A Simple Approach to Assess Group Differences in Estimated Baseline Survivor Functions from Cox Proportional Hazards Models

Abstract

We develop a heuristic approach to generate the expected events that would occur under a Cox proportional hazards model and illustrate how the results can be used to test for group differences in the resulting risk-adjusted survivor distributions. The estimated baseline survivor functions from a Cox proportional hazards models fit separately for two groups are used to construct the expected number of events occurring at each event time, the expected number of interval-censored observations, and the adjusted risk set corresponding to the expected decrements due to events and censoring for each group. Differences in the resulting expected survival distributions can be tested using log-rank and generalized Wilcoxon tests. This method should also prove useful for making other kinds of comparisons of adjusted life tables.

A Simple Approach to Assess Group Differences in Estimated Baseline Survivor Functions from Cox Proportional Hazards Models

Introduction

A considerable amount of applied research has focused on disparities and group differences in health and socio-demographic outcomes. When outcomes reflect events, or transitions from one state to another, group differences in survivorship are often of major interest. Group differences in survivor functions are often examined descriptively using life-table techniques, and are assessed more formally using various forms of the log-rank (Savage 1956), or generalized Wilcoxon (Gehan 1956) statistics, which effectively summarize the survivor distributions and lead to a test of the difference in the observed and expected number of events conditional on the total number of events occurring at a particular time. While these tests are useful in exploratory analysis, they are less suited for testing differences between survival functions that have been adjusted for many risk factors, such as fitted survivor functions from multivariate models.

In hazard rate models, the adjusted, predicted, or model-based, baseline survivor function depicts the survivorship experience that would be expected under the condition that all covariates equal 0. Thus, in the case of a model consisting of a set of factors that increase the risk of an event, we would expect the predicted baseline survivor function to be shifted towards 1. In this case, the risk-adjusted baseline survivor function represents a counterfactual survival experience of a hypothetical homogenous low-risk cohort. When such adjustments are made in two or more populations, tests of differences between the risk-adjusted survivor functions can be informative

about the possible impact that changes in the population composition on key risk factors would have on socio-demographic and health disparities across groups.

Tests of differences in survivor distributions have not thus far been extended to testing group differences in *predicted* baseline survival rates obtained from separate multivariate hazard-rate models. This paper fills that gap by extending the log-rank and generalized Wilcoxon tests to assess the difference in the estimated baseline survivor functions from Cox (1972) proportional hazards models. This approach provides tests of group differences as well as a basis for constructing a risk-adjusted life table, which may be useful in its own right for predicting the number of events that would be expected to occur at specific times or ages. In essence, we are interested in the extent to which group differences in survival would diminish or disappear if both group's risk factors could be eliminated. This paper is structured as follows: Section 1 provides background to the statistical techniques underlying tests of group differences in survivor functions; Section 2 outlines how this approach can be adapted to test the difference between fitted baseline survivor functions from Cox proportional hazard models; Section 3 provides an empirical example; and, Section 4 provides a discussion and limitations of this approach.

1.0 Background

Group differences in observed survival rates are commonly assessed using variants of the Mantel-Haenszel approach (Mantel and Haenszel 1959; Mantel 1966). While there are modern extensions of this general approach (see e.g., Heller and Venkatraman 1996), traditional approaches include the log-rank test and generalized Wilcoxon tests (Cox and Oakes 1984), which are designed to test the overall equality of survivor functions for J groups using

information about the events and the number at risk at each event time t_k. Similar approaches have been derived from the perspective of counting processes (see, e.g., Aalen 1978; Harrington and Flemming 1982).[1] These tests are intuitively appealing due to their similarity with methods for testing association in contingency tables. The two most frequently-used tests for this purpose are the log rank test (Savage 1956) and the Generalized Wilcoxon test (Prentice 1982).

The Log-rank Generalized Wilcoxon Tests

The generalized Wilcoxon and log-rank tests provide an efficient means to test for the equality of survivor distributions. We borrow from the discussions by Kalbfleisch and Prentice (1980) and Cox and Oakes (1984), who provide derivations of the log-rank test. Suppose that we wish to test the equality of survivor functions $S_1(t),\ldots S_J(t)$ ($j=1,\ldots,J$) obtained from samples from each of J populations. Let $t_1 < t_2 < \ldots < t_k < \ldots t_K$ denote the observed event times obtained by pooling over the J samples. The times constitute a set of order statistics and as such, the rankings—not the actual values of t—are relevant for the nonparametric analysis adopted here.

Tests of difference in survivor functions consider the observed number of events experienced in group j at time t_k, denoted by d_{jk}, and the observed number at risk in group j just prior to time t_k, denoted by R_{jk}. The data at time t_k can be structured in the form of a $2\times J$ contingency table containing d_{jk} events and $R_{jk}-d_{jk}$ survivors in the jth row of the table. It can be shown that the distribution of $d_{1k},\ldots,d_{Jk}$, conditional on the event and censoring experience of the samples up to time t_k, is the product of binomial distributions.

[1] The log-rank test is often referred to as the generalized Savage test). Generalized Wilcoxon tests are often referred to under the names Breslow, Tarrone-Ware, and Prentice All of these tests differ only with regard to the weights to compute the score statistic and its variance.

$$\prod_{j=1}^{J}\binom{R_{jk}}{d_{jk}}h_k^{d_k}(1-h_k)^{R_k-d_k}, \tag{1}$$

where h_k is the event probability at t_k common to each of the J samples under the null hypothesis of equality of survivor functions. It can be shown that the conditional distribution of events in the J populations, given the total number of events (d_k) at t_k follows a hypergeometric distribution

$$\frac{\prod_{j=1}^{J}\binom{R_{jk}}{d_{jk}}}{\binom{R_k}{d_k}} \tag{2}$$

From Eq.(2), the expected number of events at time t_k for each group is given by $R_{jk}\frac{d_k}{R_k}$, where $d_k=\sum_{j=1}^{J}d_{jk}$ and $R_k=\sum_{j=1}^{J}R_{jk}$.

The statistical test utilizes a score statistic as a measure of the difference between the observed and expected number of events experienced in group j at time t_k. The log-rank and generalized Wilcoxon tests differ in the weights used to compute the score statistic and its variance. The most general form of the test is due to Harrington and Flemming (1982), in which the weights are defined as

$$w_k=\hat{S}_+(t_k)^{\rho} \qquad \text{for } 0\le\rho\le 1, \tag{3}$$

where $\hat{S}_+(t_k)$ is an estimate of the pooled survivor function,

$$\hat{S}_+(t_k)=\prod_{m=1}^{k}\left(\frac{R_m}{R_m+1}\right). \tag{4}$$

The value of ρ ranges from 0 to 1 and determines the sensitivity region of the test. A choice of $\rho = 0$ leads to a log-rank test, which is more sensitive to differences in survivor functions at later durations. A value of $\rho = 1$ results in a "generalized Wilcoxon" test (Prentice 1982), which is more sensitive to differences in survivor functions at earlier durations. For each of the test statistics we construct a $J \times 1$ vector of score statistics, the *j*th element of which is

$$u_j = \sum_k w_k \left(d_{jk} - R_{jk} \frac{d_k}{R_k} \right). \tag{5}$$

The variance in the score statistic is the $J \times J$ matrix, with element *jl* given by

$$v_{jl} = \sum_k w_k^2 \frac{d_k (R_k - d_k)}{R_k - 1} \frac{R_{jk}}{R_k} \left(I(j = l) - \frac{R_{lk}}{R_k} \right), \tag{6}$$

where $I(j = l)$ is the indicator function equal to 1 if $j = l$ and 0 otherwise. The log-rank and generalized Wilcoxon test statistics are calculated as

$$S = \mathbf{u}' \mathbf{V}^{-1} \mathbf{u} \tag{7}$$

which follows a χ^2 distribution with $J - 1$ degrees of freedom. The rank of $\mathbf{V}$ is $J - 1$, thus we can omit the last dimension. For example, in a two-group comparison we can simply use

$$\chi^2_{(1)} = \frac{u_1^2}{v_{11}}. \tag{8}$$

We are aware of no analogous approaches to evaluate differences in *predicted* survival functions from multivariate statistical models such as the Cox proportional hazards model (Cox 1972), which is perhaps the most widely-used model for survival analysis in demographic and health research. Given the widespread use of this model in contexts where hazard rates are adjusted for known risk factors, extending this approach to test group differences in predicted baseline survival rates obtained from multivariate models seems warranted.

Apart from the utility of testing group differences in predicted survivor functions, the approach provides a method to generate predicted events at t_k under a specific model, thereby showing the hypothetical effect of adjusting for known risks and protective factors on the number of events occurring at time t_k. We first provide some background to the conventional approach to assessing group differences using Cox proportional hazards models and outline some the limitation of this approach.

Assessing Group Differences with Proportional Hazards Models

A somewhat more restrictive test of group difference in survivorship can be obtained from a proportional hazards model. Although these methods apply to a multi-group setting, we focus here on a two-group comparison. For example, a proportional hazards model of event occurrence for individual i at time t_k can be expressed in terms of a binary group membership variable z and a covariate x as

$$h(t_{ik}) = h_0(t_k)\exp(\beta_1 z_i + \beta_2 x_i). \quad (9)$$

A test of the null hypothesis that the log hazard ratio β_1 equals 0 for group variable z is equivalent to a test of a group difference in the log baseline hazard. It has been shown that the log-rank test outlined earlier can be carried out using a score statistic derived from evaluating the 1st-order conditions for the maximum partial likelihood function for the Cox regression model (Peto and Peto 1972; Peto and Pike 1973).

The model above is assumed to hold for each group. In particular, the baseline hazard for the group defined by $z = 1$ is identical to the baseline hazard for the $z = 0$ group shifted proportionally by $\exp(\beta_1)$ at any event time. To relax the proportionality assumption, researchers often allow group-specific baseline hazards by fitting a stratified Cox model based on the grouping variable. This approach assumes that the effects of x are the same for each stratum.

One drawback of this approach is that z is absorbed into the baseline hazard; its effect is not estimable as a consequence. Estimating a proportional hazard model separately for each group allows maximum flexibility by permitting group variation in both the baseline hazards and the covariate effects. Comparisons of group differences in covariate effects is straightforward, however differences in the baseline hazard are more difficult to assess as these are unspecified in the Cox regression model.

It would be of considerable interest to researchers to assess group differences in predictions from Cox models. A fruitful approach to do this is motivated from an evaluation of the estimated baseline survival function. While it is not generally possible to obtain estimates of the baseline hazard at time t_k directly from a Cox model, it is relatively straightforward to obtain estimates of the survivor function. Given an estimate of the survivor function and the number at risk at t_k, it is possible to work backwards to obtain an *adjusted* life table consisting of the expected number at risk and the expected number of events that would have occurred at t_k under a specific model. Next we show how the results from a Cox proportional hazard model may be used to generate predicted survivor functions that have been adjusted for risk factors and how this information may be used to generate the expected event distribution.

2.0 Applying Generalized Wilcoxon Tests to Estimated Baseline Survival Distributions

In order to apply the generalized Wilcoxon tests to predicted survival, we must first estimate the number of events that would have occurred at t_k, given the estimated survival functions and the number of individuals at risk at t_k. That is we need the estimates of the quantities given in Eqs. (4-7) provided earlier. We will assume that the number at risk at time t_k,

R_k, and the surviving faction, S_k, are known or can be estimated, where the number of subjects at risk at t_k includes those censored in the interval $[t_k, t_{k+1})$. A model-based approach to predicting the number of events that would occur at t_k considers the baseline survival function from a proportional hazards model. Specifically, the predicted survivor function at time t_k for individual i is

$$\hat{S}(t_{ik}) = S_0(t_k)^{\exp(\hat{\beta}x_i)}, \tag{10}$$

where $S_0(t_k)$ is the baseline survivor function at t_k and $\hat{\beta}$ are estimates of the effects of x.[2] Applying a –log transformation, the predicted survivor function for individual i at time t_k can be written in terms of the baseline cumulative hazard function, $H_0(t_k)$, as

$$-\log \hat{S}(t_{ik}) = H_0(t_k)\exp(\hat{\beta}x_i). \tag{11}$$

Aalen's estimator of $H_0(t_k)$ can be used to obtain an estimate of $S_0(t_k)$ (Aalen 1976). Specifically,

$$\hat{H}_0(t_k) = \sum_{t_k<t} \frac{d_k}{\sum_{i\in\mathcal{R}(t_k)} \exp(\hat{\beta}x_i)}, \tag{12}$$

where d_k denotes the number of events occurring at t_k, and $\mathcal{R}(t_k)$ denotes the risk set just prior to t_k. It follows that the estimated baseline survivor function at t_k can be estimated as

$$\hat{S}_0(t_k) = \exp[-\hat{H}_0(t_k)]. \tag{13}$$

The baseline survivor function is the model-adjusted counterpart of the empirical survivor function $S(t_k)$ and depends on the configuration of values of predictor variables under a specific model. If the predictors act uniformly to raise the risk of an event, then the predicted

[2]The baseline survivor function from a "null" Cox model results in the Kaplan-Meier estimate.

baseline survivor function will be shifted towards 1 when compared to the empirical survivor function. On the other hand, if predictors uniformly lower the risk of an event, then the predicted baseline survivor function will be shifted towards 0. The predicted baseline survivor function thus represents a counterfactual survivor function that has been adjusted upwards for *risk* factors and downward for *protective* factors. In this sense, it is most useful when the collection of predictors in a multivariate model represents an unambiguous set of *either* protective or risk factors, rather than a mixture of both types of factors.

We treat the estimated baseline survivor function, $\hat{S}_0(t)$, as analogous to the usual Kaplan-Meier estimate of $S(t)$ by assuming that the estimated baseline survival function has been generated according to the product-limit estimator as:

$$\hat{S}_0(t_k) = \prod_{k:t_k<t}\left(1-\frac{d_k}{R_k}\right) = \prod_{k:t_k<t}(1-\hat{q}_k), \qquad (14)$$

where d_k denotes the number of events which occur at t_k among those at risk (R_k) for a set of K unique event times (Kaplan and Meier 1958). It follows from this relationship that the conditional probability of an event at t_k can be obtained as

$$\hat{q}_k = \frac{\hat{S}(t_{k-1}) - \hat{S}(t_k)}{\hat{S}(t_k)}, \quad k = 2,\ldots,K-1, \qquad (15)$$

where $\hat{q}_1 = \frac{1}{\hat{S}(t_1)} - 1$.[3]

[3]The survivor function at t_1 corresponds to the first observed event time and $S(t_0) = 1.0$

An estimate of q_k is obtained as $\hat{q}_k = d_k / R_k$ according to the product limit estimator. It follows that under the current model, the predicted number of events occurring at t_k, given the number of individuals at risk at t_k can be estimated as

$$\hat{d}_k = R_k \hat{q}_k. \tag{16}$$

Thus, we estimate q_k using the estimated baseline survivor function and this in turn is used to estimate the number of events that are expected to occur at t_k under a specific proportional hazards model.

Further Considerations

Because the estimated number of events differs from the observed number of events, this will necessarily affect the composition of the risk set with the passage of time differently than what would occur under the observed pattern of events. Therefore, we adjust the subsequent number at risk according to the expected number of events occurring at t_k. This requires two decision rules. First, we shall assume that the expected number at risk initially (i.e., at t_1) is equal to the observed number at risk. To adjust the size of the risk set at points in time later than t_1 we must account for the number of censored observations that would have occurred, given the expected number of events (i.e., the counterfactual number of censored observations in the interval from t_k to t_{k+1}). The expected number of censored observations in the interval $[t_k, t_{k+1})$ is determined by adjusting the observed number of censored observations. Specifically, let c_k denote the observed number of individuals who are censored in the interval $[t_k, t_{k+1})$, and let d_k and $\hat{d}_k$ denote the observed and expected number of events at t_k, respectively. Then, the

predicted or *expected* number of observations censored at t_k is incremented or decremented as follows:

$$\hat{c}_k = c_k + (\hat{d}_k - d_k). \tag{17}$$

That is, if the model predicts more (fewer) events, this leads to an increase (decrease) the expected number of observations that would be censored in the interval $[t_k, t_{k+1})$.

Secondly, this adjustment necessarily affects the composition of the risk set at t_k. The process of predicting the number of events and the number censored changes the size of the risk set because the predictions will most assuredly differ from what is observed if the predictors have significant effects on the hazard. The expected number at risk at time t_{k-1} together with the expected number of events/censorings allows the number at risk to be decremented recursively according to the following rule:

$$\hat{R}_k = \hat{R}_{k-1} - \hat{d}_{k-1} - \hat{c}_{k-1}, \tag{18}$$

where R_1 is equal to the initial observed number at risk. In this way the number at risk is adjusted according to the model-based predictions. Note that this adjustment is similar to the way in which the number at risk in subsequent intervals is determined using the standard life table approaches, and the Kaplan-Meier estimator in particular. It can be shown that the predicted survivor function from a Cox model without covariates will exactly reproduce a cohort's survival experience. For example, the number of events at t_1 is $d_1 = S_0(t_1)R_1$ and the number of events at t_2 is $R_2[S_0(t_1) - S_0(t_2)]$. We have followed the same logic, with some modifications for censoring and risk-set size adjustment, to obtain the *predicted* events. Having obtained the predictions, we can compare the observed and predicted distributions of events, survivor distributions, and numbers at risk.

Log Rank and Generalized Wilcoxon Tests on Predicted Baseline Survival Functions

We can apply the generalized Wilcoxon test using the predicted values of d_k and R_k for each group. We then substitute the estimated quantities into the formulas for the generalized Wilcoxon tests outlined earlier yielding a model-based test. The revised score statistic based on the predictions is obtained by substituting the estimated quantities into Eqs. (4-7). In this way, we can use model-based estimates of the baseline cumulative hazard to estimate events, censorings, and the number at risk, which can then be used in statistical tests of group differences in risk-adjusted survivor functions.

3.0 Example

As an application of this approach, we consider the risk of teen pregnancy in a sample of non-Hispanic white women. This analysis uses the 1993 wave of the National Longitudinal Survey of Youth (NLSY). The NLSY is a nationally representative sample of approximately 13,000 youth aged 14-21 in 1979. Retrospective information about timing (age) of first intercourse, pregnancy, birth, and marriage are available in later years of the survey.[4] We focus on age of first pregnancy occurring while unmarried prior to age 20 (i.e., first premarital teen pregnancy). The event histories represented in this analysis reflect the early family-formation experiences of 2,277 non-Hispanic white women who were followed from 1979 until 1986, at which time the youngest member of the 1979 cohort would be about 20 years old.

[3]In the cases where a pregnancy was reported but information about age of first intercourse was missing, we imputed age at first intercourse. Numerous checks were carried out to match pregnancies with births or other pregnancy outcomes. In cases where a respondent reported a birth but was missing data on pregnancy, we imputed the pregnancy date as 9 months prior to the birth date. For a few cases, a teen pregnancy occurring close to age 20 could result in a live birth occurring about 9 months later. This information was used as a further check on the imputation of the age at first pregnancy.

There is a longstanding interest in adolescent pregnancy and its consequences. A growing body of literature explores the role of religion, as summarized in the 2001 report by the National Campaign to Prevent Teen Pregnancy (Whitehead et al. 2001). Religious participation has been shown to have general influence on sexual restraint. Moreover, specific religious subcultures may have particularly strong proscriptive sexual norms—e.g., Conservative Protestantism and Catholicism. Whitehead et al. (2001) find support for the idea that religiosity (variously defined) is associated with delayed sexual activity among some groups of teens, while at the same time suggesting that some religious teens may be less likely to use contraception when they do begin sexual activity.

Here we are interested in comparing the fitted baseline survivor functions of teen pregnancy for two groups of respondents: those respondents who were raised as Conservative Protestants, and those from other religious backgrounds, including those with no particular religious upbringing or religious preference. The multivariate models used to obtain the predicted baseline survivor functions include a set of individual-level and social background variables including: *low maternal education level* (i.e., mother's education < 12 years, $\hat{p} = 0.27$), *number changes in family structure* (family transitions) (range [0-10], $\bar{x} = 0.31$), *number of older siblings* (0-16, $\bar{x} = 1.91$), *and low frequency of church attendance* (i.e., less frequent than once per month, $\hat{p} = 0.54$.)[5]

Table 1 shows the hazard ratios and 95% confidence intervals associated with these estimates. All covariates are predictive a higher risk of teen pregnancy in this sample and the effects are similar across groups, with the exception of the effect of older siblings on the risk for conservative protestants. Figure 1 shows the observed survivor functions (a) along with the

[5] The values in parenthesis are the estimated means ($\bar{x}$) and proportions ($\hat{p}$).

predicted survivor functions (b) under the multivariate model for two groups of young women: those who were raised in Conservative Protestant families (Group 1) and those with other religious upbringing, including those with no particular religious upbringing (Group 2). Both figures show higher teen premarital pregnancy rates for women who were raised in Conservative Protestant families. As expected, the predicted baseline survivor functions are shifted towards 1 after adjusting for risk factors.

We apply the method outlined above using each group's estimated baseline survivor functions. Table 2 provides a summary of the observed and expected number of events and interval censoring along with the size of the risk set obtained by aggregating results over single-year age intervals.[6] As expected, the predicted number of events is lower and interval-censoring is higher after adjustment for risk factors. However, Conservative Protestants would be expected to experience a 46% percent reduction in teen pregnancy if risk factors were eliminated compared to an expected 30% reduction for those from other backgrounds. Thus, this approach reveals differential impacts of risk or protective factors on expected outcomes. Moreover, age-specific event counts are adjusted downward to a greater extent in earlier age intervals (ages 14-17) in the Conservative Protestant sample when compared to the second sample. Event counts in the age 14-17 interval are reduced by 65% for Conservative Protestants under the model, whereas those from other backgrounds achieve only a 26% reduction in events under this model). We can also see this pattern in the predicted survivor functions in Figure 1. In contrast to the marked reductions in incidence predicted for Conservative Protestants at early ages, we find

[6] The quantities d_a and $\hat{d}_a$ are different from those presented earlier as they represent the sum of the events (or expected events) over broad age intervals.

roughly similar predicted event reductions at later ages (34.3% among those from Conservative protestant backgrounds and 34.1% among those from other backgrounds).

We construct test statistics based on the individual-level predictions based on the approach outlined above. Table 3 provides a comparison of the observed and model-based Wilcoxon and log-rank tests. Both tests produce similar results. We can see that the survivor functions for Conservative Protestants and other groups differ at less than the 0.0001 level of significance. After adjustments for risk factors the differences are attenuated, but remain statistically significant at less than the 0.01 level.

4.0 Discussion

We provide a heuristic approach for predicting event histories under a specific proportional hazards model. The predictions consist of the event/censoring patterns and the evolving risk-set composition that would be expected to occur if risk factors could be eliminated. The approach works backwards from a predicted baseline survivor function to generate the expected events, censoring, and the number at risk that would prevail under a specific Cox regression model. This technique can be used to construct risk-adjusted life tables obtained by aggregating the predicted number of events over broader time or age intervals. The technique is also useful for evaluating group differences in survivorship, which is the main motivation in this paper. In essence, we are interested in the extent to which group differences in survival would diminish or disappear if both group's risk factors could be eliminated.

Generalized Wilcoxon tests provide a standard tool for assessing differences in empirical survivor functions. We show that these tests can be carried out using the model-based predictions. This technique inherits some limitations of the tests on which it is based. In particular, it is likely

to perform better when the ratio of hazard functions in the populations to be compared is relatively constant. It is also the case that simplifying assumptions are made regarding the independence of contingency tables formed at successive event times. The contingency tables at successive event times are clearly not statistically independent, but it would be difficult to formalize the distribution theory to take this into account. The approach is also limited insofar as the age structure of risk is fixed when making event predictions (i.e., the order statistics are not adjusted), although it is possible to predict 0 events for specific ages or times. Further work is needed to determine how this might be remedied by developing alternative procedures. We also do not take into account the sampling variability of the Cox regression estimates, although this might be overcome by carrying out replications using re-sampling procedures. Overall, this approach is more flexible than the standard approach involving stratification on auxiliary variables because, unlike stratification, it can accommodate continuous predictors. More complex model specifications involving time-varying covariates and non-proportional effects can also be considered.

References

Aalen, O. 1976. "Nonparametric Inference in Connection with Multiple Decrement Models." *Scandinavian Journal of Statistics*, 3: 15-27.

Aalen O. 1978. "Non Parametric Inference for a Family of Counting Processes." *Annals of Statistics* 6: 701-726.

Cox, D. R. 1972. "Regression Models and Life Tables (with discussion)." *Journal of the Royal Statistical Society, Series B*, 26: 103-110.

Cox, D. R. and D. Oakes .1984. *Analysis of Survival Data*. London: Chapman and Hall.

Gehan, E. A. 1956. "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples." *Biometrika,* 52: 203-223.

Harrington, D.P., and T.R. Flemming .1982. "A Class of Rank Test Procedures for Censored Survival Data," *Biometrika*, 69: 133-143.

Heller, G., and E.S. Venkatraman. 1996. "Resampling Procedures to Compare Two Survival Distributions in the Presence of Right Censored Data." *Biometrics*, 52: 1204-1213.

Kaplan, E. L., and P. Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association*, 55: 457-481.

Mantel, N. and W. Haenzel. 1959. "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease." *Journal of the National Cancer Institute*, 22: 719-748.

Mantel, N. 1966. "Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration." *Cancer Chemotherapy Report,* 50: 163-170.

Peto, R, and J. Peto. 1972. "Asymptotically Efficient Rank Invariant Test Procedures (with Discussion)." *Journal of the Royal Statistical Society*, Series A. 135: 185-206.

Peto, R., and M. C. Pike. 1973. "Conservatism of the Approximation $(O-E)^2 / E$ in the Log Rank Test for Survival Data or Tumor Incidence Data." *Biometrics*, 29: 579-584.

Prentice, R. L. 1987. "Linear Rank Tests with Right Censored Data." *Biometrika*, 65: 167-169.

Savage, I. R. 1956. "Contributions to the Theory of Rank Order Statistics—The Two Sample Case." *Annals of Mathematical Statistics*, 28: 968-977.

Tarone, R. and J. Ware. 1977. "On Distribution Free Tests for Equality of Survival Distributions." *Biometrika*, 64: 156-160.

Whitehead, B. D., Wilcox, B. L., Rostosky, S. S., Randall, B., & Wright M. L. C. 2001. *Keeping the Faith: The Role of Religion and Faith Communities in Preventing Teen Pregnancy.* Washington, DC: National Campaign to Prevent Teen Pregnancy

Table 1: Hazard Ratios and 95% Confidence Intervals from Cox Proportional Hazard Models of Premarital Pregnancy

	Conservative Protestant			Other		
		95% CI			95% CI	
Variable	**Hazard Ratio**	**lower**	**upper**	**Hazard Ratio**	**lower**	**upper**
Low Church Attendance	1.79	1.26	2.54	1.58	1.23	2.01
Number of Family Transitions	1.53	1.34	1.76	1.52	1.36	1.71
Low Maternal Education	1.64	1.19	2.27	1.90	1.51	2.39
Number of Older Siblings	1.06	0.99	1.15	1.02	1.02	1.14

Table 2: Number at Risk (R), Number of Events (d), and Number Censored (c) based on Empirical and Predicted Survivor Functions for Aggregated Single-year Age Intervals. †

Conservative Protestant

	Empirical				Model Based			
	R_a	d_a	c_a	S_a	$\hat{R}_a$	$\hat{d}_a$	$\hat{c}_a$	$\hat{S}_a$
Age Interval								
[13, 14)	606	1	4	0.998	606	0	6	0.999
[14, 15)	601	9	2	0.983	600	2	16	0.992
[15, 16)	590	17	18	0.955	582	4	44	0.978
[16, 17)	555	38	27	0.888	534	17	69	0.944
[17, 18)	490	44	66	0.803	448	28	98	0.894
[18, 19)	380	35	49	0.724	322	22	75	0.843
[19, 20)	296	20	16	0.673	225	15	26	0.807
[20+		0	260			0	184	
Number of Events		164				88		

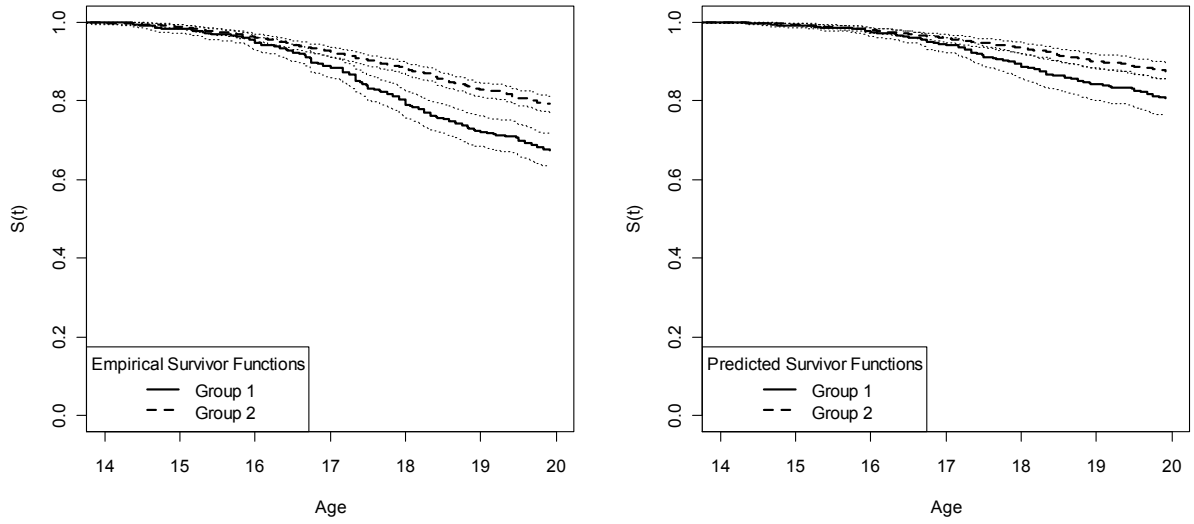
Other

	Empirical				Model Based			
	R_a	d_a	c_a	S_a	$\hat{R}_a$	$\hat{d}_a$	$\hat{c}_a$	$\hat{S}_a$
Age Interval								
[12, 13)	1671	1	0	0.999	1671	1	0	1.000
[13, 14)	1670	2	0	0.998	1670	2	0	0.999
[14, 15)	1668	17	1	0.988	1668	15	3	0.994
[15, 16)	1650	40	6	0.964	1648	31	15	0.981
[16, 17)	1604	61	17	0.927	1593	41	37	0.960
[17, 18)	1526	70	65	0.884	1495	47	88	0.935
[18, 19)	1391	84	88	0.829	1337	53	119	0.901
[19, 20)	1219	54	37	0.791	1134	39	52	0.876
[20+		0	1128			0	1043	
Number of Events		329				229		

† Note: R_a and $\hat{R}_a$ indicate the observed and expected risk set at the start of age interval a, respectively. S_a and $\hat{S}_a$ represent the empirical and predicted survivor functions at the end of age interval a. $d_a = \sum_{k \in a} d_k$, $\hat{d}_a = \sum_{k \in a} \hat{d}_k$, $c_a = \sum_{k \in a} c_k$, and $\hat{c}_a = \sum_{k \in a} \hat{c}_k$.

Table 3: Empirical and Model-based Tests of a Group Difference in Survivor Distributions Using Log-rank and Wilcoxon Tests

		Empirical				Model Based			
	N	**Obs.**	**Exp.**	χ^2	**Contrib.**	**Obs.**	**Exp.**	χ^2	**Contrib.**
Log Rank (ρ=0)									
Cons. Prot	606	164	112		23.96	88	66.15		7.22
Other	1671	329	381		7.06	229	250.85		1.90
				$\chi^2 =$	31.02			$\chi^2 =$	9.12
Wilcoxon (ρ=1)									
Cons. Prot	606	147	100		21.10	83.26	63.14		6.41
Other	1671	291	337		6.30	217.9	238.01		1.70
				$\chi^2 =$	27.40			$\chi^2 =$	8.11



(a) (b)

Figure 1: Observed (a) and predicted (b) survivor functions (a) for Conservative Protestant families (Group 1) and those from other religious backgrounds, including those with no particular religion (Group 2).