-4356(02)00536-X

Conroy, R., Pyorala, K., Fitzerald, A., Sans, S., Menotti, A., De Backe, G., …Graham, I. M. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Europian Heart Journal*, *24*, 987-1003. http://dx.doi.org/10.1016/S0195-668X(03)00114-3

Cox, D. (1972). Regression models and life tables (with discussion). *J R Stat Soc B*., *34*, 187-220.

Cox, D. (1975). Partial Likelihood. *Biometrika*, *62*, 269-276.

Duchesne, T., & Lawless, J. (2000). Alternative time scales in failure time models. *Lifetime Data Analysis*, *6*, 157-179. http://dx.doi.org/10.1023/A:1009616111968

Farewell, V., & Cox, D. (1979). A note on multiple time scales in life testing. *Applied Statistics*, *28*, 73-75. http://dx.doi.org/10.2307/2346815

Flegal, K., Graubard, B., Williamson, D., & Gail, M. (2007). Cause-specific excess deaths associated with underweight, Overweight and Obesity. *JAMA, 298*, 2028-2037. http://dx.doi.org/10.1001/jama.298.17.2028

Hung, H, & Chiang, C. (2010). Estimation methods for time-dependent AUC with survival data. *Canadian Journal of*

Statistics, 38, 8-26. http://dx.doi.org/10.1002/cjs.10046

Korn, E., Graubard, B., & Midthune, D. (1997). Time-to-event Analysis of Longitudinal Follow-up of a survey: Choice of the time scale. *Am J Epidemiol*, *145*, 72-80.

McGee, D. L., & The Diverse Populations Collaboration. (2005). Body mass index and mortality. A meta-analysis based on person-level data from twenty-six observational studies. *Annals of Epidemiology*, *15*, 87-97. http://dx.doi.org/10.1016/j.annepidem.2004.05.012

Oakes, D. (1995). Multiple time scales in survival analysis. *Lifetime Data Analysis*, *1*, 7-18. http://dx.doi.org/10.1007/BF00985253

Pencina, M., Larson, M, & D'Agostino, R. (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine*, *26*, 1343-1359. http://dx.doi.org/10.1002/sim.2699

Thiebaut, A., & Benichou, J. (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine, 24*, 3803-3820. http://dx.doi.org/10.1002/sim.2098

Wilson, P., D'Agostino, R., Levy, D., Belanger, A., Silbershatz, H., & Kannel, W. (1998). Prediction of coronary heart disease using risk factor categories. *Journal of The American Heart Association, 97*, 837-1847. http://dx.doi.org/10.1161/01.CIR.97.18.1837

Appendix

Table A: Table showing the coefficient estimates from cox proportional hazards model, standard errors of the coefficients, pairwise comparison p-values with bootstrap method, correlation coefficients between the age at entry and covariate, and test of exponentiality of the baseline hazard function for **M1**, **M2** and **M3**. (**M1**: Time-on-study Model, **M2**: Left truncated age model, **M3**: Unadjusted age scale model)

	M1		M2		M3		p-values					Corr	Expo
SI	β_1	se_1	β_2	se_2	β_3	se_3	M1 and M2		M1 and M3		M2 and M3		Corr	Expo
1	0.0062	0.0065	0.0076	0.0067	0.0015	0.0066	0.6974		0.2162		0.0000	*	0.2369	1
2	0.0022	0.0035	0.0027	0.0034	0.0006	0.0035	0.2113		0.0014	*	0.0027	*	0.1321	0
3	0.0025	0.0019	0.0036	0.0019	-0.0017	0.0019	0.0000	*	0.0000	*	0.0000	*	0.2066	0
4	0.0120	0.0039	0.0131	0.0035	0.0083	0.0034	0.6722		0.2020		0.0006	*	0.5027	1
5	0.0138	0.0046	0.0114	0.0042	0.0068	0.0040	0.4388		0.0727		0.0365	*	0.3717	0
6	0.0184	0.0038	0.0168	0.0034	0.0119	0.0033	0.6058		0.0489	*	0.0011	*	0.5785	1
7	0.0060	0.0036	0.0100	0.0034	0.0041	0.0031	0.0568		0.4473		0.0002	*	0.4359	0
8	0.0087	0.0051	0.0076	0.0051	0.0028	0.0050	0.3974		0.0000	*	0.0000	*	0.2528	0
9	0.0035	0.0038	0.0033	0.0038	-0.0003	0.0037	0.8026		0.0113	*	0.0011	*	0.1455	0
10	0.0267	0.0045	0.0253	0.0045	0.0200	0.0045	0.3507		0.0008	*	0.0009	*	0.3933	0
11	0.0123	0.0027	0.0126	0.0026	0.0106	0.0026	0.8303		0.2880		0.0124	*	0.3981	1
12	0.0123	0.0034	0.0133	0.0032	0.0108	0.0033	0.6892		0.5922		0.0372	*	0.3681	0
13	0.0168	0.0028	0.0167	0.0026	0.0134	0.0026	0.9468		0.0589		0.0002	*	0.5358	1
14	0.0179	0.0025	0.0171	0.0023	0.0123	0.0023	0.6171		0.0077	*	0.0001	*	0.3939	0
15	0.0206	0.0024	0.0217	0.0021	0.0201	0.0021	0.4320		0.7389		0.0001	*	0.4704	0
16	0.0217	0.0022	0.0220	0.0021	0.0192	0.0021	0.7389		0.0124	*	0.0000	*	0.2325	0
17	0.0048	0.0058	0.0081	0.0058	0.0053	0.0057	0.0989		0.8026		0.0001	*	0.2633	0
18	0.0073	0.0047	0.0076	0.0049	-0.0025	0.0044	0.8745		0.0001	*	0.0000	*	0.6280	0
19	0.0094	0.0034	0.0091	0.0035	-0.0009	0.0033	0.8303		0.0000	*	0.0000	*	0.4660	0
20	0.0328	0.0048	0.0308	0.0045	0.0303	0.0045	0.4047		0.3173		0.2113		0.4964	0
21	0.0311	0.0037	0.0275	0.0036	0.0268	0.0036	0.0244	*	0.0114	*	0.0196	*	0.3558	0
22	0.0126	0.0043	0.0133	0.0043	0.0058	0.0042	0.5597		0.0001	*	0.0000	*	0.3383	0
23	0.0152	0.0049	0.0143	0.0049	0.0074	0.0049	0.4132		0.0000	*	0.0000	*	0.3896	1
24	0.0095	0.0053	0.0092	0.0053	0.0036	0.0052	0.8026		0.0002	*	0.0010	*	0.3312	0
25	0.0116	0.0046	0.0112	0.0045	0.0024	0.0045	0.7389		0.0000	*	0.0000	*	0.4168	1
26	0.0218	0.0020	0.0220	0.0020	0.0204	0.0019	0.6892		0.0196	*	0.0000	*	0.1917	0
27	0.0179	0.0021	0.0177	0.0021	0.0154	0.0021	0.6171		0.0000	*	0.0000	*	0.2985	1
28	0.0142	0.0014	0.0146	0.0014	0.0118	0.0014	0.3173		0.0000	*	0.0000	*	0.2186	0
29	0.0212	0.0013	0.0209	0.0012	0.0186	0.0012	0.5485		0.0000	*	0.0000	*	0.3217	1
30	0.0179	0.0053	0.0174	0.0052	0.0099	0.0051	0.8118		0.0014	*	0.0007	*	0.5272	0

| 31 | 0.0086 | 0.0044 | 0.0090 | 0.0044 | 0.0008 | 0.0043 | 0.8242 | | 0.0000 | * | 0.0000 | * | 0.3849 | 1 |
32	0.0083	0.0066	0.0072	0.0065	0.0014	0.0063	0.6722		0.0553		0.0203	*	0.5318	0
33	0.0117	0.0047	0.0136	0.0047	0.0058	0.0047	0.2912		0.0032	*	0.0000	*	0.4051	1
34	0.0263	0.0082	0.0254	0.0081	0.0184	0.0079	0.6357		0.0024	*	0.0000	*	0.1744	0
35	0.0038	0.0093	-0.0024	0.0091	-0.0072	0.0089	0.1777		0.0145	*	0.0002	*	0.1904	0
36	0.0098	0.0045	0.0088	0.0045	0.0041	0.0044	0.2113		0.0000	*	0.0000	*	0.2293	0
37	0.0130	0.0042	0.0117	0.0042	0.0065	0.0041	0.1042		0.0000	*	0.0000	*	0.2487	0
38	0.0095	0.0035	0.0097	0.0034	0.0066	0.0035	0.8557		0.0383	*	0.0019	*	0.5518	1
39	0.0138	0.0034	0.0134	0.0034	0.0095	0.0034	0.7389		0.0114	*	0.0053	*	0.3837	1
40	0.0126	0.0018	0.0121	0.0018	0.0096	0.0018	0.3173		0.0000	*	0.0000	*	0.5763	0
41	0.0122	0.0017	0.0128	0.0016	0.0084	0.0016	0.2301		0.0000	*	0.0000	*	0.3994	0
42	0.0140	0.0029	0.0144	0.0029	0.0116	0.0028	0.5677		0.0077	*	0.0000	*	0.4904	1
43	0.0108	0.0023	0.0117	0.0023	0.0086	0.0023	0.0719		0.0002	*	0.0000	*	0.3183	1
44	0.0273	0.0032	0.0252	0.0032	0.0246	0.0032	0.0196	*	0.0027	*	0.0027	*	0.2412	1
45	0.0210	0.0017	0.0211	0.0017	0.0206	0.0017	0.6171		0.1824		0.0000	*	0.0963	0
46	0.0241	0.0037	0.0233	0.0036	0.0210	0.0035	0.4670		0.0268	*	0.0010	*	0.2260	0
47	0.0220	0.0021	0.0223	0.0020	0.0198	0.0020	0.5485		0.0002	*	0.0000	*	0.2393	0
48	0.0139	0.0015	0.0140	0.0015	0.0126	0.0015	0.8026		0.0012	*	0.0000	*	0.2771	0
49	0.0134	0.0013	0.0137	0.0013	0.0119	0.0012	0.3173		0.0000	*	0.0000	*	0.1978	0
50	0.0182	0.0018	0.0192	0.0017	0.0171	0.0017	0.0956		0.0668		0.0000	*	0.2689	0
51	0.0148	0.0028	0.0148	0.0028	0.0112	0.0027	1.0000		0.0455	*	0.0056	*	0.5532	1
52	0.0152	0.0027	0.0165	0.0027	0.0104	0.0025	0.2787		0.0115	*	0.0007	*	0.2619	1
53	0.0277	0.0048	0.0279	0.0047	0.0240	0.0047	0.8940		0.0515		0.0004	*	0.3510	0
54	0.0244	0.0034	0.0237	0.0033	0.0196	0.0032	0.6171		0.0027	*	0.0000	*	0.3933	0
							1 - Exponential,				**0 - non-exponential**			**18**
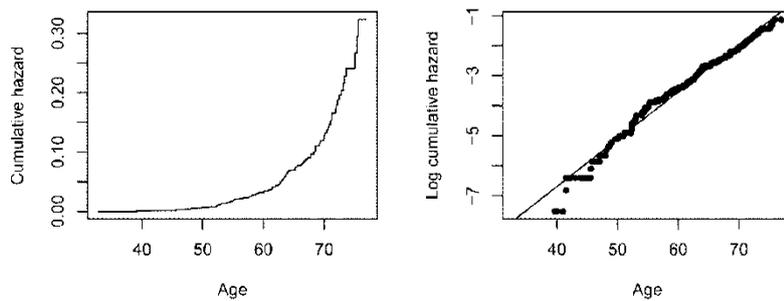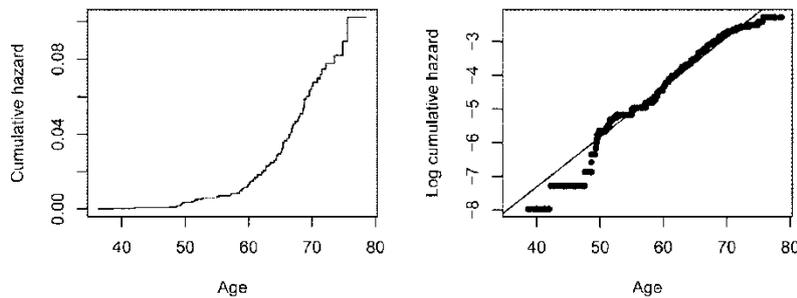
Figure A. Two examples of cumulative baseline hazard and corresponding log-cumulative hazard plots. Example 1 represents an example of exponential case and Example 2 represents an example of non-exponential case.

Table B. Table showing the time dependent area under the ROC curves estimated at 1, 3 and 5 years for **M1** and **M2** (**M1**: Time-on-study Model, **M2**: Left truncated age model)

sl.no	M1			M2		
	T = 1	T = 3	T = 5	T = 1	T = 3	T = 5
1	0.6700	0.6920	0.5612	0.4619	0.4976	0.5097
2	0.6341	0.6311	0.6137	0.4716	0.4691	0.4820
3	0.5603	0.5984	0.6171	0.5023	0.5213	0.5312
4	0.8221	0.7867	0.8056	0.5330	0.6436	0.7023
5	0.7703	0.8252	0.7726	0.5618	0.6248	0.6350
6	0.8451	0.9294	0.8936	0.8860	0.9062	0.8409
7	0.6826	0.7159	0.7255	0.7580	0.6774	0.6181
8	0.6926	0.6664	0.6818	0.6316	0.5824	0.6043
9	0.6762	0.6524	0.6674	0.5550	0.5672	0.5378
10	0.8935	0.7804	0.7754	0.8129	0.7014	0.6727
11	0.5850	0.8513	0.7984	0.4631	0.7373	0.7590
12	0.7199	0.6904	0.7386	0.5723	0.6393	0.6846
13	0.8956	0.8636	0.8582	0.9254	0.8591	0.8436
14	0.8183	0.7910	0.8456	0.7973	0.7329	0.7688
15	0.7446	0.7320	0.7085	0.5847	0.6278	0.6438
16	0.7946	0.8008	0.7640	0.8152	0.7233	0.6908
17	0.7253	0.6936	0.7298	0.5724	0.6090	0.6155
18	0.8144	0.8224	0.8247	0.6896	0.7381	0.7320
19	0.8345	0.8294	0.8233	0.5773	0.6594	0.6818
20	0.8825	0.7214	0.6959	0.7207	0.6054	0.6110
21	0.6178	0.7047	0.7165	0.3705	0.6896	0.6614
22	0.7393	0.6871	0.6676	0.6432	0.5758	0.5841
23	0.7421	0.6846	0.7458	0.6564	0.6334	0.6666
24	0.7560	0.6921	0.6741	0.6615	0.5772	0.5927
25	0.6892	0.6910	0.7026	0.6435	0.5814	0.6059
26	0.5956	0.6870	0.7049	0.4734	0.5846	0.5860
27	0.5858	0.6652	0.6969	0.6202	0.6475	0.6417
28	0.6359	0.6780	0.7045	0.5809	0.5643	0.5700
29	0.7212	0.7189	0.7302	0.7195	0.6266	0.6581
30	0.8432	0.8381	0.7970	0.8426	0.7683	0.7361
31	0.8160	0.7887	0.8030	0.7165	0.6491	0.6950
32	0.7721	0.7879	0.8015	0.8145	0.7940	0.7717
33	0.8190	0.8258	0.8346	0.7102	0.7808	0.7520
34	0.5008	0.5837	0.5914	0.5047	0.5464	0.5634
35	0.4891	0.5127	0.5647	0.6574	0.5775	0.4970
36	0.7852	0.7024	0.6649	0.6330	0.5957	0.5770
37	0.6223	0.6432	0.6193	0.5719	0.5774	0.5562
38	0.6758	0.7485	0.7543	0.4429	0.5950	0.6262
39	0.7026	0.7302	0.7526	0.5898	0.6358	0.6394
40	0.8273	0.8293	0.8406	0.7272	0.7294	0.7273
41	0.7452	0.7561	0.7735	0.5512	0.6095	0.6250
42	0.7015	0.7555	0.7502	0.5329	0.6857	0.6758
43	0.7776	0.7707	0.7767	0.5523	0.5956	0.6078
44	0.5567	0.7102	0.6534	0.5358	0.6135	0.5813
45	0.6288	0.6357	0.6389	0.5585	0.5532	0.5632
46	0.6042	0.6522	0.6533	0.5176	0.5354	0.5768
47	0.7309	0.6667	0.6793	0.7470	0.5936	0.6249
48	0.6592	0.6580	0.6390	0.6141	0.5962	0.5846
49	0.6184	0.6689	0.6707	0.5367	0.5845	0.5704
50	0.6966	0.6768	0.7064	0.6887	0.6174	0.6318
51	0.7659	0.8615	0.8506	0.6161	0.7372	0.7566
52	0.8729	0.8662	0.8725	0.6423	0.6341	0.6541
53	0.6217	0.6428	0.6876	0.6070	0.5564	0.5704
54	0.7037	0.7073	0.7307	0.5418	0.5766	0.6100