Fall 2021

Correcting for Measurement Error in the Outcome When Estimating the Distribution of Time to Pregnancy With the Current Duration Approach

Nicole Nasrallah

CORRECTING FOR MEASUREMENT ERROR IN THE OUTCOME WHEN ESTIMATING THE DISTRIBUTION OF TIME TO PREGNANCY WITH THE CURRENT DURATION APPROACH

by

Nicole Nasrallah

Bachelor of Science in Biochemistry
North Carolina State University, 2018

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Public Health in

Biostatistics

Arnold School of Public Health

University of South Carolina

2021

Accepted by:

Alexander C. McLain, Director of Thesis

Stella Self, Reader

Jiajia Zhang, Reader

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

ABSTRACT

The current duration approach to modeling time-to-pregnancy (TTP) models the length of pregnancy attempt for women that are currently attempting pregnancy. There is a scarcity of studies, let alone TTP studies, that account for measurement error in the outcome. Previously, the benefits of a piecewise constant model with regards to bias in estimates of the survival function with measurement error and the parametric modelling of TTP was shown. In this thesis, correcting for measurement error in the outcome with the current duration approach is explored through piecewise constant models with log-normal measurement error. Five different methods are compared to determine the optimal method in reducing the bias associated with the estimating the survival function. Specifically, we compared three naïve approaches (untransformed, shifted, and transformed), and two different simulation and extrapolation, or SIMEX, methods. The SIMEX method uses increasing measurement error variance to generate data with increasing measurement error variance that is fit to a piecewise constant model to approximate a quadratic model that is extrapolated for the approximate true coefficient value. The SIMEX method was tested with log-transformed and untransformed hazard rates from the parametric piecewise model, which we refer to as log SIMEX and regular SIMEX methods, respectively. The methods are further compared through their confidence intervals to determine the extent of precision when correcting

for measurement error in the outcome. The log SIMEX method has somewhat high variance and instances of the highest and lowest bias depending on the Weibull parameters. Nevertheless, it is shown to be an inconsistent method in correcting. For measurement error in the outcome. The transformed method performs the worst with regards to bias. Overall, there is no consistently best method to correct for measurement error in the outcome as the extent of bias and MSE depend on the measurement error variance and Weibull parameters.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFT ....Accelerated Failure Time Model

CDUI .... Current Duration of Unprotected Intercourse

CPC ....Cumulative Probability of Conception

HR...Hazard Ratio

SIMEX .... Simulation and Extrapolation Method

TTFP.. Time to First Pregnancy

TTP... Time to Pregnancy

CHAPTER 1: INTRODUCTION

1.1 INFORMATION ON INFERTILITY AND TIME TO PREGNANCY

Infertility, the inability to get pregnant after 12 months of unprotected intercourse, is a common reason for visits to the OB/GYN. About 6% of women aged 15-44 are considered infertile and 12% have difficulty getting pregnant ("Infertility | Reproductive Health | CDC" 2021). There are many causes of infertility; in women this includes hormonal imbalances, weight problems, and a variety of serious health conditions. In men, infertility can also be a result of health conditions, environmental toxins, and radiation treatment.

Although there has been research on infertility, it has been focused on treatments. By focusing on TTP, or the time it takes for a couple attempting to become pregnant to succeed, a more consistent prevalence of infertility can be determined. By studying the likelihood of conception, the management of fertility can be balanced, avoiding inaccurate estimations of pregnancies (Gnoth et al. 2003). Nevertheless, more attention can be drawn to infertility issues.

Evidence shows that TTP can be influenced by reproductive history, and contraceptive methods (Olsen et al. 1998) in addition to negative lifestyle choices such as smoking and alcohol (Hassan and Killick 2004; Ricci et al. 2017).

It can be challenging to accurately model TTP due to a plethora of latent factors. In addition, many studies in clinical research only account for error in the covariates or simply do not account for error at all. Much of the statistical literature on measurement error has focused on situations where there exists measurement in the covariate values (Carroll et al. 2006). There has been much less research focusing on methods correcting for error in the failure time outcome itself (Oh et al. 2018). Part of the reason for this is that with linear models, measurement error in the outcome will not impact estimates of regression coefficients. The aim of this thesis is to explore and evaluate the effectiveness of methods intending to correct for error in the TTP outcome.

1.2 TIME TO PREGNANCY STUDY DESIGNS

The main types of designs for studying TTP are retrospective, prospective, and cross-sectional. The most common study design in recent TTP studies is the retrospective study design. The most common ways data are collected are through structured questionnaires and surveys whether through the telephone (Jacobson et al. 2018), questionnaires (Jensen et al. 2001), or interviews (Juul et al. 2000). Retrospective studies have the advantage that they are less expensive, take less time and money compared to prospective studies, and represent a well-defined population (Joffe et al. 2005). However, one of the main disadvantages of this study design is the presence of recall bias. Radin et. al (2015) found recall bias was greater in pregnancy planners that conceived within 12 months, although not associated with maternal age. In addition to recall bias, these studies do not account for sterile couples or couples who give up an attempt to become pregnant in their analysis which can lead to biases in estimates of

the distribution of TTP. These selection effects have been associated with an opposite trend between age and fecundity when estimating the effect of age on TTP (Juul et. al 2000).

The second most common study design in TTP studies is the prospective study design in which a selected cohort is followed overtime to determine time to conception. Some studies, such as Gnoth et. al (2003), recruit women from the beginning of their attempt whilst keeping in mind whether or not the attempt coincides with the beginning of their menstrual cycle (Baird 2013). Other studies recruit women who have already started to conceive prior to the beginning of the study, which raises the issue of left-truncation (Radin et al. 2015). Unlike retrospective studies, prospective studies offer the advantage with respect to collecting data on additional factors such as the timing and frequency of intercourse in addition to other characteristics that effect pregnancy outcomes (Bonde et al. 2006). However, they are known to be impacted by attrition bias in addition to being expensive and time-consuming. Like retrospective studies, selection bias does exist in prospective studies since non-planners are missed when selecting a cohort (McLain et al. 2014).

Cross sectional studies differ from prospective and retrospective studies in that women are not recruited before they are at risk of pregnancy, like in the classic prospective approach, or after they are have achieved pregnancy, like in the retrospective approach (Slama et al. 2006). In the cross-sectional study design, women that are sampled are asked if they are currently having unprotected intercourse with the intent of getting pregnant and for how many months if applicable (Keiding et al. 2012).

Women taking part in studies with a cross-sectional approach are sometimes followed up in what is referred to as a prevalent cohort design or not followed up in the so-called current duration design.

Like the other study approaches, the cross-sectional study design has its advantages and disadvantages. Advantages include the ability to include subfertile and infertile couples (missed in retrospective studies) in addition to including accidental or non-planned pregnancies (missed in prospective studies). However, the observations from the current duration approach have length bias since those with longer durations are more likely to be included in the study as they are more likely to be attempting to get pregnant when surveyed. Further, all of the observations are right censored as the current duration of attempt, not the TTP, are observed (McLain et al. 2014). This results in the cross-sectional studies requiring special statistical models to estimate the distribution of TTP. Another potential problem is that the current duration design results in right censored observations. Therefore, there is no distinguishment between attempts ending with the event of pregnancy and attempts ending due to giving up (Slama et al. 2006).

While the current duration design has been a popular method to estimate the distribution of TTP, one of the main disadvantages of this study design is the presence of recall bias. This results in measurement error in the observed current duration values.

The aim of this thesis is to: **Investigate methods to correct for measurement error in the outcome with current duration data. Different methods will be evaluated**

using simulation studies. The various methods for estimating time to pregnancy in women will be discussed in Chapter 2.

CHAPTER 2: STATISTICAL METHODS FOR TTP AND CURRENT DURATION DATA

2.1 PARAMETRIC METHODS

The most common method in previous studies to analyze TTP is the parametric approach. Previous studies have used parametric approach with discrete TTP (van Eekelen et al. 2020; Weinberg and Gladen 1986) in addition to continuous TTP (Keiding 2011; Keiding et al. 2012; Zare et al. 2017).

Weinberg and Gladen (1986) based their model off the idea that sexually active couples vary in their fecundability, defined as a couple's ability to get pregnant regardless of intention. Therefore, they assumed that the variation in pregnancy rates are the result of heterogeneity in fecundability between couples and not due to time effects. Weinberg and Gladen (1986) assume that a couple's fecundability is a random variable Q that follows the beta distribution. They found that the number of cycles needed to get pregnant, *X*, follows a geometric distribution conditional on a couple's fecundity, Q=q. The marginal distribution of *X* was then obtained through:

$$\int_0^1 P(X = x|q) f(q)\, dq. \quad (1)$$

where they found that fecundability, *q,* equals, *1-p*. They found the marginal distribution to be:

$$\Pr(X = x) = E[q^{x-1}(1 - q)], \quad (2)$$

which can also be expressed in terms of moments:

$$\mathrm{E}(\mathrm{q}^{\mathrm{x}-1}) - \mathrm{E}(\mathrm{q}^{x}). \quad (3)$$

By assuming that *X*, the discrete number of menstrual cycles required to get pregnant, follows a geometric distribution they found that the probability that a randomly selected couple conceives at cycle *x*, conditional on their fecundability is

$$\Pr(X{=}x|p) = (1{-}p)^{x-1}p. \quad (4)$$

When they assumed that the fecundability follows a Beta distribution, the marginal probability of conception for a randomly selected couple at cycle *x* is

$$\Pr(\mathrm{X} = \mathrm{x}) \;=\; \mu \prod_{i=1}^{x-1} \frac{[1-\mu+(i-1)\theta]}{[1-(i-1)\theta]}, \quad (5)$$

where μ is mean parameter of the Beta distribution and θ is the shape parameter of the Beta distribution.

Additionally, Weinberg and Gladen (1986) found the conditional density of fecundability, *p*, for couples who have experienced *j* unsuccessful events cycles is

$$f(p|X>j) = p^{(\mu-\theta)/\theta}(1-p)^{[1-\mu+(j-1)\theta}. \quad (6)$$

Van Eekelen et. al (2020) also chose to model TTP using a beta-geometric model. However, Weinberg and Gladen's (1986) model is considered 'static', which according to van Eekelen, refers to models that do not predict over more than one time horizon, and will therefore, overestimate the probability of conception due to the failure of accounting for decreasing natural conception (Van Eekelen et al. 2017).

Eekelen et. al. (2017) introduced statistical methods to allow for predictions to be updated over time (McLernon et al. 2019). Therefore, unlike Weinberg and Gladen, who do not take into account differences in probability of conception between time

periods, the cumulative probability of pregnancy is calculated over prediction window, w. This results in the probability to conceive in $s+w$ cycles given that the couple did not conceive at or before s cycles shown as

$$\mathrm{P}(X \leq s + w | X > s) = 1 - \prod_{j=s}^{s+w-1} \frac{1-\mu+j\theta}{1+j\theta}, \quad (7)$$

where μ and θ are the respective mean and second shape parameter of the reparameterized beta distribution.

Zare et. al (2017) modeled TTP using a parametric frailty, or mixed effects survival model, to measure continuous time-to-first-pregnancy, TTFP. The two sources of variability in the study were within women attempting pregnancy in rural Iran, and within villages/families.

Along with the above parametric approaches for modeling TTP, the parametric approach has also been used for current duration analyses. Keiding et. al (2002) estimated TTP from a subset collected by the European Infertility and Subfecundity Study Group where the most recent time to pregnancy was recorded in what is known as a historically prospective cohort study, a study design where past data of women who were trying to conceive, regardless of outcome, is used to determine whether it ends in pregnancy attempts. An event was defined as one that ended in pregnancy and those that dropped out or stopped attempting pregnancy were censored. The subset Keiding selected was a cross sectional sample of women who were still attempting pregnancy after being followed up. By comparing the truth, which are those censored in the historically prospective cohort to the estimates from the current duration approach by

fitting a Pareto (λ + 1, μ) distribution and censoring at 15 months, the extent of direct likelihood inference can be determined.

Keiding (2011) used accelerated failure time regression on current duration observations of unprotected intercourse, or CDUI, taken from Slama et. al 2012, to build a TTP survival function using a generalized gamma distribution and a pareto distribution separately. At TTP $\geq$ 1 month, the survival functions relating to these underlying distributions were nearly identical as are their covariate frequency of sexual intercourse.

Although the parametric approach to analyzing fecundability is common, its restrictiveness and lack of flexibility is problematic when the distribution of TTP is unknown (McLain et. al 2014). It can be especially restrictive when TTP is continuous and can be more flexible if represented through piecewise models. (Zare et. al 2017).

2.2 NONPARAMETRIC METHODS

Although less common, many studies use non-parametric approaches in modeling TTP due to their flexibility as compared to the parametric approach. Various approaches and methods include the nonparametric Bayesian approach, Kaplan Meier estimator, and Aalen Johanssen estimator.

Although nonparametric approaches allow for flexibility with an unknown TTP distribution, an obstacle in non-parametric estimation is that the NPMLE, or nonparametric maximum likelihood estimation, of a decreasing TTP is inconsistent at 0 (Keiding et. al 2012). Therefore, an extra step is needed to counteract the inconsistency as seen in studies like Gasbarra et. al (2015). McLain et. al (2014) also mentioned that

methods measuring continuous TTP using the current duration approach is unable to integrate zeros.

Gnoth et. al (2003) estimated the cumulative probability of conception separately for including sub fertile and truly infertile couples, and for the subgroup of those who finally conceived using the Kaplan Meier method. His aim was to prevent overestimation of fertility among couples who drop out due to inability to get pregnant. Through the log rank test, the CPC for those who never conceived was significantly lower than those who finally did get pregnant

To account for the length-bias in current duration data and solve the inconsistency problem at a length of pregnancy attempt equal to 0 months of the NPMLE, Gasbarra et. al (2015) suggested a nonparametric Bayesian model and inference framework for determining TTP.

Friedrich et. al (2017) aimed to develop a Nelson Aalen estimator based off of the awareness by Meister and Schaefer (2008) that TTP estimation needs to account for competing risks in addition to left truncation.

2.3 SEMIPARAMETRIC METHODS

Semiparametric approaches are methods that include both parametric and nonparametric components. The advantage of semi-parametric methods over parametric and non-parametric methods is that they are flexible enough to accommodate the unknown TTP distribution in addition to providing a proper estimate of the survivor function of TTP (McLain et. al 2014).

McLain et. al 2014 proposed two semiparametric backward recurrence proportional hazards to estimate regression coefficients and survivor function, $\bar{F}$, when working with current duration data: one when the total duration of pregnancy attempt is discrete and the other when total duration is continuous but has been grouped according to known intervals. The discrete current duration observations in months, *Y*, have probability mass function:

$$g(y|\mathbf{Z}) = \mathrm{g}(0|\mathbf{Z})\exp\left\{-\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\sum_{j=0}^{y}\alpha_j\right\}, \quad (8)$$

where g represents the mass function of *Y*, α_j represents the baseline hazard rate for timepoint *j* and $\mathbf{Z}$ **and** $\boldsymbol{\beta}$ represent an p-dimensional vector of exposures and parameters, respectively. Grouped current durations, which are created from continuous *T,* have probability mass function

$$g(y) = \frac{1}{\mu_T}\int_y^{y+1}\bar{F}(u)du \text{ for } y = 1,2,\ldots, \quad (9)$$

where μ_T = expected time to the end of an attempt, whether it ends in a pregnancy or not.

A piecewise backward recurrence cox model was also proposed to account for digit preference that overestimates length of attempts for 6, 12, and 24 months. The piecewise estimator of *T's* survivor function is

$$\exp\left[-\exp(\boldsymbol{\beta}^T\boldsymbol{Z})\sum_{\{j:t_{j-1}<y\}}\gamma_j\{(y \wedge t_j) - t_{j-1}\}\right], \quad (10)$$

where t_j represents time at respective chosen knots and y represents the reported length of pregnancy attempt.

Oberhofer and Reichsthaler (2004) estimated a semi-parametric approach using the local likelihood method. Their goal was to model the childbearing behavior of the population while taking into account different census dates and mother's age. For fitting over age, an exponential function and parametric spline function were used to model the response and predictor respectively. When fitting over time, a non-parametric approach was the optimal method as it allowed for the most flexibility with regards to parameters' movements.

Eekelen et. al (2020) used landmarking with the Cox proportional hazards model to derive dynamic predictions. The three approaches van Eekelen et. al (2020) used to build a prediction model landmark were: fitting a Cox model to each landmark dataset, fitting a Cox model to each strata, and fitting a Cox model that uses linear and quadratic covariates to represent landmark number. When comparing to the parametric beta geometric model for deriving dynamic predictions, landmarking was found to be a better model for smaller sample size.

CHAPTER 3: ESTIMATION METHODS

Section 3.1 discusses two current duration studies that do not correct for measurement error in the outcome which pave the way to the methods used to correct for measurement error in the outcome. Section 3.2 gives existing literature on the SIMEX method. Section 3.3 examines the type of data, distribution, and models used to correct for measurement error in the outcome. Sections 3.4 and 3.5 discuss the specific methods used to create naïve survival curves that are then compared to the true current duration survival to determine bias and nevertheless, their effectiveness at correcting for measurement error in the outcome.

3.1 STUDIES THAT DO NOT CORRECT FOR MEASUREMENT ERROR IN THE OUTCOME

Previous methods of determining the TTP distribution do not correct for measurement error in the outcome. This includes McLain et. al (2014) who analyzed properties of their proposed semiparametric Cox model and piecewise constant models. Observations denoted as Y represent observed lengths of pregnancy attempts, also referred to as current durations. The total duration of pregnancy attempts, T, is represented by the lesser value between TTP, denoted by X, and end of pregnancy attempt without becoming pregnant, U. Their goal was to estimate the distribution of T.

Given that Y and T are discrete random variables, McLain et. al (2014) found that the survivor function of T, given that time is discrete, is represented by

$$\bar{F}_{\mathrm{T}}(\mathrm{t}|\mathbf{Z}) = \exp\{-\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}) \sum_{j=0}^{t} \alpha_j\},$$

where $\alpha_j \geq 0$, β is a vector of parameter estimates, and $\mathbf{Z}$ is a vector of the different exposures. This results in *Y* having probability mass function

$$g(y|\boldsymbol{Z}) = g(0|\boldsymbol{Z})\exp\{-\exp(\boldsymbol{\beta}^T\boldsymbol{Z}) \sum_{j=0}^{y} \alpha_j\}.$$

The piecewise constant baseline model creates a joint to control for digit preference. For this approach, McLain et. al found that the probability mass function of *Y* is

$$g_P(y|\mathbf{Z}) = g_P(0|\boldsymbol{Z}) \exp\left[-\exp(\boldsymbol{\beta}^T\boldsymbol{Z}) \sum_{\{j:t_{j-1}<y\}} \gamma_j\{(y \wedge t_j) - t_{j-1}\}\right],$$

where $g_P(0|\boldsymbol{Z}) = \left[\sum_{y=0}^{\infty} exp\{-exp(\boldsymbol{\beta}^T\boldsymbol{Z}) \sum_{\boldsymbol{j=0}}^{\boldsymbol{y}} \boldsymbol{\alpha_j}\}\right]^{-1}$. Here, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha_j}$ in the above models can be estimated using maximum likelihood. To test the probability mass functions of *Y* with and without accounting for digit preference, simulations studies were performed. Results showed that both the piecewise and semiparametric models showed minimal bias for the estimates of $\boldsymbol{\beta}$**,** but the piecewise method had much less bias in the estimates of $\bar{F}_{\mathrm{T}}(\mathrm{t}|\mathbf{Z})$.

Keiding et. al (2002) proposed a parametric approach to model current duration data, by specifying that the observed current length of pregnancy attempt, *Y,* follows Pareto(λ, μ) distribution with survival function $(1 + \mu y)^{-\lambda-1}$. The Pareto has a conjugacy between *Y* and *T,* as the survival function of total duration of pregnancy attempts, *T*, also follows a Pareto(1+λ, μ) distribution.

3.2 PREVIOUS LITERATURE ON THE SIMEX METHOD

A method used to correct for measurement error is the simulation extrapolation, or SIMEX, method. SIMEX is a method that was originally proposed to correct for the impact of measurement error in the covariates when estimating regression coefficients (Stefanski and Cook 1995). The popularity of SIMEX was driven by the generality and ability to work with different types of measurement error problems. Oh et. al (2018) recently proposed using SIMEX to account for measurement error in survival outcomes through a Weibull accelerated failure model

$$\log(Y') = \alpha_0 + \beta X + \sigma\epsilon = \log(Y) + v, \quad (11)$$

where *Y* is true length of pregnancy attempt, *Y'* is the observed length of pregnancy attempt (*Y* with error), and *v* is the measurement error with mean 0 and known variance, σ_v^2. $\log(Y^*) = \log(Y') + \sqrt{\lambda}v$ where v is simulated to have variance σ_v^2. Since additional measurement error is needed in addition to the default measurement error, σ_v^2, the total measurement error for $\log(Y^*)$ is

$$\sigma_v^2 + \lambda\sigma_v^2 = (1 + \lambda)\sigma_v^2,$$

where $\lambda > 0$. Through simulation, the estimated β parameter value at each of λ is found and is used to fit a linear model which is extrapolated to $\lambda = -1$ for a total of B iterations to get the total measurement error shown above.

Given that σ_v^2 is constant, simulations are run to determine bias at different values of true β. The bias is the difference between the coefficients from the model that corrects for total measurement error using the SIMEX method and the coefficients from the model with true time to pregnancy attempt. By also comparing the SIMEX method

to the naïve methods introduced in the previous section, the performances between methods can be determined. Oh et. al found that the SIMEX method maintained a lower MSE than the naïve method in a variety of settings.

3.3 METHOD OVERVIEW

Due to the suitability of the parametric approach in modelling current duration data as discussed in Keiding (2011), the methods of correcting for measurement error in the outcome will be applied to parametric models built from current duration data. Nevertheless, current duration data, *Y*, is generated from the Weibull distribution. Because a parametric approach is being used to model the current duration data, the true TTP function is represented by 1-$F(t)$, where $F(t)$ represents Weibull's cumulative distribution function:

$$e^{(t/\delta)^{\beta}} \quad for\ t > 0,$$

and $\beta > 0$ is a shape parameter, $\delta > 0$ is a scale parameter, and *t* is the true length of a pregnancy attempt.

McLain et. al (2014) found that failure to account for digit preference can overestimate the percentage of women not achieving pregnancy at 12 and 24 months. Therefore, the estimated survival curves for each method discussed are created from piecewise constant baseline models to account for digit preference. The piecewise constant models for each method are fit with 6 and 10 discrete and unique knots respectively. These knots are generated from the quantile function of the Weibull distribution with probabilities $\frac{1:K}{(K+1)}$, where K represents the number of knots.

3.4 NAÏVE METHODS

The first method, called the untransformed method, estimates the survival function of length of pregnancy attempt with respect to time from zero to 50 months. No transformations are made to *Y*, which is used to generate the piecewise model. Additionally, no modifications are made to the time variable when creating the survival function. Nevertheless, the survival function created from the piecewise model is represented by:

$$\hat{S}(t).$$

The second naïve method, known as the shifted method assumes that on average, couples find out about their pregnancy μ_v months after they actually become pregnant. Here, μ_v is the mean of the measurement error variance, i.e., $\mu_v = E(e^{\nu})$. We assume that the measurement error follow a log-normal distribution where μ_v is given by

$$exp(\sigma_v^2/2),$$

where σ_v^2 is the measurement error variance.

Shown in Equation 1 is the expected true length of pregnancy where *Y* denotes the true current length of pregnancy in months and *Y'* is the reported current length of pregnancy in months. The shifted method attempts to correct for the measurement error through shifting the observed *Y* values. For example, if on average couples find out that their pregnant 1 month after actually becoming pregnant, their reported length of pregnancy is actually 1 month greater than the true length of time to pregnancy.

$$\mathbf{E}(Y) = \mathbf{E}(Y') - \mu_v. \qquad (1)$$

The mean probability of length of pregnancy attempt at each time point is divided by the probability of length of pregnancy attempt at the mean measurement error variance, that is,

$$\hat{S}(t) = \frac{\hat{S}(t+\mu_v)}{\hat{\mathbf{S}}(\mu_v)}. \qquad (2)$$

The numerator represents the survival of the reported time to pregnancy at each time point while the denominator represents the adjusted baseline survival according to how many months off the average subject reports compared to the truth.

Unlike the untransformed and shifted methods, which build survival functions based off of unmodified *Y*, the transformed methods builds a piecewise model off of Y/μ_v, where μ_v is defined previously and is assumed to be known. However, like the untransformed method, no modifications are done to the time variable when creating the survival function. Therefore, the survival function created from the transformed method is simply the estimated survival function from the transformed data.

3.5 SIMEX METHODS

In the SIMEX method the current duration survival model is created using a modified version of *Y* with associated lognormally distributed measurement error, ν with density function

$$\frac{1}{\sqrt{2\pi}\sigma x} exp(\frac{(-ln(x)-\mu)^2}{2\sigma^2}),$$

where μ = 0 and $\sigma^2 = \sigma_v^2$. Specifically, $\log(Y^*) = \log(Y') + \sqrt{\lambda}\nu$ is generated for a total of B times per value of λ. In our simulation we used $B = 10$ and $\lambda =$ $(0.1,0.2, \ldots ,0.9,1,1.25,1.5,1.75, 2, 3)$. For each simulated dataset the piecewise constant

model survival model is fitted with no covariates which results in (K+1) α values, where K is the number of knots. This results in a Bx(K+1) matrix of estimated α values for each value of λ (we used 15 values of lambda).

The α values along with their corresponding λ are used to create a linear model as follows:

$$\alpha_{kb} = \beta_0 + \beta_1\lambda + \beta_2\lambda^2 + \varepsilon \qquad \text{where } k=1,\ldots,K+1 \text{ and } b=1,\ldots,B.$$

The linear model is extrapolated to predict the α_k values at $\lambda = -1$ for all k=1,…,K. Note that $\alpha_k > 0$ which could cause issues when fitting the linear model as predicted values can be negative. As a result, we also tested fitting the linear model based on log transformed values. That is, we fit the above linear model to $\log(\alpha_k)$. The predicted $\log(\alpha_k)$ were then back transformed to get the estimated α_k's. This we refer to as the log transformed SIMEX method, or log SIMEX method throughout the paper. For the untransformed SIMEX method, otherwise known as the regular SIMEX method throughout the paper, when a predicted α value was less than or equal to 0 we kept them at 0. For both methods, the resulting α predictions are used to estimate the survival function.

CHAPTER 4: SIMULATION STUDY

4.1 BIAS FUNCTION AND SIMULATIONS

To test the different methods R was used. Each method is tested over four different $\sigma_{\upsilon}{}^2$ values, 0.25, 0.5, 0.75, and 1, two different combinations of Weibull parameter, and two different knot values. The sample size used when generating *Y* is 200.

Within R, we create a bias function which depends on the measurement error variance, Weibull parameters, knot values, λ values, time vector, and B, which represents iterations per λ with regards to the SIMEX methods. This bias function is executed for 1000 iterations for the transformed, shifted, untransformed, log SIMEX, and regular SIMEX given each possible combination of knot value, Weibull parameters, and measurement error variance. The average of the iterations at each time point is taken and used to create plots to compare with the true survival curve and between methods. The standard deviation at each time point is also taken to create individual plots with 95% confidence intervals and determine mean squared error.

4.2 SIMULATION RESULTS

Figures 4.1 and 4.2 depict the comparison of the survival curves from each method fit with 6 knots for each measurement error variance value in increasing order from left to right with shape and scale combinations of 0.5 and 6, and 2 and 10

respectively (see A.1 and A.2 for bias values). Figures 4.3 and 4.4 are identical to figures 4.1 and 4.2 respectively but use 10 knots to build the survival curves. Tables 4.1-4.4 represent the figures' respective bias values at months 3, 6, 12, 24, and 36 (see A.3 and A.4 for bias values).

The results from the graphs show that the difference in using 6 knots versus 10 knots is trivial to all methods except the log SIMEX method. Comparing Figures 4.1 to 4.2 and 4.3 to 4.4, it is evident that fitting the piecewise model created from the log SIMEX method with 6 knots results in not only a smoother fit, but also less overall bias for both sets of Weibull parameters at each measurement error variance value. Nevertheless, further analysis is done with survival curves fit with 6 knots.

When comparing the survival curves that depict the SIMEX method versus the log SIMEX method shown in Figure 4.1, the regular SIMEX method results in more bias at errors of 0.75 and 1 but less bias at 0.25 and 0.5 than its log counterpart. When fitting the survival curves to a Weibull distribution of shape=2 and scale=10 as shown in Figure 4.3, the log SIMEX version performs better overall at each measurement error variance value for each set of knots.

When comparing the survival curves with Weibull shape and scale of 0.5 and 6 respectively, a measurement error variance values of 0.75 and 1, the transformed method for correcting for measurement error exhibits the most bias. At a measurement error variance value of 0.25, the largest biases are seen in both the shifted and the log SIMEX methods. The SIMEX method exhibits the least amount of bias. At a measurement error variance value of 0.5, the log SIMEX exhibits the most amount of

bias while the untransformed method, although very comparable to the bias of the SIMEX method, exhibits the least amount of bias. The untransformed and log SIMEX methods exhibit the least amounts of bias at a value of 0.75. At a measurement error value of 1, the log SIMEX method and shifted method seem to exhibit the least amounts of bias.

When looking at the survival curves with the Weibull shape and scale of 2 and 10 respectively in Figure 4.3, the transformed method exhibits the largest amount of overall bias at each measurement error variance value. The untransformed, shifted, and regular SIMEX (nonlog) methods survival curves are almost identical to each other at each measurement error variance. Additionally, these curves change very little with an increase in measurement error variance. With regards to bias, the log SIMEX method seems to result in the least amount of overall bias at 0.25, 0.5, and 0.75 before around 12 months. However, after 12 months, the survival curves from this method resulted in either similar or higher bias values from the untransformed, shifted, and regular SIMEX methods. At a measurement error value of 1, log SIMEX survival curve is almost identical to the untransformed, shifted, and regular SIMEX methods.

When comparing the performance of correcting for measurement error between the different Weibull distributions, the shape and scale of 0.5 and 6 result in less overall bias across all methods, except the log SIMEX, and measurement error variances before approximately 20 months. After 20 months, the extent of bias is similar between the Weibull distributions.

Because there were many instances where determining the best method was difficult, individual graphs with their 95% confidence intervals were generated to further compare variance among methods. Figure 4.5-4.8 compare the individual graphs at all four measurement error variances using the Weibull distribution with parameters 0.5 and 6. Figures 4.9-4.12 compare the graphs using Weibull distribution with parameters 2 and 10. The untransformed, shifted, transformed, and regular SIMEX methods had similar confidence interval widths between each other with regards to measurement error variance and across both distributions. The log SIMEX method had similar confidence intervals width as the aforementioned methods across the distribution with shape and scale of 2 and 10 but a somewhat higher interval across the distribution with shape and scale of 0.5 and 6, especially across 0.25 and 0.5. In fact, the measurement error variances that resulted in the highest bias when looking at this method correspond to the measurement error variances, 0.25 and 0.5, that result in the widest confidence intervals. This, nevertheless, results in a relatively higher MSE (see Table B.1). Coincidentally, the log SIMEX method had the lowest MSE values across measurement error variances when looking at distribution of shape=2 and scale=10 (see Table B.2). Contrarily, the somewhat higher MSE values seen from the transformed method are a result of the high bias discussed previously because their confidence intervals are relatively narrow as shown in Figures 4.5-4.12.

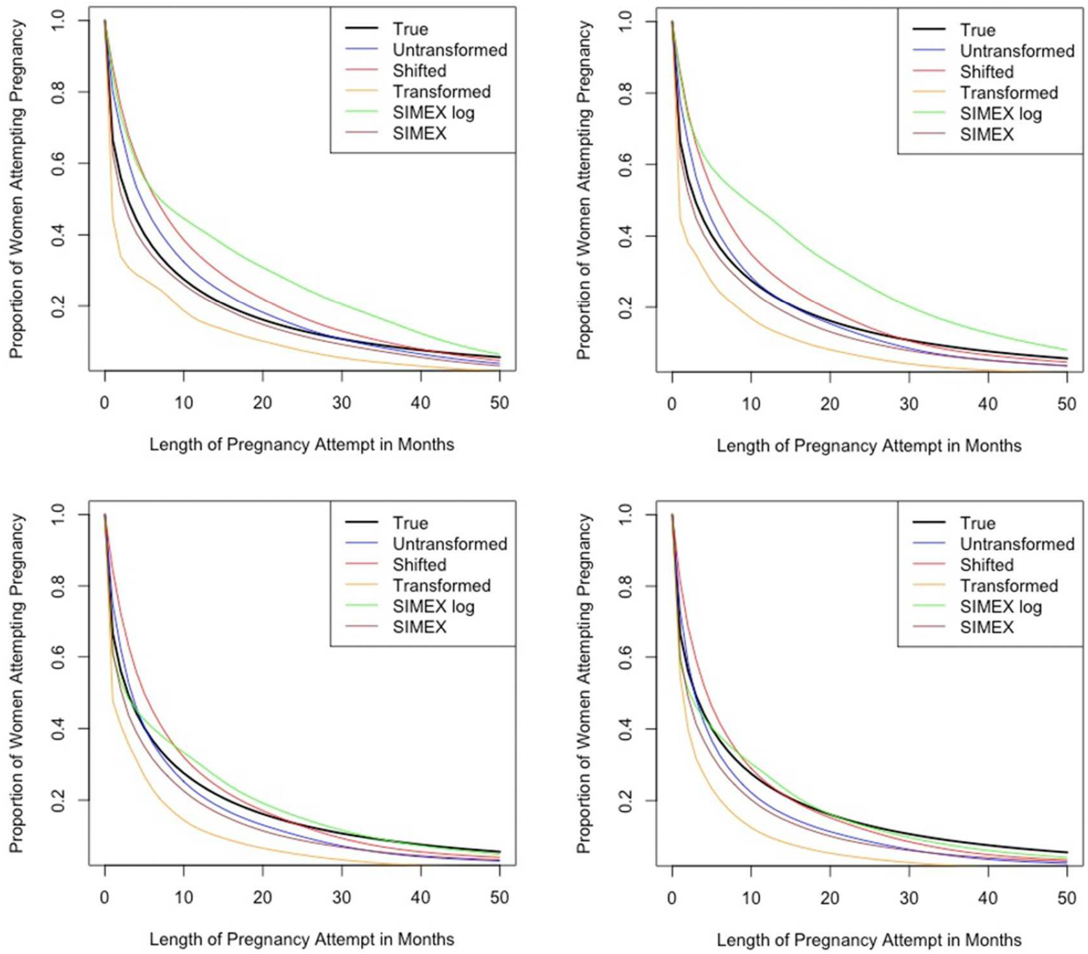

Figure 4.1: Comparison of survival curves with measurement error variance of 0.25 (top left), 0.5 (top right), 0.75 (bottom left), and 1 (bottom right) fit with 6 knots from the Weibull distribution with scale=0.5 and shape=6.

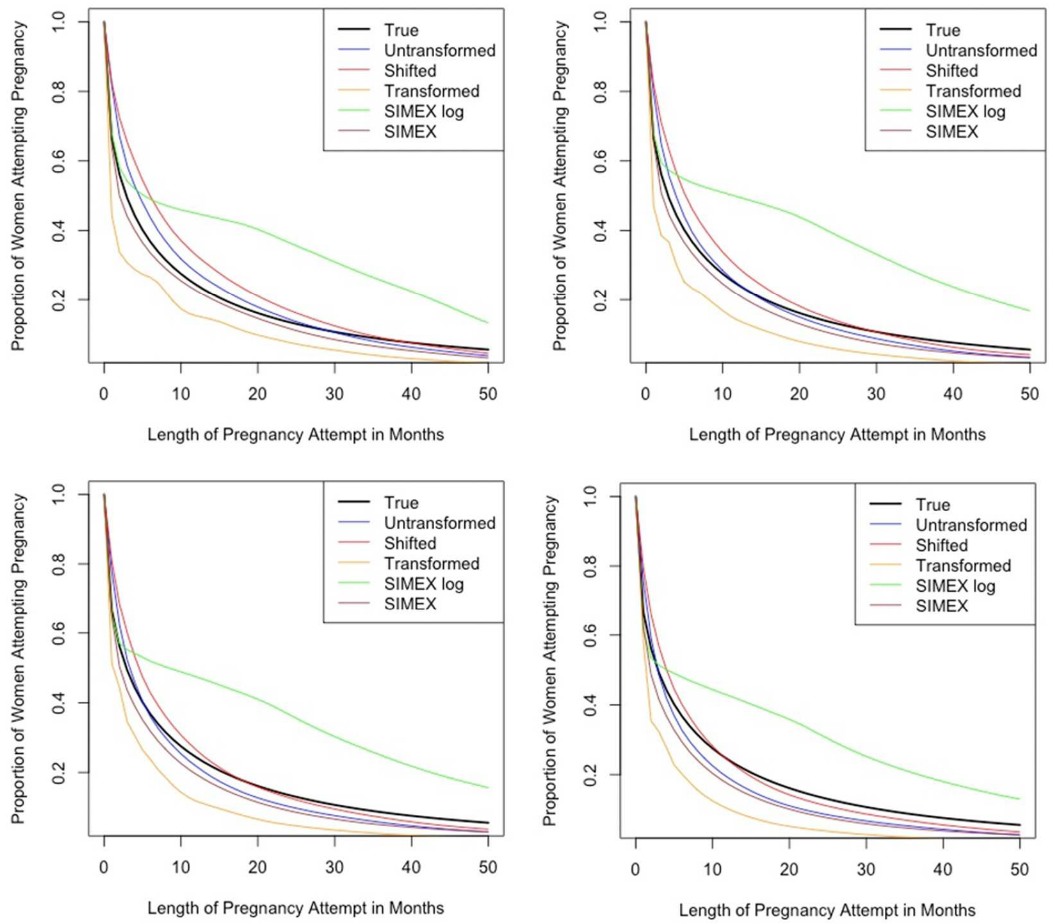

Figure 4.2: Comparison of survival curves with measurement error variance of 0.25 (top left), 0.5 (top right), 0.75 (bottom left), and 1 (bottom right) fit with 10 knots from the Weibull distribution with scale=0.5 and shape=6.

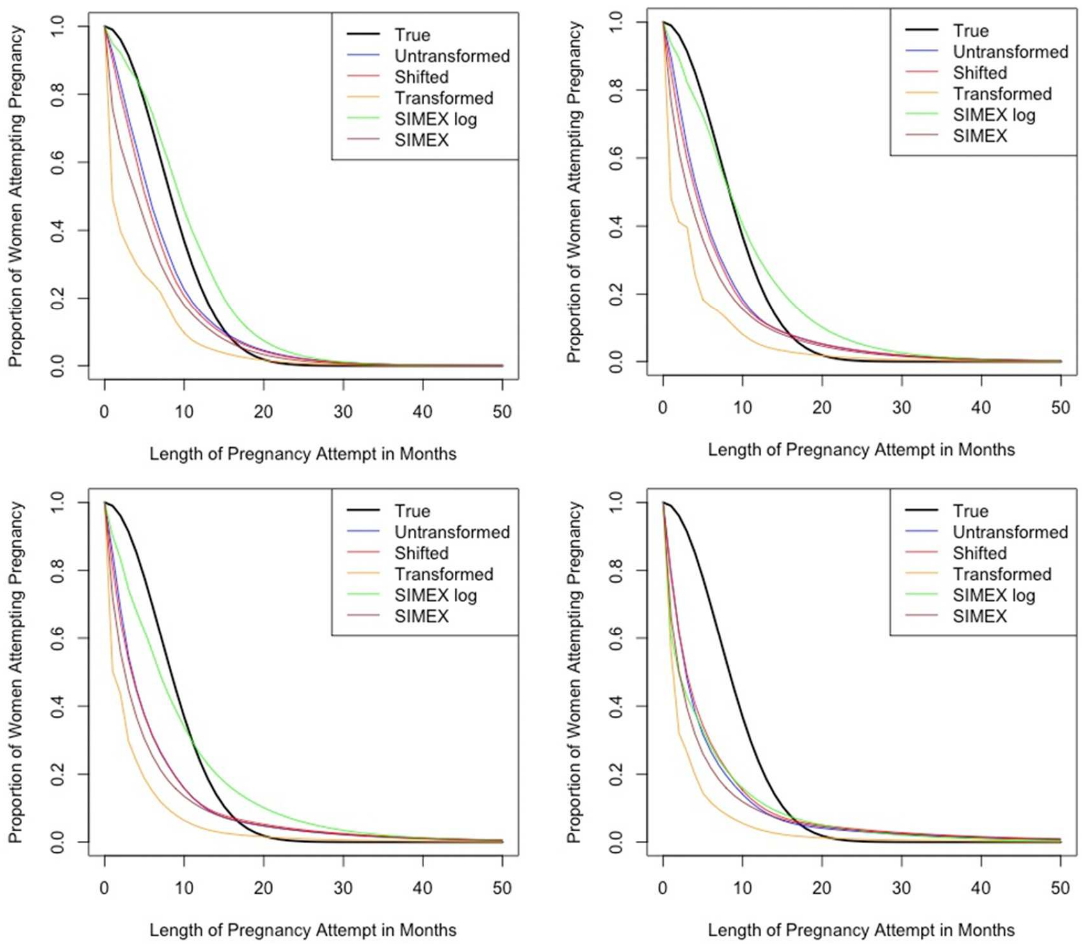

Figure 4.3: Comparison of survival curves with measurement error variance of 0.25 (top left), 0.5 (top right), 0.75 (bottom left), and 1 (bottom right) fit with 6 knots from the Weibull distribution with scale=2 and shape=10..

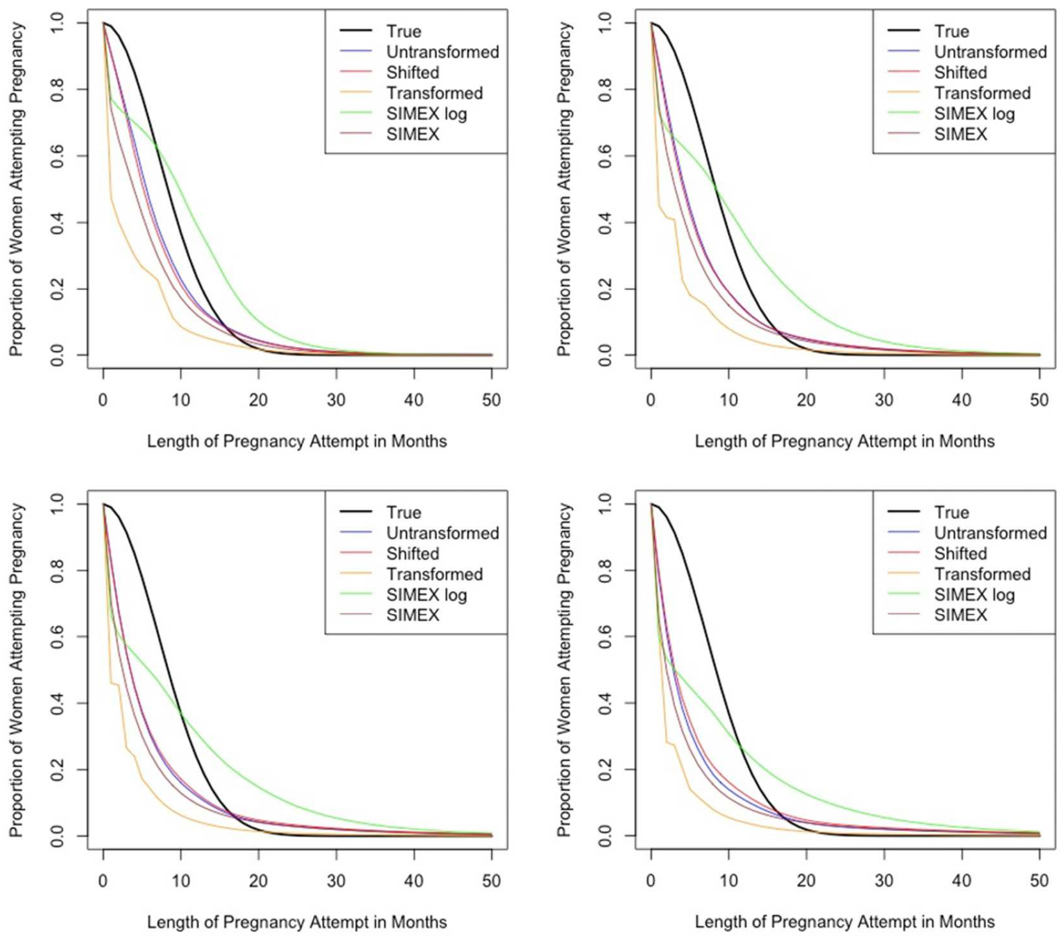

Figure 4.4: Comparison of survival curves with measurement error variance of 0.25 (top left), 0.5 (top right), 0.75 (bottom left), and 1 (bottom right) fit with 10 knots from the Weibull distribution with scale=2 and shape=10.

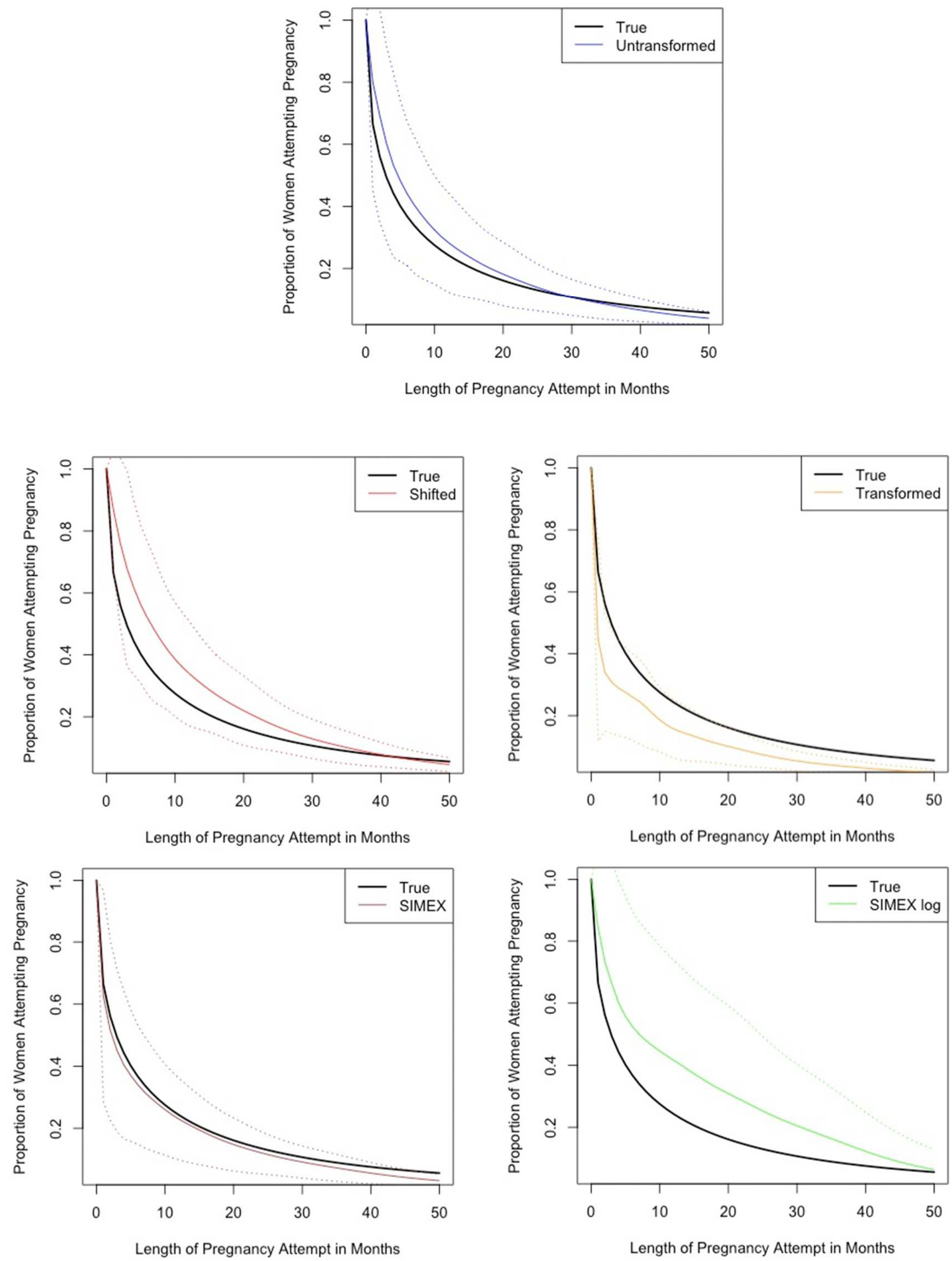

Figure 4.5: Comparison of survival curves with 95% confidence intervals at measurement error variance of 0.25 from distribution with shape and scale of 0.5 and 6.

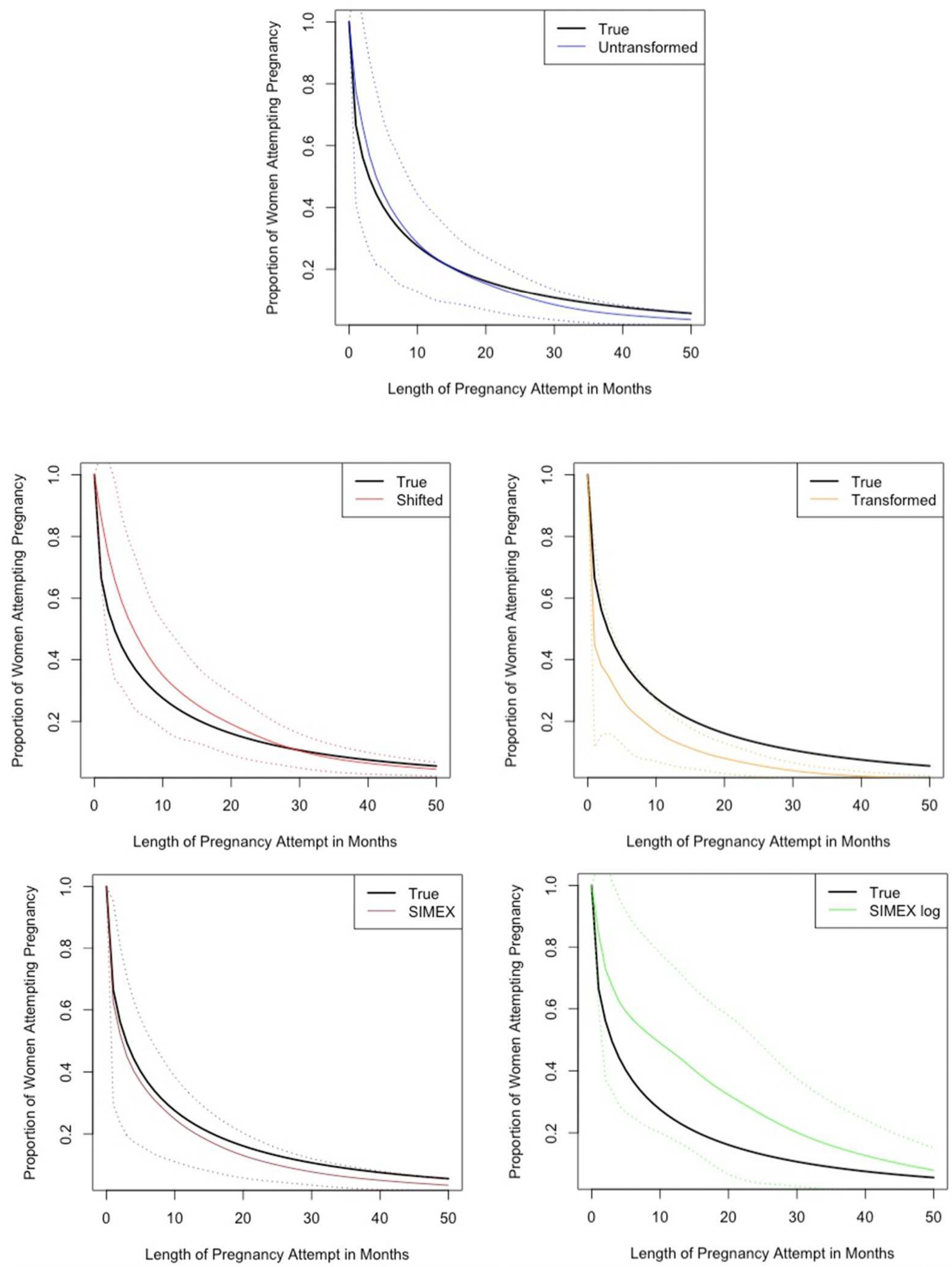

Figure 4.6: Comparison of survival curves with 95% confidence intervals at measurement error variance of 0.5 from distribution with shape and scale of 0.5 and 6.

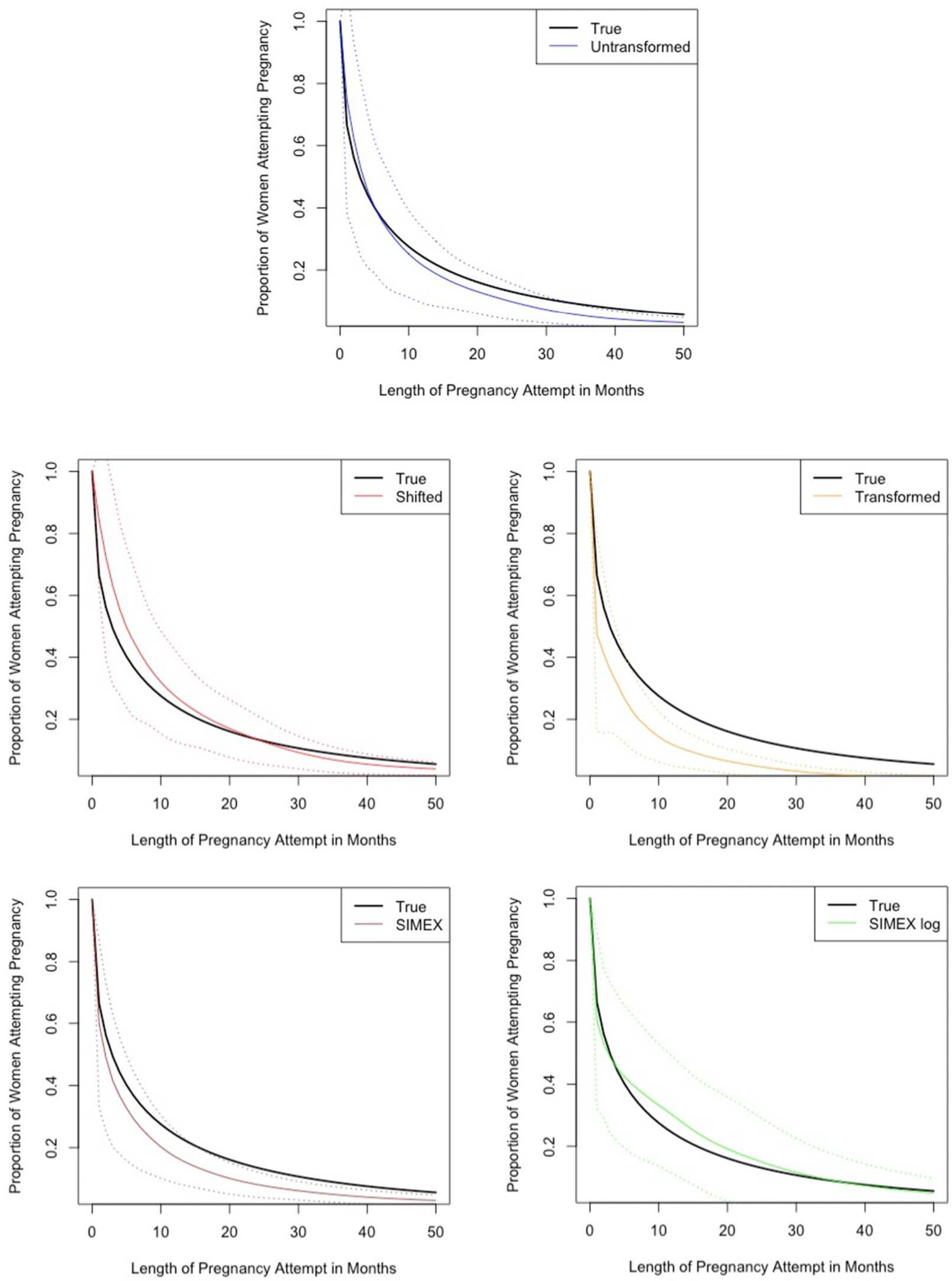

Figure 4.7: Comparison of survival curves with 95% confidence intervals at measurement error variance of 0.75 from distribution with shape and scale of 0.5 and 6.

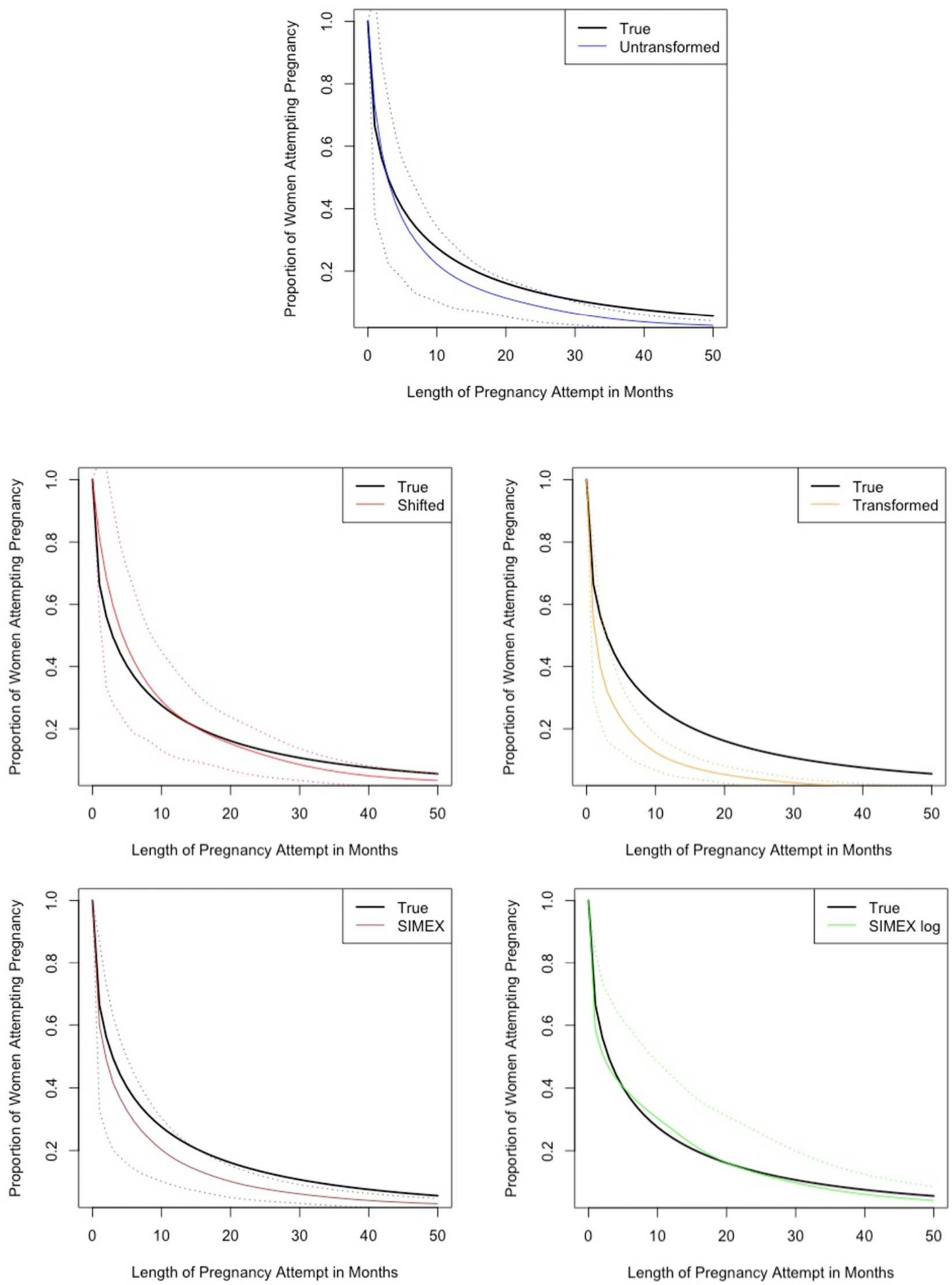

Figure 4.8: Comparison of survival curves with 95% confidence intervals at measurement error variance of 1 from distribution with shape and scale of 0.5 and 6.

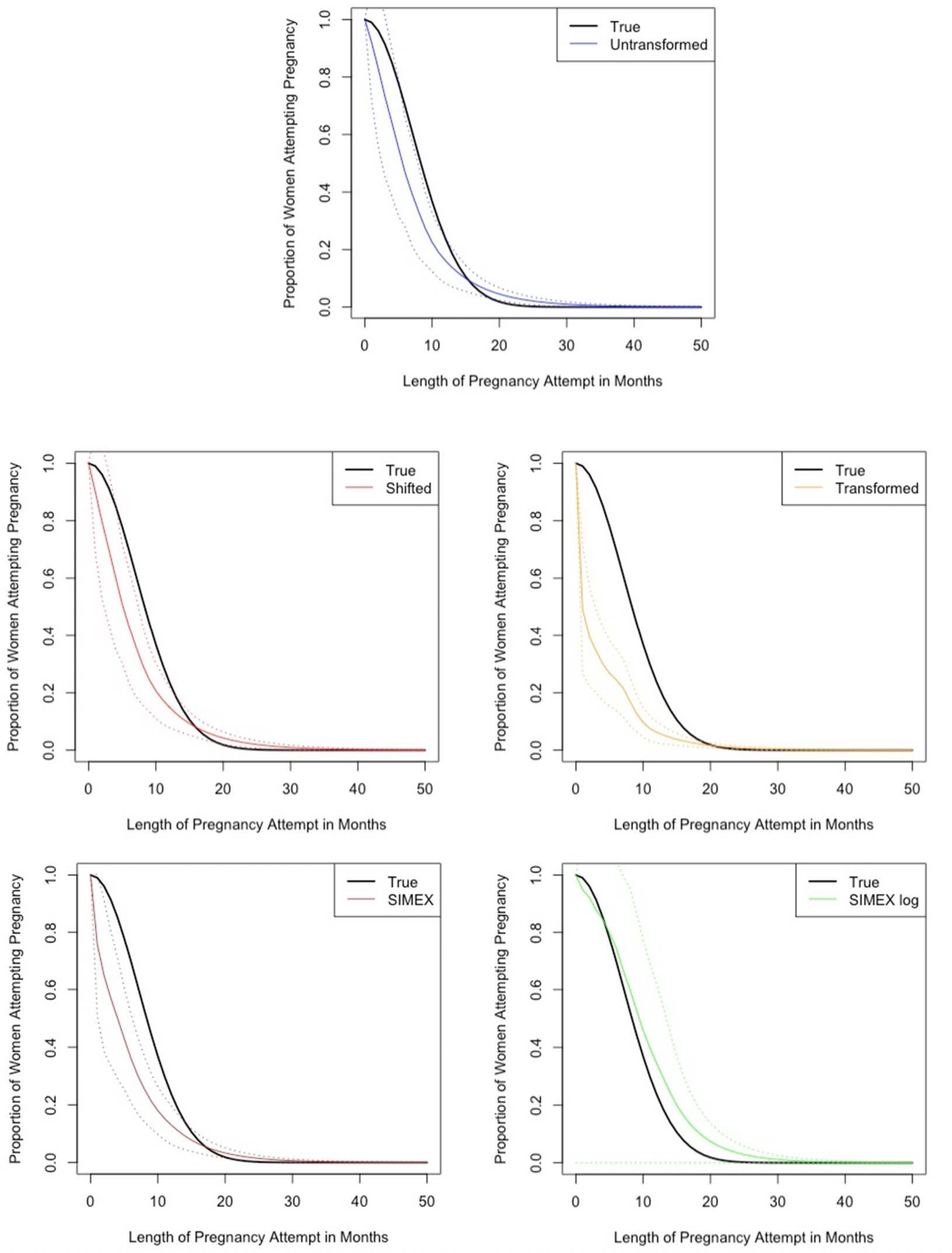

Figure 4.9: Comparison of survival curves with 95% confidence intervals at measurement error variance of 0.25 from distribution with shape and scale of 2 and 10.

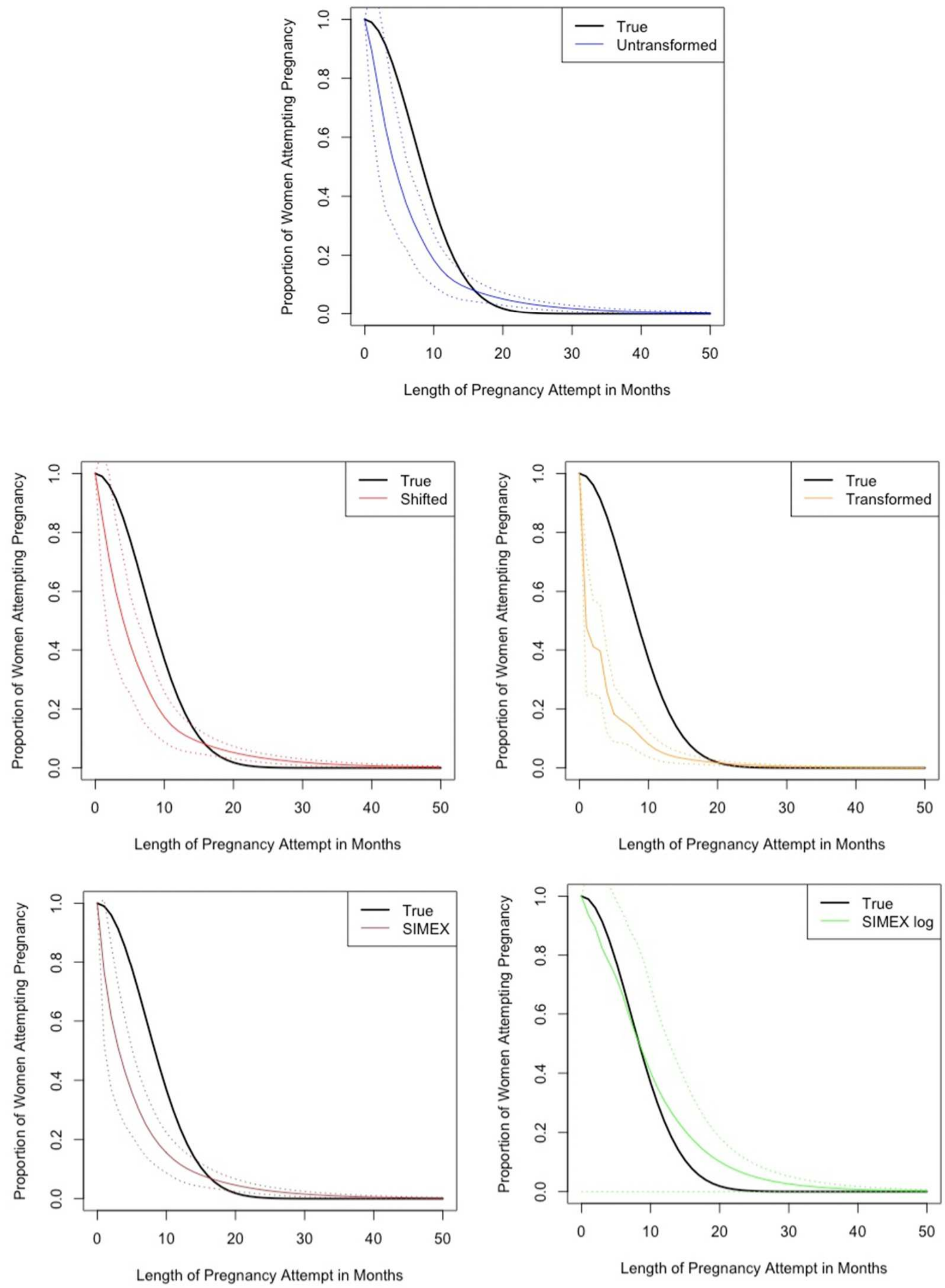

Figure 4.10: Comparison of survival curves with 95% confidence intervals at measurement error variance of 0.5 from distribution with shape and scale of 2 and 10.

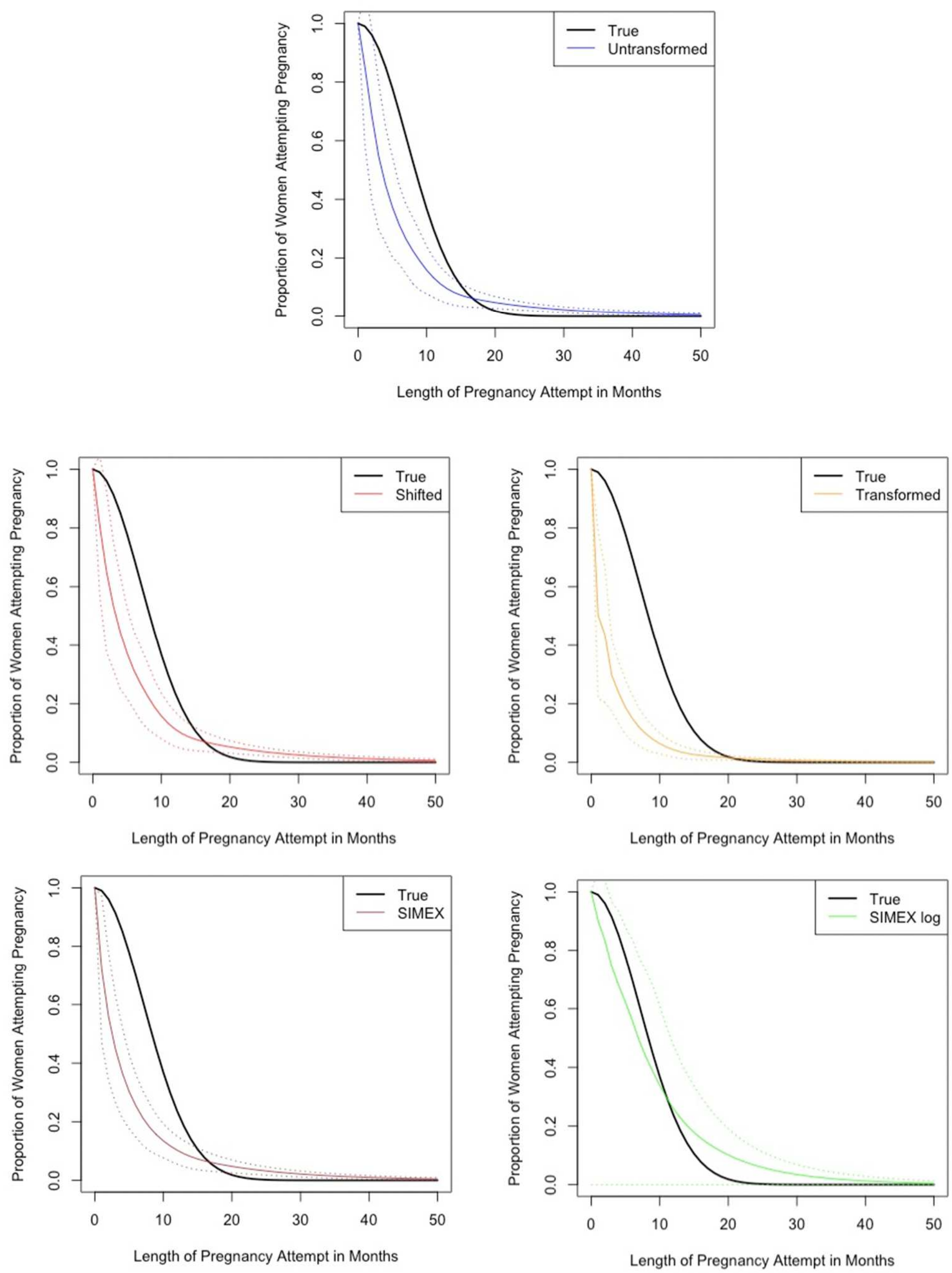

Figure 4.11: Comparison of survival curves with 95% confidence intervals at measurement error variance of 0.75 from distribution with shape and scale of 2 and 10.

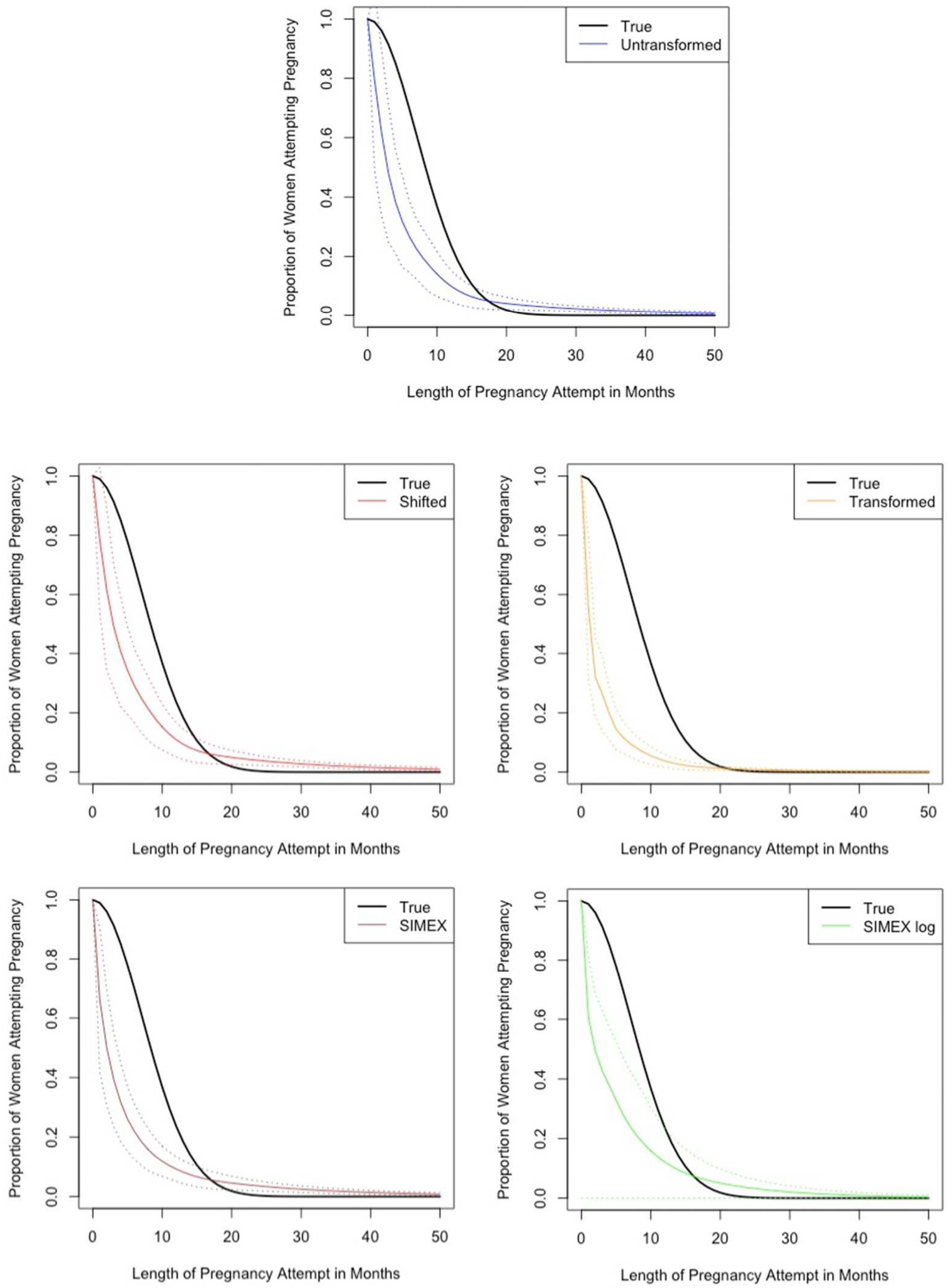

Figure 4.12: Comparison of survival curves with 95% confidence intervals at measurement error variance of 1 from distribution with shape and scale of 2 and 10.

CHAPTER 5: DISCUSSION

Correcting for measurement error in the outcome is not common in most survival analysis studies, let alone time to pregnancy studies. Therefore, the purpose of this thesis was to compare methods used to correct for measurement error in the outcome when dealing with survival curves generated from current duration parametric piecewise models. The methods included the untransformed, shifted, transformed, and the two SIMEX methods; one which uses the log alpha values in the creation of the piecewise model and the other uses the untransformed alpha values. Because current duration data was being modelled, the curves represent the probability of getting pregnant at each month of attempting. Although this approach includes subfertile, infertile, and accidental pregnancies, it cannot distinguish between an attempt ending due to pregnancy or just giving up.

Using 6 versus 10 knots only made an impact on the log SIMEX method where it appeared that 6 knots fit better than 10 knots. When comparing methods with regards to bias, the transformed method of correcting for measurement error in the outcome was almost always the worst with regards to bias across both distributions and all measurement error variances. The log SIMEX method was worse in only two instances. Therefore, transforming *Y* by dividing it by μ_v when creating piecewise models using current duration data is not optimal. Although the worst method was distinctive, the

best method was not. To further compare the methods, their respective curves with the 95% confidence intervals are generated to evaluate for precision. All the methods do not differ much in confidence interval width except for the log SIMEX method which result in a wide upper interval. It is kept in mind that the lower interval was, in many cases, partly or fully at or below zero due to the nonsymmetric distribution. Because the log SIMEX method had the highest and lower MSE values depending on the Weibull distribution, this method is inconsistent for correcting for measurement error in the outcome. The untransformed, shifted, transformed, regular SIMEX methods do not differ much between each other with regards to precision.

Studying the different methods to correct for measurement error in the outcome when dealing with current duration data involves determining the extent of bias in addition to variance. Bias is associated with accuracy of the model while variance or deviation is associated with precision. Both are crucial components to avoid under or overfitting a model. By just looking at bias, the log SIMEX method could be considered a good method in correcting for error. However, the confidence intervals produced are wide in some cases. The transformed method has high bias overall but relatively narrow confidence intervals. Both methods contain high MSE at many measurement error variances and are therefore, not the best in correcting for measurement error in the outcome. In addition, neither method results in consistently smooth curves when being fit to 6 knots and 10 knots.

Because extent of bias and MSE depend on the Weibull parameters and measurement error variances, we cannot say that there is an optimal method for

correcting for measurement error in the outcome. However, because the untransformed method never had the greatest bias, greatest MSE, was most consistent with regards to both of these measurement, and always resulted in smooth survival curves, this method could be used to correct for measurement error in the outcome when knowing little to nothing about the measurement error variance and underlying distribution.

Although we did not find an obviously optimal method, other methods could be explored to determine whether one could unarguably be the best at correcting for measurement error in the outcome.

REFERENCES

Baird, D. D. (2013), “Chapter 14 - Women’s Fecundability and Factors Affecting It,” in *Women and Health (Second Edition)*, eds. M. B. Goldman, R. Troisi, and K. M. Rexrode, Academic Press, pp. 193–207. https://doi.org/10.1016/B978-0-12-384978-6.00014-5.

Bonde, J. P., Joffe, M., Sallmén, M., Kristensen, P., Olsen, J., Roeleveld, N., and Wilcox, A. (2006), “Validity issues relating to time-to-pregnancy studies of fertility,” *Epidemiology*, LWW, 17, 347–349.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement error in nonlinear models: a modern perspective*, Chapman and Hall/CRC.

Gasbarra, D., Arjas, E., Vehtari, A., Slama, R., and Keiding, N. (2015), “The current duration design for estimating the time to pregnancy distribution: a nonparametric Bayesian perspective,” *Lifetime data analysis*, Springer, 21, 594–625.

Gnoth, C., Godehardt, D., Godehardt, E., Frank-Herrmann, P., and Freundl, G. (2003), “Time to pregnancy: results of the German prospective study and impact on the management of infertility,” *Human reproduction*, Oxford University Press, 18, 1959–1966.

Hassan, M. A., and Killick, S. R. (2004), “Negative lifestyle is associated with a significant reduction in fecundity,” *Fertility and sterility*, Elsevier, 81, 384–392

“Infertility | Reproductive Health | CDC” (2021), Available athttps://www.cdc.gov/reproductivehealth/infertility/index.htm.

Jacobson, M. H., Chin, H. B., Mertens, A. C., Spencer, J. B., Fothergill, A., and Howards, P. P. (2018), “‘Research on infertility: definition makes a difference’ revisited,” *American journal of epidemiology*, Oxford University Press, 187, 337–34

Jensen, T. K., Slama, R., Ducot, B., Suominen, J., Cawood, E. H. H., Andersen, A. G., Eustache, F., Irvine, S., Auger, S., and Jouannet, P. (2001), “Regional differences in waiting time to pregnancy among fertile couples from four European cities,” *Human reproduction*, Oxford University Press, 16, 2697–2704.

Joffe, M., Key, J., Best, N., Keiding, N., Scheike, T., and Jensen, T. K. (2005), "Studying time to pregnancy by use of a retrospective design," *American journal of epidemiology*, Oxford University Press, 162, 115–124.

Juul, S., Keiding, N., and Tvede, M. (2000), "Retrospectively sampled time-to-pregnancy data may make age-decreasing fecundity look increasing," *Epidemiology*, JSTOR, 717–719.

Keiding, N. (2011), "The current duration (backward recurrence time) approach to estimating the distribution of time to pregnancy."

Keiding, N., HØJBJERG HANSEN, O. K., Sørensen, D. N., and Slama, R. (2012), "The current duration approach to estimating time to pregnancy," *Scandinavian Journal of Statistics*, Wiley Online Library, 39, 185–204.

Keiding, N., Kvist, K., Hartvig, H., Tvede, M., and Juul, S. (2002), "Estimating time to pregnancy from current durations in a cross-sectional sample," *Biostatistics*, Oxford University Press, 3, 565–578.

McLain, A. C., Sundaram, R., Thoma, M., and Buck Louis, G. M. (2014), "Semiparametric modeling of grouped current duration data with preferential reporting," *Statistics in medicine*, Wiley Online Library, 33, 3961–3972.

McLernon, D. J., Lee, A. J., Maheshwari, A., Van Eekelen, R., van Geloven, N., Putter, H., Eijkemans, M. J., van der Steeg, J. W., Van Der Veen, F., and Steyerberg, E. W. (2019), "Predicting the chances of having a baby with or without treatment at different time points in couples with unexplained subfertility," *Human Reproduction*, Oxford University Press, 34, 1126–1138.

Meister, R., and Schaefer, C. (2008), "Statistical methods for estimating the probability of spontaneous abortion in observational studies—analyzing pregnancies exposed to coumarin derivatives," *Reproductive Toxicology*, Elsevier, 26, 31–35.

Oberhofer, W., and Reichsthaler, T. (2004), "Modelling fertility: a semi-parametric approach," *Regensburger Diskussionsbeiträge zur Wirtschaftswissenschaft*, 396.

Oh, E. J., Shepherd, B. E., Lumley, T., and Shaw, P. A. (2018), "Considerations for analysis of time-to-event outcomes measured with error: Bias and correction with SIMEX," *Statistics in medicine*, Wiley Online Library, 37, 1276–1289.

Olsen, J., Juul, S., and Basso, O. (1998), "Measuring time to pregnancy. Methodological issues to consider.," *Human reproduction (Oxford, England)*, Citeseer, 13, 1751–1753.

Radin, R. G., Rothman, K. J., Hatch, E. E., Mikkelsen, E. M., Sorensen, H. T., Riis, A. H., Fox, M. P., and Wise, L. A. (2015), "Maternal recall error in retrospectively reported time-to-pregnancy: an assessment and bias analysis," *Paediatric and perinatal epidemiology*, Wiley Online Library, 29, 576–588.

Ricci, E., Al Beitawi, S., Cipriani, S., Candiani, M., Chiaffarino, F., Viganò, P., Noli, S., and Parazzini, F. (2017), "Semen quality and alcohol intake: a systematic review and meta-analysis," *Reproductive biomedicine online*, Elsevier, 34, 38–47.

Slama, R., Ducot, B., Carstensen, L., Lorente, C., de La Rochebrochard, E., Leridon, H., Keiding, N., and Bouyer, J. (2006), "Feasibility of the current-duration approach to studying human fecundity," *Epidemiology*, JSTOR, 440–449.

Stefanski, L. A., and Cook, J. R. (1995), "Simulation-extrapolation: the measurement error jackknife," *Journal of the American Statistical Association*, Taylor & Francis, 90, 1247–1256.

van Eekelen, R., Putter, H., McLernon, D. J., Eijkemans, M. J., and van Geloven, N. (2020), "A comparison of the beta-geometric model with landmarking for dynamic prediction of time to pregnancy," *Biometrical Journal*, Wiley Online Library, 62, 175–190.

Van Eekelen, R., Scholten, I., Tjon-Kon-Fat, R. I., van der Steeg, J. W., Steures, P., Hompes, P., van Wely, M., Van Der Veen, F., Mol, B. W., and Eijkemans, M. J. (2017), "Natural conception: repeated predictions over time," *Human Reproduction*, Oxford University Press, 32, 346–353.

Weinberg, C. R., and Gladen, B. C. (1986), "The beta-geometric distribution applied to comparative fecundability studies," *Biometrics*, JSTOR, 547–560.

Zare, N., Nouri, B., Moradi, F., and Parvareh, M. (2017), "The study of waiting time to first pregnancy in the south of Iran: A parametric frailty model approach," *International Journal of Reproductive BioMedicine*, International Journal of Reproductive BioMedicine, 15, 11–16.

APPENDIX A: BIAS TABLES

Table A.1: Average Bias Values From Simulation When Curves Fit to 6 Knots Using Data Generated from Weibull Shape and Scale of 0.5 and 6.

		Method				
ME var	Time	Untransformed	Shifted	Transformed	SIMEX log	SIMEX
0.25	3	-0.11	-0.19	0.19	-0.17	0.04
	6	-0.07	-0.15	0.11	-0.16	0.03
	12	-0.04	-0.1	0.09	-0.17	0.01
	24	-0.01	0.04	0.06	-0.13	0.01
	36	0.01	-0.01	0.05	-0.07	0.02
0.5	3	-0.07	-0.16	0.15	-0.18	0.04
	6	-0.03	-0.12	0.13	-0.2	0.03
	12	0	-0.06	0.1	-0.21	0.03
	24	0.01	0.02	0.07	-0.13	0.03
	36	0.03	0.01	0.06	-0.07	0.03
0.75	3	-0.03	-0.14	0.14	0.01	0.06
	6	0.01	-0.08	0.14	-0.03	0.05
	12	0.03	-0.03	0.12	-0.06	0.05
	24	0.03	0	0.09	-0.02	0.04
	36	0.04	0.02	0.06	0	0.03
1	3	0.01	-0.1	0.17	0.03	0.08
	6	0.04	-0.05	0.16	-0.01	0.07
	12	0.05	-0.01	0.14	-0.03	0.07
	24	0.04	0.01	0.09	0	0.05
	36	0.04	0.03	0.07	0.01	0.04

Table A.2: Average Bias Values From Simulation When Curves Fit to 10 Knots Using Data Generated from Weibull Shape and Scale of 0.5 and 6.

		Method				
ME var	Time	Untransformed	Shifted	Transformed	SIMEX log	SIMEX
0.25	3	-0.09	-0.16	0.19	-0.05	0.05
	6	-0.07	-0.13	0.1	-0.12	0.03
	12	-0.04	-0.09	0.09	-0.21	0.03
	24	-0.01	-0.03	0.06	-0.23	0.03
	36	0.01	0	0.05	-0.17	0.03
0.5	3	-0.06	-0.14	0.13	-0.08	0.05
	6	-0.03	-0.09	0.13	-0.17	0.03
	12	0	-0.05	0.1	-0.25	0.03
	24	0.02	-0.01	0.07	-0.26	0.03
	36	0.02	0.01	0.06	-0.18	0.03
0.75	3	-0.03	-0.11	0.15	-0.06	0.06
	6	0.01	-0.06	0.13	-0.15	0.05
	12	0.03	-0.02	0.13	-0.23	0.05
	24	0.03	0.01	0.09	-0.23	0.04
	36	0.03	0.02	0.06	-0.16	0.04
1	3	-0.03	-0.11	0.15	-0.02	0.08
	6	0.01	-0.06	0.13	-0.11	0.07
	12	0.03	-0.02	0.13	-0.18	0.07
	24	0.03	0.01	0.09	-0.18	0.05
	36	0.03	0.02	0.06	-0.12	0.04

Table A.3: Average Bias Values From Simulation When Curves Fit to 6 Knots Using Data Generated from Weibull Shape and Scale of 2 and 10.

		Method				
ME var	Time	Untransformed	Shifted	Transformed	SIMEX log	SIMEX
0.25	3	0.18	0.21	0.57	0.04	0.34
	6	0.23	0.27	0.45	-0.05	0.34
	12	0.08	0.09	0.17	-0.11	0.11
	24	-0.02	-0.02	-0.01	-0.03	-0.01
	36	0	0	0	0	0
0.5	3	0.28	0.32	0.52	0.09	0.4
	6	0.32	0.34	0.53	0.04	0.4
	12	0.11	0.11	0.19	-0.07	0.12
	24	-0.03	-0.03	-0.01	-0.05	-0.03
	36	-0.01	-0.01	0	-0.01	-0.01
0.75	3	0.37	0.38	0.62	0.17	0.47
	6	0.38	0.38	0.55	0.14	0.44
	12	0.13	0.12	0.19	-0.02	0.13
	24	-0.03	-0.04	-0.01	-0.06	-0.03
	36	-0.01	-0.02	0	-0.02	-0.01
1	3	0.43	0.42	0.65	0.49	0.52
	6	0.43	0.41	0.58	0.42	0.48
	12	0.14	0.13	0.2	0.12	0.14
	24	-0.03	-0.04	-0.01	-0.03	-0.03
	36	-0.02	-0.02	0	-0.01	-0.02

Table A.4: Average Bias Values From Simulation When Curves Fit to 10 Knots Using Data Generated from Weibull Shape and Scale of 2 and 10.

		Method				
ME var	Time	Untransformed	Shifted	Transformed	SIMEX log	SIMEX
0.25	3	0.18	0.2	0.57	0.19	0.34
	6	0.24	0.26	0.45	0.04	0.34
	12	0.08	0.09	0.18	-0.16	0.12
	24	-0.02	0.02	-0.01	0.04	-0.01
	36	0	0	0	-0.01	0
0.5	3	0.28	0.3	0.51	0.26	0.4
	6	0.33	0.34	0.53	0.12	0.4
	12	0.1	0.1	0.18	-0.13	0.13
	24	-0.03	-0.03	-0.01	-0.08	-0.02
	36	-0.01	-0.01	0	-0.02	-0.01
0.75	3	0.37	0.36	0.65	0.34	0.47
	6	0.39	0.38	0.55	0.2	0.45
	12	0.12	0.11	0.19	-0.07	0.14
	24	-0.03	-0.03	-0.01	-0.1	-0.03
	36	-0.01	-0.02	0	-0.03	-0.01
1	3	0.43	0.41	0.64	0.41	0.52
	6	0.44	0.41	0.58	0.28	0.48
	12	0.13	0.11	0.2	-0.02	0.15
	24	-0.02	-0.03	-0.01	-0.09	-0.03
	36	-0.01	-0.02	0	-0.03	-0.02

APPENDIX B: MSE TABLES

Table B.1: MSE Values from Weibull Parameters of 2 and 10 Between Methods When Fitting Survival Curves with 6 Knots.

		Method				
ME var	Time	Untransformed	Shifted	Transformed	SIMEX log	SIMEX
0.25	3	0.037	0.060	0.041	0.070	0.019
	6	0.019	0.038	0.017	0.061	0.010
	12	0.008	0.017	0.010	0.057	0.005
	24	0.002	0.004	0.004	0.033	0.001
	36	0.001	0.001	0.002	0.011	0.001
0.5	3	0.030	0.054	0.030	0.062	0.019
	6	0.014	0.030	0.021	0.065	0.010
	12	0.005	0.010	0.012	0.066	0.004
	24	0.002	0.002	0.006	0.032	0.002
	36	0.001	0.001	0.018	0.009	0.001
0.75	3	0.022	0.046	0.029	0.016	0.017
	6	0.010	0.022	0.023	0.014	0.010
	12	0.005	0.007	0.017	0.013	0.005
	24	0.002	0.002	0.008	0.006	0.003
	36	0.002	0.001	0.004	0.002	0.001
1	3	0.018	0.036	0.037	0.015	0.018
	6	0.010	0.016	0.029	0.011	0.012
	12	0.006	0.005	0.021	0.008	0.007
	24	0.003	0.002	0.009	0.005	0.003
	36	0.002	0.001	0.005	0.002	0.002

Table B.2: MSE Values from Weibull Parameters of 2 and 10 Between Methods When Fitting Survival Curves with 6 Knots.

		Method				
ME Var	Time	Untransformed	Shifted	Transformed	SIMEX log	SIMEX
0.25	3	0.054	0.063	0.33	0.018	0.13
	6	0.063	0.081	0.21	0.027	0.12
	12	0.0074	0.0089	0.031	0.0011	0.013
	24	0.00051	0.00044	0.00004	0.0013	0.00022
	36	0.000023	0.000021	0.0000029	0.000025	0.0000082
0.5	3	0.099	0.12	0.27	0.022	0.17
	6	0.11	0.12	0.29	0.022	0.16
	12	0.013	0.014	0.035	0.019	0.016
	24	0.001	0.001	0.000068	0.0036	0.00074
	36	0.00012	0.00013	0.00001	0.00019	0.000075
0.75	3	0.15	0.16	0.39	0.042	0.23
	6	0.15	0.15	0.3	0.039	0.2
	12	0.017	0.016	0.038	0.013	0.019
	24	0.0011	0.0013	0.000066	0.0047	0.0011
	36	0.00023	0.00028	0.000016	0.00046	0.0002
1	3	0.2	0.19	0.43	0.25	0.28
	6	0.19	0.17	0.34	0.18	0.23
	12	0.02	0.017	0.04	0.017	0.021
	24	0.0009	0.0014	0.000043	0.0013	0.0011
	36	0.054	0.063	0.33	0.018	0.13