INTERMEDIATE MARKERS: SURROGACY ASSESSMENT USING PRINCIPAL STRATIFICATION AND MULTI-STATE MODELS

by

Anna S.C. Conlon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2013

Doctoral Committee:

Professor Jeremy M.G. Taylor, Co-Chair
Professor Michael R. Elliott, Co-Chair
Assistant Professor Yun Li
Assistant Professor Rafael Meza

to Shean, my statistically significant other.

ACKNOWLEDGMENTS

There are many people I would like to thank for their support and encouragement throughout the process of writing this dissertation. First, I would like to thank my dissertation committee. Jeremy Taylor, my co-chair, was always available to discuss my work and provide feedback. I have gained a vast amount of experience and knowledge through working with him and being on his training grant. Jeremy has also supported me financially over the past years. Michael Elliott, also my co-chair, has provided valuable insight and ideas on all of my work, and has encouraged me to attend and present my work at many conferences. I feel very fortunate to have had both Jeremy and Mike to work with throughout my time as a student. Yun Li, in addition to serving on my committee, has aided in fostering ideas for future research in topics related to my dissertation. Rafael Meza has generously given his time to serve as a committee member and provided much helpful feedback.

I would also like to thank other professors, mentors and peers. I greatly profited by working with Tim Johnson, and by being on Jack Kalbfleisch's training grant and a GSI for Bhramar Mukherjee. I also immensely enjoyed the classes that I had with all three of them. Harry Haber helped me through the beginning of the Master's program and has provided helpful insights on working as a Biostatistician. Heather Schroeder was my study partner through the Master's program and is still the first person I go to for all of my SAS programming questions. My office mate, Jared Foster, was

always ready for discussions on any topic and to provide helpful feedback. At Oberlin College, Jason Hudson showed me what can be achieved through perseverance and dedication and helped shape my work ethic for all of my post collegiate pursuits.

My family has also been tremendously supportive through my entire academic career. My grandparents, Ruth and Sam provided the financial means for me to pursue all educational opportunities. My parents, Angie and Dave instilled in me a love of learning from a young age and have been endlessly supportive in all of my endeavours. And, lastly, I am greatly thankful to my husband Shean for his constant moral support. This dissertation would not have been possible without his never ending encouragement and confidence in me.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Intermediate markers: surrogacy assessment using principal stratification and multi-state models

by

Anna S.C. Conlon

Co-chairs: Jeremy M.G. Taylor and Michael R. Elliott

Intermediate markers can be useful in clinical trials as either surrogate markers intended to replace the true outcome of interest or as auxiliary variables intended to improve efficiency in the analysis of the true outcome. We explore methods pertaining to both of these roles of intermediate markers. First, we propose methods for assessing the validity of a potential surrogate marker. Working under the principal stratification approach for surrogacy validation proposed by Frangakis and Rubin (2002), we propose quantities to evaluate surrogacy when the joint distribution of the potential surrogate and final outcomes is multivariate normal. The multivariate normality assumption is then relaxed and a Gaussian copula model is used to model the joint distribution of surrogate and final outcomes, and quantities are derived from this model to determine surrogacy. For both the multivariate normal model and the Gaussian copula model, a Bayesian estimation strategy is used and, as some parameters are not identifiable from the data, we explore the use of informative priors that

are consistent with reasonable assumptions in the surrogate marker setting to aid in estimation.

Methods for utilizing an intermediate marker as an auxiliary variable to improve efficiency in the analysis of the true outcome are then considered. A multi-state model with an incorporated cured fraction is used to model recurrence and death in colon cancer. The model is used to assess how individual covariates affect the probability of being cured of disease and the transition rates between the various disease states. Once parameter estimates from the model are obtained, survival probabilities can be estimated with gains in efficiency obtained as compared to Kaplan-Meier estimates. The model is then used in a multiple imputation strategy which imputes death times for censored subjects. By using the joint model, recurrence is used as an auxiliary variable in predicting survival times. We explore the use of a hierarchical model and model adaptations that can be made to potentially further the efficiency gains obtained through the multiple imputation procedure. We demonstrate the potential use of the proposed methods in shorting the length of a trial and reducing sample sizes.

CHAPTER I

Introduction

There is much interest in the use of intermediate outcome variables as either surrogate endpoints or as auxiliary variables for the true outcome of interest in randomized clinical trials, as they may allow trials to be run more quickly and inexpensively. A surrogate endpoint (S) is one that is intended to replace the true endpoint (T) in evaluating therapy and an auxiliary variable is one that is intended to be used to improve the efficiency of the analysis of the true endpoint. In this dissertation we will consider both of these uses of an intermediate marker. First, we explore methods of validating a surrogate endpoint that is intended to replace a true endpoint. We propose surrogate validation measures for multivariate normal surrogate and outcome data, with extensions to non-normally distributed data through the use of a Gaussian copula model. We then look at methods to jointly model an intermediate outcome and final outcome with a goal of utilizing the information from the intermediate variable to improve the analysis on the true endpoint of interest. We propose a multi-state model with an incorporated cured fraction to jointly model colon cancer recurrence and survival and employ the model to utilize recurrence as an auxiliary variable for overall survival.

Chapters II and III consider the use of an intermediate marker, S, as a surrogate maker for the true endpoint, T, and explore methods of validating S as a surrogate. In both of these chapters, we work under the principal surrogacy framework of causal inference proposed by Frangakis and Rubin (2002). In Chapter II, we propose quantities to evaluate surrogacy when the joint distribution of the potential outcomes of S and T follow a multivariate normal distribution. Many previous methods of surrogate validation rely on models for the conditional distribution of T given the treatment (Z) and S. However, S is a post-randomization variable, and unobserved, simultaneous predictors of S and T may exist. When such confounders exist, these methods will not have a causal interpretation. Therefore, there has been much recent work in the area of surrogacy assessment under the principal surrogacy approach, which looks at the distribution of the potential outcomes of T conditional on principal strata based on the joint distribution of the potential outcomes of S, which are pre-randomization variables. Treatment effect estimates that condition on these principal strata are therefore causal estimates. Existing literature on methods of surrogacy assessment using principal stratification has examined settings in which both S and T are binary (Li, *et al.* 2010), or in which S is continuous with binary T (Gilbert and Hudgens, 2008; Zigler and Belin, 2012). Work in the principal stratification setting when both S and T are continuous has been discussed in the application to partial compliance (Bartolucci and Grilli, 2011; Schwartz, *et al.* 2011), where the conditional distribution for each of the potential outcomes of T are modeled separately. Here, we consider the entire joint distribution of the potential outcomes of S and the potential outcomes of T when both S and T are continuous and their joint distribution is multivariate normal. Once parameter estimates from this model are obtained, we examine various causal quantities that may aid in the assessment of S as a surrogate marker for

T. As the model is not fully identifiable from the data, we propose some reasonable prior distributions and assumptions that can be placed on non-identified parameters to aid in the Bayesian estimation scheme. We explore the relationship between our surrogacy measures and the surrogacy measures proposed by Prentice (1989). The method is applied to data from a macular degeneration study where change in visual acuity at six months is assessed as a surrogate for change in visual acuity at one year, and to data from an ovarian cancer study where progression free survival is assessed as a surrogate for overall survival.

In Chapter III, we again consider surrogacy validation measures using the principal stratification framework, but relax the multivariate normality assumption. In this setting, a Gaussian copula model can be used to model the joint distribution of the counterfactual surrogate and final outcome measures. We address the scenario of an ordinal categorical variable as a surrogate for a censored failure time true endpoint. The use of a copula model to assess surrogacy in this setting was explored by Burzykowski, *et al.* (2004), where a Plackett copula was used to jointly model observed tumor response and survival in advanced colorectal cancer. De Leon and Wu (2011) explored the use of the Gaussian copula to jointly model a bivariate discrete and continuous outcome. Here, we extend these ideas by employing a four dimensional Gaussian copula model to jointly model the potential outcomes of an ordinal surrogate marker and the potential outcomes of a censored time to event final outcome and derive quantities from this model to assess surrogacy. A Bayesian estimation strategy is used to aid in the estimation of non-identified parameters, and the use of some prior distributions that are consistent with reasonable assumptions in the surrogacy setting are assessed. We apply the method to data from an advanced colorectal cancer clinical trial where tumor response is assessed as a surrogate for

overall survival.

Chapters IV and V consider the use of S as an auxiliary variable that is meant to aid in the efficiency of the estimation of T. In Chapter IV, we propose a multi-state model with an incorporated cured fraction to jointly model recurrence and death in colon cancer. The multi-state model and cure model have each been separately considered with both parametric and non-parametric assumptions. Here our proposed model combines aspects of both of these models, providing insight into how individual covariates affect the probability of being cured of disease and the time to recurrence, time to death and time to death after recurrence, as well as the association of the two endpoints of interest, recurrence and death. A Bayesian MCMC estimation strategy is used to obtain parameter estimates. Checks for the adequacy of the model fit and for the functional forms of covariates are explored through the use of Cox-Snell residual plots and deviance residual plots, respectively. These model assessments are natural to consider for multi-state models, but we are unaware of literature on using them in cure models. The methods are applied to data from 12 randomized Phase III trials of colon cancer, where there is interest in exploring common covariate effects on each aspect of the disease process across all 12 trials.

Chapter V uses the proposed multi-state model with a cured fraction detailed in Chapter IV to explore the use of recurrence as an auxiliary variable for improving efficiency in estimating the treatment effect on overall survival. Estimates of overall survival and disease free survival can be derived from the model with efficiency gains, as compared to Kaplan-Meier estimates, obtained by utilizing recurrence information and the parametric assumptions of the model. Alternatively, efficiency gains can be achieved by using the model in a multiple imputation procedure to impute death times for censored subjects. Treatment effect estimates on overall survival are then obtained

by combining results from the multiply imputed data sets. As the multi-state model jointly models time to recurrence and time to death, recurrence is an auxiliary variable in the imputation procedure, and treatment effect estimates from the imputed data sets often result in an efficiency gain as compared to the original data, resulting in the potential to shorten the length of the trial. The multiple imputation procedure explored here extends that proposed by Conlon, *et al.* (2011), where recurrence and death were modeled separately. In their procedure, a cure model was used to model recurrence and for death, a proportional hazards model with Weibull baseline hazard and recurrence as a time dependent covariate was used. Here, our proposed model jointly models recurrence and death, and through the Bayesian estimation procedure, the use of some restrictive priors can be explored as a way to further the efficiency gains obtained through the imputation procedure. Additionally, the use of a hierarchical model to facilitate information sharing of common covariate effects across the 12 trials is considered as a modeling strategy, with the goal of improving upon the efficiency gains from the imputed data.

CHAPTER II

Surrogacy Assessment Using Principal Stratification When Surrogate and Outcome Measures are Multivariate Normal

2.1 Introduction

A surrogate endpoint (S) is an intermediate outcome variable occurring in between the treatment (Z) and the outcome of interest (T). The surrogate is usually known to be involved in the mechanism of the disease process and can be measured at an earlier time than the desired outcome. Therefore, there is considerable interest in the use of surrogate markers in clinical trials, as they offer the potential to run trials more cheaply and quickly by extracting information regarding the treatment effect on T through the earlier measured S. Examples of established surrogate markers include blood pressure under anti-hypertensive drug treatment as a surrogate for cardiovascular disease (Weir and Walley, 2006), and three year disease free survival as a surrogate for five year overall survival in colorectal cancer (Sargent, *et al.* 2007). We examine two data examples in the application of our method. The first concerns patients with age-related macular degeneration and considers the use of change in visual acuity at 6 months after starting treatment as a surrogate marker for change

in visual acuity at 1 year. The second concerns ovarian cancer and assesses progression free survival as a surrogate for overall survival.

Before a surrogate can be used in practice, it must be shown to be a valid surrogate for the outcome of interest. In a landmark paper, Prentice (1989) proposed a formal definition of surrogacy along with a validation strategy. Prentice's criteria require that S and T be correlated and the treatment effect on T be fully captured by S. Other methods for surrogacy evaluation have since been proposed, including the proportion of treatment effect explained by S (Freedman, Graubard, and Schatzkin, 1992), and individual-level and trial-level surrogacy association measures in meta-analyses (Buyse, *et al.* 2000).

Surrogacy assessments like these that rely on adjusting for surrogate markers measured after randomization result in estimates that will not have a causal interpretation since the markers are measured after randomization (Rosenbaum, 1984). Therefore, Frangakis and Rubin (2002) (henceforth FR) introduced a definition of a surrogate endpoint, called a "principal surrogate", based on a principal stratification approach. In this framework, each subject has two potential outcomes corresponding to each treatment, denoted $S(Z)$ and $T(Z)$, for $Z = 0, 1$. The principal surrogacy approach looks at the distribution of the potential outcomes of T conditional on principal strata based on the joint distribution of $S(0)$ and $S(1)$. The principal strata are unaffected by treatment, and are thus pre-randomization variables. Treatment effect estimates that condition on these principal strata are therefore causal estimates when treatments are randomly assigned.

The rationale for considering whether the principal stratification approach is appropriate for assessing surrogacy has been discussed in the literature, with some support provided in the discussion by VanderWeele (2011) and by Zigler and Belin

(2012). In this approach, the value of S as a surrogate for T is determined by the extent to which the causal effect of treatment on S can reliably predict the causal effect of treatment on T. The rationale for considering principal surrogacy or more generally considering the joint distribution of $S(0), S(1), T(0), T(1)$ is most easily explained in the case where S and T are binary. In this case, the joint distribution of $S(0), S(1), T(0), T(1)$ amounts to a partition of the population into cells with a probability attached to each cell. These probabilities completely characterize the population and from them an assessment of surrogacy can be made. For example, one can consider the fraction of the population for which $T(0)$ is not equal to $T(1)$ amongst those who have $S(0)$ not equal to $S(1)$. Then additionally, this fraction might be contrasted with the fraction of the population for which $T(0)$ is not equal to $T(1)$ amongst those who have $S(0)$ equal to $S(1)$. As we will describe below, other summary measures that can be obtained from the joint distribution might also be considered. When S and T are continuous, the joint distribution of $S(0), S(1), T(0), T(1)$ again completely characterizes the population, from which summary measures for assessing surrogacy, such as the distribution of $T(1) - T(0)$ given $S(1) - S(0)$, can be obtained. If one accepts that the joint distribution completely characterizes the population, then the challenges are determining what useful summary measures to extract from this distribution, and the estimation of this distribution.

We note that the principal stratification approach to assessing surrogacy uses a causal framework, but the causal framework it uses differs from the framework presented by Pearl (1995) and discussed in Joffe and Greene (2009). In the principal stratification framework, there are only two causal effects, one on S and one on T and we are interested in the association between these two. The other causal framework, while it may also be interesting to consider, does require additional consideration of

the effect of S on T, requiring hypothetical manipulations of S. This alternative causal framework is more mechanistic and allows notions of direct and indirect effects of Z on T. We will not pursue it in this paper.

Existing literature on methods for surrogacy assessment using the principal stratification approach has examined settings in which both S and T are binary (Li *et al.* 2010), or in which S is continuous with binary T (Gilbert and Hudgens, 2008; Zigler and Belin, 2012). For a binary S and T, Li, *et al.* (2010) developed an estimation method for the causal quantities associated with the cross classification of the potential outcomes using a log-linear model and Bayesian estimation procedure. Gilbert and Hudgens (2008) (henceforth GH) used the framework of FR to develop an estimand, termed the causal effect predictiveness (CEP) surface for evaluating surrogacy when S is continuous or categorical and T is binary. Work in the principal surrogacy framework when both S and T are continuous has been discussed in the application to partial compliance (Bartolucci and Grilli, 2011; Schwartz, *et al.* 2011). In this context, the joint distribution of the potential outcomes of the intermediate variable, in this case degree of compliance, is modeled either parametrically or semi-parametrically with principal causal effects measured by comparisons of the potential outcomes of T conditional on S, where the conditional distributions for $T(0)$ and $T(1)$ are modeled separately. Qin, *et al.* (2008) used a principal stratification approach in the assessment of a continuous surrogate with a time to event outcome.

Here, we consider the entire joint distribution of $(S_i(0), S_i(1), T_i(0), T_i(1))$ and propose estimands to evaluate principal surrogacy when both S and T are continuous and the joint distribution of the potential outcomes is multivariate normal. Once parameter estimates for this distribution are obtained, various causal quantities that may aid in the assessment of S as a surrogate marker for T may be examined. Specific

quantities of interest include $E[T(1) - T(0)|S(1) - S(0)]$, $P(T(1) - T(0) > 0|S(1) - S(0))$, and the correlation between $T(1)-T(0)$ and $S(1)-S(0)$. The use of $cor(T(1)-T(0), S(1)-S(0))$ has been discussed by Wang, *et al.* (2012), who specifically contrast it with the observable correlation between S and T, given the treatment group.

Because some parameters of the joint distribution are not fully identifiable from the data, we use a Bayesian estimation procedure with plausible prior distributions and some reasonable constraints on model parameters to reduce the non-identifiability problem of modeling counterfactual observations and to aid in estimation of the quantities of interest. In order to facilitate the consideration of reasonable constraints we found it convenient to decompose the covariance matrix, Σ of $(S_i(0), S_i(1), T_i(0), T_i(1))$ as $\Sigma = QRQ$ (Barnard, McCulloch and Meng, 2000), and place constraints on the correlations R, rather than on the covariance terms in Σ. We also explore the relationship between some of the proposed surrogacy assessment quantities and those based on the Prentice criteria. In Section 2.2, we describe the model and possible constraints that could be made to facilitate estimation. In Section 2.3, we introduce surrogacy measures based on the potential outcomes framework. Section 2.4 describes the Bayesian estimation procedure that we use and Section 2.5 provides simulation results from this procedure. In Section 2.6 we apply these methods to the macular degeneration data and ovarian cancer data. Section 2.7 concludes with a discussion.

2.2 Potential Outcomes Model

For a randomized trial with treatment assignment Z ($Z = 1$ or 0), continuous surrogate marker S and continuous true endpoint T, each subject i, $i = 1, ..., n$, has two potential outcomes for each of S_i and T_i, denoted by $S_i(z_i)$ and $T_i(z_i)$.

Only one outcome, corresponding to the received treatment for subject i in each of the pairs $(S_i(0), S_i(1))$ and $(T_i(0), T_i(1))$ can be observed. The joint distribution of $(S_i(0), S_i(1), T_i(0), T_i(1))$ describes the causal associations between Z, S and T. In the continuous setting where $(S_i(0), S_i(1), T_i(0), T_i(1))$ is multivariate normal with mean μ and covariance matrix Σ, we have the following joint distribution:

$$\begin{pmatrix} S_i(0) \\ S_i(1) \\ T_i(0) \\ T_i(1) \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{S_0} \\ \mu_{S_1} \\ \mu_{T_0} \\ \mu_{T_1} \end{pmatrix}, \begin{pmatrix} \sigma^2_{S_0} & \rho_s \sigma_{S_0} \sigma_{S_1} & \rho_{00} \sigma_{S_0} \sigma_{T_0} & \rho_{01} \sigma_{S_0} \sigma_{T_1} \\ & \sigma^2_{S_1} & \rho_{10} \sigma_{S_1} \sigma_{T_0} & \rho_{11} \sigma_{S_1} \sigma_{T_1} \\ & & \sigma^2_{T_0} & \rho_t \sigma_{T_1} \sigma_{T_0} \\ & & & \sigma^2_{T_1} \end{pmatrix} \right)$$

The mean μ and the variances corresponding to the diagonal elements of Σ, along with the correlations between $(S_i(0), T_i(0))$ and $(S_i(1), T_i(1))$ corresponding to ρ_{00} and ρ_{11}, are fully identifiable from the data. Because only one of the counterfactual pairs of outcomes is observed for each subject, ρ_s, ρ_t, ρ_{01}, and ρ_{10} are not identifiable. However, the identifiable correlation parameters together with the requirement that Σ be positive definite places boundary constraints on these non-identified parameters, which, along with other plausible assumptions that we can make, aids in their estimation.

We make the standard assumptions of ignorable treatment assignments (Rubin, 1978) and the stable unit treatment value assumption (SUTVA). Ignorable treatment assignment implies that Z is independent of $(S(0), S(1), T(0), T(1))$ and holds for blinded, randomized trials. SUTVA implies that the potential outcomes $(S_i(0), S_i(1), T_i(0), T_i(1))$ are independent of the treatment assignments of other subjects. This allows us to write the potential outcomes for subject i as a function of Z_i rather than of the entire vector of subject treatment assignments.

Other context specific constraints can be added, such as all ρ's ≥ 0, a plausible assumption for most variables S that would be under consideration as a potential surrogate for T, and especially when the identifiable Pearson correlation coefficients, $\hat{\rho}_{00}$ and $\hat{\rho}_{11}$, are positive. Other plausible assumptions are $\rho_{01} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$, and $\rho_{10} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$, indicating a belief that the correlation between the surrogate response and final outcome response in opposite treatment arms is less than the correlation between the surrogate response and final outcome response within the same treatment arm, or the correlation between the surrogate responses or final treatment responses across treatment arms.

2.3 Assessing Surrogacy Using Potential Outcomes Framework

2.3.1 Definitions of Surrogacy

Because S is a post-randomization variable, unobserved simultaneous predictors of both S and T may exist. In this case, methods of surrogacy assessment that require conditioning on S do not result in causal estimates (Rosenbaum, 1984). When baseline covariates account for all common causes of S and T, surrogacy measures that condition on S will be causal. However, the assumption of no unmeasured confounders of S and T is untestable, potentially leading to noncausal estimates (Gilbert, *et al.* 2009). Therefore, FR proposed a definition of principal surrogacy (PS), which uses a principal stratification approach to assess the validity of a surrogate marker. This framework focuses on the distribution of $p(T(0), T(1)|S(0), S(1))$. Since $S(1)$ and $S(0)$ are unaffected by treatment assignment, they can be treated as baseline covariates. Quantities estimated from this distribution will therefore always have

a causal interpretation. FR proposed two measures of surrogacy, the "dissociative effect" given by $E(T_i(1) - T_i(0)|S_i(1) = S_i(0))$, and the "associative effect" given by $E(T_i(1) - T_i(0)|S_i(1) \neq S_i(0))$.

For the multivariate normal distribution, the distribution of $(T(1) - T(0)|S(1) - S(0) = s)$ is normal with mean given by $E[T_i(1) - T_i(0)|S_i(1) - S_i(0) = s] = \gamma_0 + \gamma_1 s$, where

$$\gamma_0 = (\mu_{T_1} - \mu_{T_0}) - \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)(\mu_{S_1} - \mu_{S_0})$$

$$\gamma_1 = \left(\frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\right)$$

and variance given by

$$\sigma^2_{T_0} + \sigma^2_{T_1} - 2\rho_t\sigma_{T_0}\sigma_{T_1} - \frac{\left(\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}\right)^2}{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}.$$

The value of γ_0 is then a measure of the "dissociative effect". Values of γ_0 near zero indicate that the causal effect of treatment on the final outcome is near zero when the causal effect of treatment on the surrogate is near zero, a characteristic that a good principal surrogate should possess. When $\gamma_0 \neq 0$, there can be a causal effect of the treatment on the final outcome even if there is no causal effect of the treatment on the surrogate, implying that the treatment affects the outcome through pathways that do not involve the surrogate. The value of $\gamma_0 + \gamma_1 s$ is a measure of the "associative effect". A good principal surrogate should result in a large associative effect, indicating that as the treatment effect on the surrogate increases, the treatment effect on the final outcome increases as well, thus this measures the extent to which the effect of Z on S is associated with an effect of Z on T (VanderWeele, 2011).

GH suggest a refined definition of a principal surrogate endpoint. In their setting with binary T they define two properties, "average causal necessity" (ACN) and "average causal sufficiency" (ACS), that a valid surrogate marker should satisfy. ACN

is satisfied if $risk_{(1)}(s_1, s_0) = risk_{(0)}(s_1, s_0)$ for all $s_1 = s_0$, where $risk_{(z)}(s_1, s_0) = p(T(Z) = 1|S(1) = s_1, S(0) = s_0)$. ACS is satisfied if there exists some constant $C \geq 0$ such that $risk_{(1)}(s_1, s_0) \neq risk_{(0)}(s_1, s_0)$ for all $|s_1 - s_0| > C$. In our setting of continuous T, we can consider the joint conditional distribution of $(T(0), T(1))$. Specific summaries of this joint distribution which are of major interest include $E[T(1) - T(0)|S(1) - S(0) = s]$ for $s = 0$ and $|s| > C$ for some $C \geq 0$, $P(T(1) > T(0)|S(1) - S(0) = s)$ and the correlation between $T(1) - T(0)$ and $S(1) - S(0)$. Also of interest is the "causal effect predictiveness (CEP) surface" proposed by GH which considers the entire surface of $E[T(1) - T(0)|S(1), S(0)]$. In the case of a binary outcome, the definitions of ACN and ACS are equivalent to the conditional expectations of $T(0)$ and $T(1)$. GH suggest that their framework is also applicable in the setting of continuous endpoint, with the expressions for $P(T(z) = 1 \mid \cdot)$ replaced by $E(T(z) \mid \cdot)$. We can therefore consider ACN satisfied if $E[T(1) - T(0) \mid S(1) - S(0) = 0] = 0$ and ACS satisfied if $E[T(1) - T(0) \mid S(1) - S(0) = s] \neq 0$ for all $|s| > C$, corresponding to $\gamma_0 = 0$ and $\gamma_1 \neq 0$. In the setting of a continuous endpoint, we may also consider the entire conditional distribution of $T(1)$ and $T(0)$. In this case, ACN in distribution is satisfied if $P(T(1) - T(0) > 0 \mid S(1) - S(0) = 0) = 0.5$ and ACS in distribution is satisfied if $P(T(1) - T(0) > 0 \mid S(1) - S(0) = s) \neq 0.5$. For multivariate normal data this conditional probability is:

$$\Phi_{10}(s) = P(T(1) - T(0) > 0|S(1) - S(0) = s) = \Phi\left(\frac{\gamma_0 + \gamma_1 s}{\sqrt{\sigma^2_{T_0} + \sigma^2_{T_1} - 2\rho_t\sigma_{T_0}\sigma_{T_1} - \gamma_1^2(\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1})}}\right)$$

In the multivariate normal setting, the metrics of ACN and ACS based on the entire conditional distribution and based only on the conditional expectation are

closely related. If ACN in expectation holds ($\gamma_0 = 0$), then ACN in distribution will also hold ($\Phi_{10}(0) = 0.5$ when $S(1) - S(0) = 0$). If both ACN and ACS in expectation hold ($\gamma_0 = 0$ and $\gamma_1 \neq 0$), then $\Phi_{10}(s) \neq 0.5$ for $s \neq 0$, satisfying ACS in distribution. Therefore, validation of S as a surrogate can be done by evaluating either $\gamma_0 = 0$, $\gamma_1 \neq 0$ or $\Phi_{10}(0) = 0.5$, $\Phi_{10}(s) \neq 0.5$ for $s \neq 0$.

Another potentially useful measure to assess surrogacy is the correlation between $T(1) - T(0)$ and $S(1) - S(0)$, which we denote by ρ_{ST}. It can be shown that ρ_{ST} is given by

$$\rho_{ST} = \frac{\rho_{11}\sigma_{S_1}\sigma_{T_1} - \rho_{10}\sigma_{S_1}\sigma_{T_0} - \rho_{01}\sigma_{S_0}\sigma_{T_1} + \rho_{00}\sigma_{S_0}\sigma_{T_0}}{\sqrt{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}\sqrt{\sigma^2_{T_0} + \sigma^2_{T_1} - 2\rho_t\sigma_{T_0}\sigma_{T_1}}} = \gamma_1\sqrt{\frac{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}{\sigma^2_{T_0} + \sigma^2_{T_1} - 2\rho_t\sigma_{T_0}\sigma_{T_1}}}$$

When $\rho_{ST} = 0$, γ_1 will also be 0, hence ACS in expectation will not be met, and S cannot be a valid principal surrogate for T. When $\rho_{ST} > 0$, $\gamma_1 > 0$, thus satisfying ACS. A positive value of ρ_{ST} does not, however, provide information about γ_0, and therefore cannot alone determine whether or not S is a valid surrogate marker. A final way that we consider summarizing the conditional distribution of $T(1) - T(0)$ given $S(1) - S(0) = s$, is through the CEP graph, which is a plot of $E[T(1) - T(0)|S(1) - S(0) = s]$ versus s, which in the multivariate normal setting, is simply a plot of $\gamma_0 + \gamma_1 s$ versus s.

One issue that arises in the validation of surrogate markers is the presence of the "surrogate paradox" (Chen *et al.*, 2007; VanderWeele, 2013) where there is a positive effect of treatment on the surrogate, the surrogate and outcome are positively correlated, but there is a negative effect of treatment on the outcome. VanderWeele, (2013) notes that the principal surrogacy criteria capture the notion of surrogacy well and conceptually avoid the surrogate paradox. However, he also points out that while theoretically appealing, due to lack of identifiability, the criteria may be difficult to

use in practice. It is easy to see that if both ACN and ACS hold ($\gamma_0 = 0$ and $\gamma_1 \neq 0$, respectively), then $E[T(1) - T(0) \mid S(1) - S(0) = s]$ will be in the same direction as $S(1) - S(0) = s$ as long as $\gamma_1 > 0$, thus avoiding the surrogate paradox. If ACN is not perfectly satisfied (i.e. $\gamma_0 \neq 0$), then there is a small range of s for which the surrogate paradox can occur. If $\gamma_0 < 0$, then $E[T(1) - T(0) \mid S(1) - S(0) = s] < 0$ for $s \in [0, -\gamma_0/\gamma_1]$. If $\gamma_0 > 0$, then $E[T(1) - T(0) \mid S(1) - S(0) = s] > 0$ for $s \in [-\gamma_0/\gamma_1, 0]$. Wu *et al.* (2011) provide methods for detecting the presence of the surrogate paradox based on the observed data.

2.3.2 Relationship Between Principal Surrogacy Measures and Prentice Surrogacy Criteria

The ACN and ACS measures corresponding to conditional expectation can be linked to the original surrogacy definition proposed by Prentice (1989). Prentice's criteria for a valid surrogate require that $f(T|Z) \neq f(T)$, $f(S|Z) \neq f(S)$, $f(T|S) \neq f(T)$, and $f(T|S, Z) = f(T|S)$. In the normal setting, the observed variables have the following distributions:
$\begin{pmatrix} S_i \\ T_i \end{pmatrix} \mid Z_i = 1 \sim N\left(\begin{pmatrix} \mu_{S_1} \\ \mu_{T_1} \end{pmatrix}, \begin{pmatrix} \sigma^2_{S_1} & \rho_{11}\sigma_{S_1}\sigma_{T_1} \\ & \sigma^2_{T_1} \end{pmatrix}\right)$ and $\begin{pmatrix} S_i \\ T_i \end{pmatrix} \mid Z_i = 0 \sim N\left(\begin{pmatrix} \mu_{S_0} \\ \mu_{T_0} \end{pmatrix}, \begin{pmatrix} \sigma^2_{S_0} & \rho_{00}\sigma_{S_0}\sigma_{T_0} \\ & \sigma^2_{T_0} \end{pmatrix}\right)$.
While Z is independent of the (pre-randomization) joint distribution of $(S(0), S(1))$, it is not independent of the observed S. Thus, assuming the probability of being randomized to either $Z = 1$ or $Z = 0$ is 0.5, the conditional expectation of T is linear in S, and can be written as $E[T_i \mid S_i] = \mu_0 + \mu_1 S_i$, where $\mu_0 = \frac{1}{2}(\mu_{T_1} + \mu_{T_0}) - \frac{1}{2}(\frac{\rho_{11}\sigma_{T_1}}{\sigma_{S_1}}\mu_{S_1} + \frac{\rho_{00}\sigma_{T_0}}{\sigma_{S_0}}\mu_{S_0})$ and $\mu_1 = \frac{1}{2}(\frac{\rho_{11}\sigma_{T_1}}{\sigma_{S_1}} + \frac{\rho_{00}\sigma_{T_0}}{\sigma_{S_0}})S_i$. Furthermore, $E[T_i|Z_i] = \theta_0 + \theta_1 Z_i$ where $\theta_1 = \mu_{T_1} - \mu_{T_0}$, $E[S_i|Z_i] = \alpha_0 + \alpha_1 Z_i$ where $\alpha_1 = \mu_{S_1} - \mu_{S_0}$, and $E[T|S, Z] = \beta_0 + \beta_1 Z + \beta_2 S + \beta_3 SZ$ where $\beta_1 = (\mu_{T_1} - \mu_{T_0}) - \left(\frac{\rho_{11}\sigma_{T_1}}{\sigma_{S_1}}\mu_{S_1} - \frac{\rho_{00}\sigma_{T_0}}{\sigma_{S_0}}\mu_{S_0}\right)$, $\beta_2 = \frac{\rho_{00}\sigma_{T_0}}{\sigma_{S_0}}$, and $\beta_3 = \frac{\rho_{11}\sigma_{T_1}}{\sigma_{S_1}} - \frac{\rho_{00}\sigma_{T_0}}{\sigma_{S_0}}$. The Prentice criteria are satisfied when $\theta_1 \neq 0$,

$\alpha_1 \neq 0$, $\mu_1 \neq 0$, $\beta_1 = 0$, $\beta_2 \neq 0$, and $\beta_3 = 0$. It can be shown that when

$$\frac{\rho_{11}\sigma_{T_1}}{\sigma_{S_1}} = \frac{\rho_{00}\sigma_{T_0}}{\sigma_{S_0}} \tag{II.1}$$

and

$$\rho_{00}\rho_s = \frac{1}{2}\left(\rho_{10} + \rho_{01}\frac{\sigma_{S_0}\sigma_{T_1}}{\sigma_{S_1}\sigma_{T_0}}\right) \tag{II.2}$$

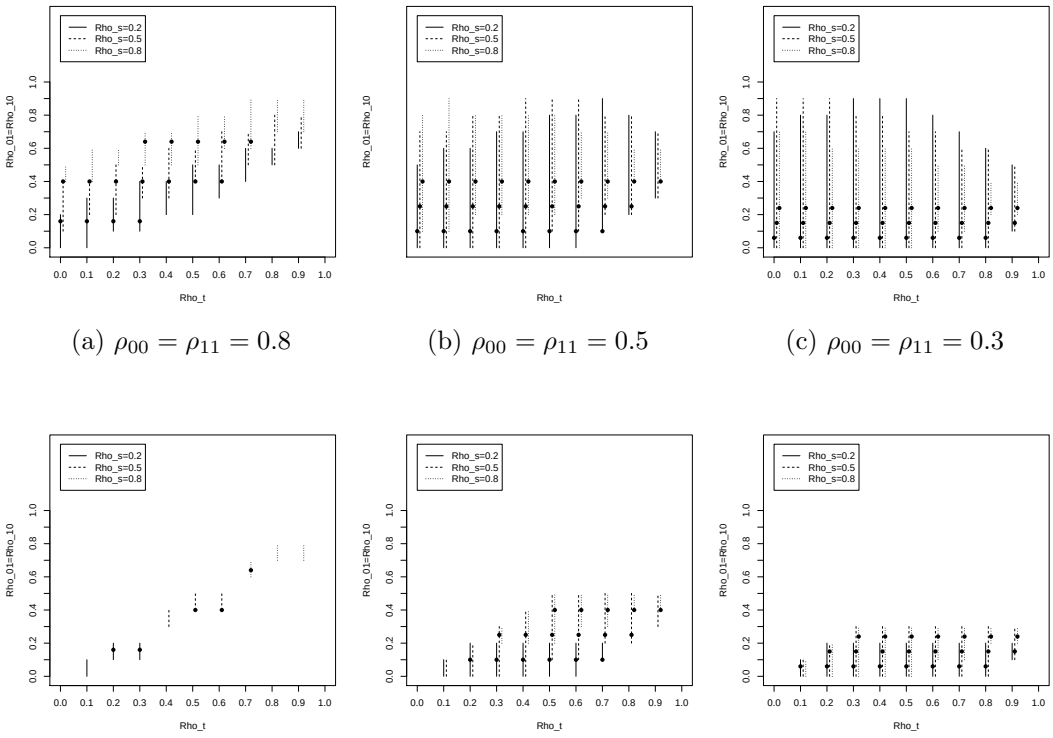
we have $\gamma_1 = \beta_2 = \mu_1$, $\gamma_0 = \beta_1$ and $\beta_3 = 0$. Therefore, under these conditions, the Prentice criteria and the principal surrogacy criteria requiring that both ACN and ACS be met (or $\gamma_0 = 0$ and $\gamma_1 \neq 0$) will reach the same conclusions regarding the validity of S as a surrogate. When the above conditions are not met, conflicting conclusions may be drawn by the Prentice criteria and principal surrogacy criteria. As we regard principal surrogacy to be the main objective in surrogacy assessment, approaching the question of surrogacy using the Prentice criteria in this case may lead to erroneous conclusions.

In any real setting we would not expect the conditions in equations II.1 and II.2 to be exactly satisfied. However, in many settings we can see that the Prentice criteria and principal surrogacy criteria will reach similar conclusions. Often $\sigma_{S_0} \approx \sigma_{S_1}$, $\sigma_{T_0} \approx \sigma_{T_1}$ and we might expect ρ_{00} to be similar to ρ_{11}, thus equation II.1 is approximately satisfied. Similarly for any candidate surrogate, we may expect the average of the "across treatment arm" correlations, ρ_{01} and ρ_{10}, to be less than the "within treatment arm" correlation ρ_{00}, and the correlation of the surrogate marker across treatment arms, ρ_s; thus departures from equality in equation II.2 may not be large.

2.3.3 Parameter Identifiability and Restrictions

Given the identified parameters, the positive definite restriction on R, and plausible assumptions about correlation values, we can gain some insight into the possible ranges, or "identification regions" (Gustafson, 2010) for the partially identified parameters and examine scenarios within this space which lead to different surrogacy conclusions. Under the restriction that all ρ's are non-negative, and the simplifying assumptions that $\rho_{01} = \rho_{10}$, $\rho_{11} = \rho_{00}$, and $\sigma_{S_0} = \sigma_{S_1} = \sigma_{T_0} = \sigma_{T_1}$, the top half of Figure 2.1 displays the possible ranges for $\rho_{01} = \rho_{10}$ across different values of ρ_s and ρ_t for a given $\rho_{11} = \rho_{00}$, where ρ_{11} and ρ_{00} are the identifiable Pearson correlation coefficients between $S_i(1)$ and $T_i(1)$, and $S_i(0)$ and $T_i(0)$, respectively. The length of the identification region for ρ_{01} and ρ_{10} is smallest when ρ_{11} and ρ_{00} are large. For all values of ρ_{11} and ρ_{00}, the length of the identification region for ρ_{01} and ρ_{10} decreases as ρ_s and ρ_t increase. The bottom half of Figure 2.1 provides ranges for these parameters under the additional restriction that $\rho_{01} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$. This restriction greatly reduces the range of possible values for the partially identified parameters, and has implicit effects on the possible ranges for γ_0 and γ_1. Under these restrictions, γ_1 must be greater than 0, implying that ACS always holds. In this scenario where ACS always holds, poor principal surrogates can be characterized by large values of γ_0, implying that the treatment can effect the outcome without effecting the surrogate. Alternatively, a poor surrogate would have a small value of γ_1, implying that there is still a positive, but weak association between causal effects on the surrogate and causal effects on the outcome. These restrictions seems reasonable, as S is typically known to somehow be associated with or a relevant aspect of the disease process, so even if it is not a valid principal surrogate from an ACN and ACS

perspective, we expect there to be at least a small association of treatment effects on S with treatment effects on T. The solid points in each figure are parameter values under which the Prentice criteria and PS criteria are in agreement. In this restricted space the deviation between the Prentice criteria and the PS criteria are less than in the unrestricted space, however we see that scenarios can arise in which the Prentice criteria lead to incorrect conclusions regarding the validity of a principal surrogate.



(a) $\rho_{00} = \rho_{11} = 0.8$ (b) $\rho_{00} = \rho_{11} = 0.5$ (c) $\rho_{00} = \rho_{11} = 0.3$

(d) $\rho_{00} = \rho_{11} = 0.8$ (e) $\rho_{00} = \rho_{11} = 0.5$ (f) $\rho_{00} = \rho_{11} = 0.3$

Figure 2.1: Identification regions of unidentified parameters in MVN model
Plots (a), (b), and (c): under restriction ρ's ≥ 0
Plots (d), (e), and (f): under restriction ρ's ≥ 0, $\rho_{01} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$
Solid points: PS criteria and Prentice criteria in agreement

2.4 Estimation Procedure

A Bayesian approach is used to estimate parameters. Unobserved potential outcomes are treated as missing data and imputed from the appropriate posterior distribution at each iteration of the Markov chain. The covariance matrix Σ is decomposed as QRQ, where Q is the diagonal matrix of standard deviations and R is the correlation matrix. Assuming a priori independence, this allows us to factor the prior distribution $p(\mu, \Sigma)$ as $p(\mu)p(R)p(Q)$ and to place non-informative priors on the fully identified parameters μ, Q, ρ_{00}, and ρ_{11}. Specifically, the prior for μ is $N_4(0, \Sigma_0)$, where $\Sigma_0 = diag(10^6)$, and the prior for each diagonal element of Q is $p(\sigma_j) \propto 1$, for $j = (S(0), S(1), T(0), T(1))$. We place marginal priors on each of the correlation parameters in R and explore the use of four different prior assumptions. For each of these there is the additional assumption that R must be positive definite. The four priors are

(a) Jointly uniform prior such that for each of the six correlations $p(\rho) \sim Unif(-1, 1)$

(b) Jointly uniform prior such that for each of the six correlations $p(\rho) \sim Unif(0, 1)$

(c) All $\rho's \geq 0, \rho_{01} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$, and $\rho_{10} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$

(d) Beta priors such that:

- $p(\rho_{11}) \sim Unif(0, 1)$
- $p(\rho_{00}) \sim Unif(0, 1)$
- $p(\rho_{10})$ and $p(\rho_{01}) \sim Beta(3\alpha_0, 3 - 3\alpha_0)$ such that $P(\rho_{01}, \rho_{10} \leq min(\hat{\rho_{00}}, \hat{\rho_{11}})) = 0.80$
- $p(\rho_s)$ and $p(\rho_t) \sim Beta(3\alpha_1, 3 - 3\alpha_1)$ such that $P(\rho_s, \rho_t \geq E[\rho_{10}]) = 0.80$

where $\hat{\rho}_{00}$ and $\hat{\rho}_{11}$ are the Pearson correlation coefficients estimated from the observed data and $\mathrm{E}[\rho_{10}]$ is the expected value under the $Beta(3\alpha_0, 3-3\alpha_0)$ distribution. Prior assumption (a) is a non-informative prior on all of the correlations. Under scenario (b), all correlations are constrained to be positive, a plausible assumption especially when $\hat{\rho}_{00}$ and $\hat{\rho}_{11}$ are positive. In scenario (c), in addition to the positivity assumption, we restrict ρ_{01} and ρ_{10} to be smaller than the other four correlation parameters. This seems reasonable as ρ_{01} and ρ_{10} are measures of the correlation between the surrogate response and final outcome response in opposite treatment arms, which is unlikely to be larger than the correlation between the surrogate response and final outcome response within the same treatment arm, or the correlation between the surrogate responses or final treatment responses across treatment arms. Finally, prior assumption (d) places similar restrictions on the correlations as assumption (c), but is a little more flexible as ρ_{01} and ρ_{10} are only assumed to be smaller than the other correlations with a probability of 0.8. Appendix A provides density plots of the Beta priors when $\hat{\rho}_{00}$ and $\hat{\rho}_{11}$ are equal to 0.8, 0.5, and 0.3.

Posterior estimates of the unobserved potential outcomes, parameter values, and the causal quantities of interest, γ_0, γ_1, $\Phi_{10}(0)$, ρ_{ST}, and the CEP curve at the points $(\mu_{S_1} - \mu_{S_0}) \pm 2SD(S(1) - S(0))$, where $\mathrm{SD}(S(1) - S(0))$ is the standard deviation of $(S(1) - S(0))$, are obtained using the Gibbs sampler. Each component of Q and R are drawn one at a time. When drawing each element of R, the range of possible values must first be determined in order to satisfy the positive definite requirement, given that the other correlations are held fixed. The range of values corresponding to a positive definite matrix are those in the interval determined by the roots of the quadratic equation that result from solving $|R| = 0$. The specific equations solved to obtain parameter ranges are provided in Appendix B.

As the posterior distributions for the components of Q and R can not be easily sampled from, draws are made using the griddy Gibbs sampler (Ritter and Tanner, 1992). Details of the Gibbs sampler are provided in Appendix C.

2.5 Simulations

We conduct simulations to evaluate the performance of the above methods of surrogacy assessment. We consider the scenarios where under the true parameter values of the simulated data, surrogate validity is the same (S is valid, or S is invalid) under both the Prentice criteria and PS criteria. We also consider the two cases where, under the true parameter values of the simulated data, S is valid under Prentice but not under PS, and S is valid under PS but not under Prentice. In this paper we interpret the results from the perspective that principal surrogacy is the correct approach. We investigate whether the wrong conclusions would be reached if the Prentice criteria were used instead, and whether it is easier to validate a principal surrogate depending on whether or not the Prentice criteria are also satisfied.

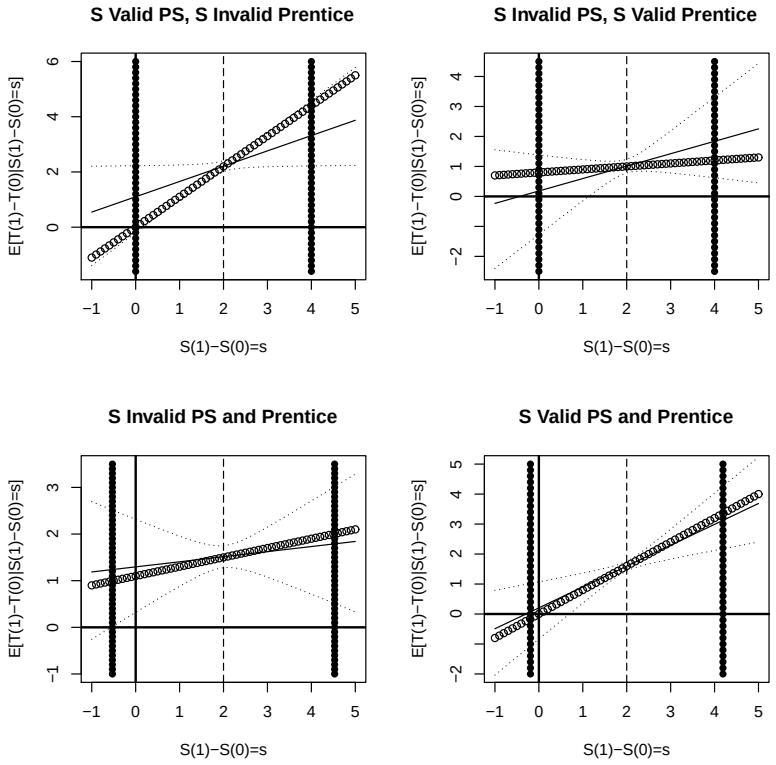
We first explore the sensitivity of the estimation to the plausible prior restrictions on R that we might make. For each simulation, we simulate 200 data sets, each with a sample size of 300. For each of the four different surrogacy scenarios we perform four simulations, with the estimation procedure done using each of the priors outlined in Section 2.4. Table 2.1 provides the posterior means and standard deviations of the Bayesian estimates and means of the posterior standard deviations ($P\bar{S}D$) for the model parameters, the quantities of interest from the Prentice model, and the causal quantities of interest, γ_0, γ_1, ρ_{ST}, $\Phi_{10}(0)$ and the CEP curve at $(\mu_{S_1} - \mu_{S_0}) \pm 2SD(S(1) - S(0))$. The identified parameters are not sensitive to

changes in the prior specifications (only results under the Beta priors shown) while the unidentified parameters are quite sensitive to prior assumptions. In all four scenarios, the standard deviation of the Bayesian estimates is smaller than $P\bar{S}D$ for the unidentified parameters. There is very little bias in estimating β_1, β_2, and β_3, while there is some bias in estimating γ_0, γ_1, ρ_{ST}, $\Phi_{10}(0)$ and the CEP points. The estimation performed using Beta priors appears to provide the best estimation for the unidentified parameters across these four models. While this prior does not always perform best in terms of bias, it has on average better coverage of the parameters across the different scenarios than the other models.

Using the Beta priors, Table 2.2 provides an estimate of the proportion of times that S would be considered a valid principal surrogate based on the proposed measures. This means that 0 is in the 95% credible interval for γ_0, and outside of the 95% credible interval for γ_1. For $\Phi_{10}(0)$ this means that 0.5 is in the 95% credible interval. For ρ_{ST}, we look at the proportion of times that its credible interval is outside of 0, and for the CEP curve we look at the proportion of times that the 95% credible intervals at the points $(\mu_{S_1} - \mu_{S_0}) + 2SD(S(1) - S(0))$ and $(\mu_{S_1} - \mu_{S_0}) - 2SD(S(1) - S(0))$ do not overlap (denoted by $CEP^U_{-2SD} < CEP^L_{+2SD}$). Table 2.2 also provides an estimate of the proportion of times that S would be a valid surrogate based on the Prentice criteria (0 in the 95% confidence interval for $\hat{\beta}_1$, and $\hat{\beta}_3$ and 0 outside of the 95% confidence interval for $\hat{\beta}_2$) The entire CEP curve, shown in Figure 2.2, is also used to visually assess principal surrogacy and the expected treatment effect on T at relevant values of $S(1) - S(0)$.

Our estimation procedure for γ_0 and γ_1 reaches the correct conclusion regarding surrogate validity when principal surrogacy is unmet, regardless of whether or not the Prentice criteria are met under the true parameters. We correctly identify S as

Figure 2.2: Simulation results: CEP curves



KEY: $\circ\circ\circ$ True line, $-$ Mean CEP, $\cdots$ CEP 95% CI, $|$ $(\mu_{S_1} - \mu_{S_0})$, $\bullet$ $(\mu_{S_1} - \mu_{S_0}) \pm 2SD(S(1) - S(0))$

an invalid principal surrogate 99% of the time in the scenario in which S is invalid under the Prentice criteria, and 85% of the time when S is valid under the Prentice criteria. In comparison, the Prentice criteria incorrectly determine S to be a valid surrogate 26% and 92% of the time, respectively, in these two scenarios. When S is a valid principal surrogate, our procedure most reliably determines surrogate validity when the Prentice criteria would also conclude that S is a valid surrogate. In this scenario, we correctly identify S as a valid principal surrogate 94% of the time, while the Prentice criteria conclude S to be a valid surrogate 95% of the time. When S is a valid principal surrogate but the Prentice criteria show S to be invalid, our estimation procedure and the Prentice approach have a similar ability to detect surrogacy, with

neither approach providing reliable surrogacy conclusions.

We note that by basing surrogacy assessment on the criteria that $\gamma_0 = 0$, we do not avoid the problem in the Prentice criteria of proving a null hypothesis, namely that certain parameters assume the value of 0. Therefore, we can also examine the other proposed estimands to aid in validating S as a surrogate. The tests of $\rho_{ST} = 0$ and $CEP^U_{-2SD} < CEP^L_{+2SD}$ have similar power to correctly determine surrogacy, and are nearly equivalent to evaluating surrogacy based on the requirement of ACS in expectation that $\gamma_1 \neq 0$. The criterion of $\Phi_{10}(0) = 0.5$ being included in the 95% credible interval is equivalent to evaluating surrogacy based on the requirement of ACN in expectation that $\gamma_0 = 0$ and does reasonably well at determining surrogacy when the Prentice criteria and PS criteria are in agreement, but is unable to reliably distinguish valid principal surrogates from invalid ones with the two criteria disagree.

We perform additional simulations to assess the robustness of our procedure when joint normality does not hold. We consider three scenarios. In the first, joint normality does not hold, but each of the marginal distributions are normal. In the second, the joint distribution of the counterfactuals is a multivariate t_3 distribution, and in the third, each of the marginal distributions are lognormally distributed. For each of these scenarios, we generate multivariate normal data with the same mean and covariance structure as the non-normal data to compare the performance of the estimation procedure. When multivariate normality does not hold, the point estimates of the Prentice and PS parameters are nearly identical to the multivariate normal models, but with larger posterior standard deviations and lower coverage rates. There is little difference between the non-multivariate normal models and the multivariate normal models in terms of the assessment of S as a valid surrogate marker, indicating that the procedure is fairly robust to model misspecification. Appendix D provides details

of these simulations.

2.6 Applications

2.6.1 Visual acuity in age-related macular degeneration

We apply our estimation method to a clinical trial of interferon-α for 183 patients with age-related macular degeneration (Buyse, *et al.* 2000). These data come from a multicenter trial comprised of 36 different centers. The number of patients per center ranges from 2 to 18. The treatment indicator (Z_i) equals 0 for placebo and 1 for the treatment. The surrogate marker (S_i) is change in visual acuity at 6 months after starting treatment and the final endpoint (T_i) is change in visual acuity at 1 year. We subtract off the Best Linear Unbiased Predictor estimates from S_i and T_i to account for random center effects. Appendix E provides histograms and normal QQ plots of the observed data to assess the marginal normality of S and T in each treatment group and the bivariate normality of S and T within each treatment group. The assumption of marginal normality appears to hold, except potentially for T in the control group. Bivariate normality of S and T within each treatment group appears to hold approximately. The estimates used in assessing the Prentice criteria are as follows: $\hat{\theta}_1 = -3.34(SE = 2.13, P = 0.12)$, $\hat{\alpha}_1 = -2.03(SE = 1.90, P = 0.29)$, $\hat{\mu}_1 = 0.65(SE = 0.07, P < 0.0001)$, $\hat{\beta}_1 = -2.67(SE = 1.94, P = 0.17)$, $\hat{\beta}_2 = 0.69(SE = 0.09, P < 0.0001)$, and $\hat{\beta}_3 = -0.11(SE = 0.14, P = 0.44)$.

As θ_1 and α_1 are not statistically significant, the Prentice criteria are not met. Using our approach with Beta priors for the correlation parameters, we get the following posterior estimates for the principal surrogacy parameters of interest, $\gamma_0 = -1.62(-5.49, 2.16)$, and $\gamma_1 = 0.60(-0.24, 1.43)$. As γ_1 contains 0 within its 95%

credible interval, we conclude that change in visual acuity at 6 months is not a valid principal surrogate for change in visual acuity at 12 months. The average Pearson correlation, $\bar{\rho_{ST}}$ of $T_i(1) - T_i(0)$ and $S_i(1) - S_i(0)$ was 0.48 (-0.16, 0.92), also indicative of a poor principal surrogate. This is in agreement with the conclusion reached by Buyse, *et al.* (2000). Figure 2.3(a) shows a plot of the (CEP) curve, where $CEP = E[T(1) - T(0)|S(1) - S(0) = s]$ with a 95% credible interval for each value of s. The middle dashed line indicates the posterior mean of $\mu_{S_1} - \mu_{S_0}$, and the outer two dashed lines show the posterior means of $\mu_{S_1} - \mu_{S_0} \pm 2SD_{S(1)-S(0)}$, where $SD_{S(1)-S(0)}$ is the standard deviation of $S(1) - S(0)$, given by $\sqrt{\sigma^2_{S_0} + \sigma^2_{S_1} - 2\rho_s\sigma_{S_0}\sigma_{S_1}}$. The plot shows that 0 is contained within the credible interval at almost all values of s, indicating that there could be large effects of treatment on the surrogate with no expected effect of treatment on the outcome. Similarly, when there is no treatment effect on S, there could still be a sizeable treatment effect on T.

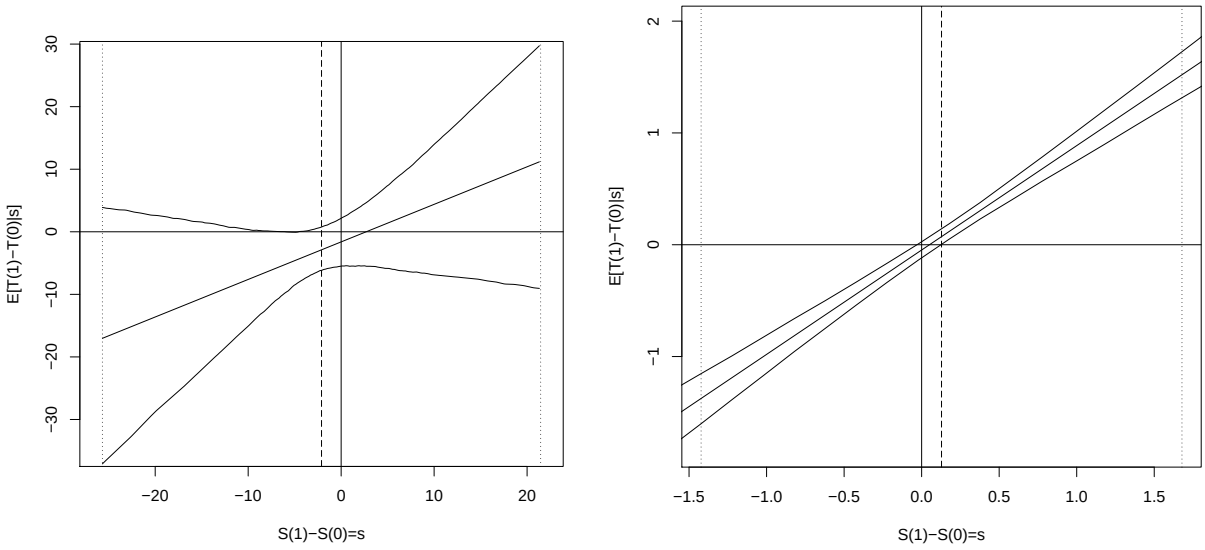
2.6.2 Progression free survival as a surrogate for overall survival in an ovarian cancer trial

This trial was analyzed by Buyse, *et al.* (2000) using a meta-analytic validation method. A total of 274 women were treated for ovarian cancer in two treatment arms. Of these patients, 201 experienced a clinical progression of the disease prior to death, and 43 died without a clinical disease progression. The remaining 30 patients were censored for death and not considered in the analysis. There are 126 subjects in the control arm and 118 in the treatment arm. The surrogate marker is progression free survival (PFS) time, in months and the final endpoint is overall survival (OS) time, in months. As both of these outcomes were right skewed, the fourth root of each was taken to approximately normalize the data, as shown in Appendix F.

Estimates of parameters used to assess the validity of the Prentice criteria are as follows: $\hat{\theta}_1 = 0.08(SE = 0.10, P = 0.41)$, $\hat{\alpha}_1 = 0.14(SE = 0.09, P = 0.14)$, $\hat{\mu}_1 = 0.95(SE = 0.02, P < 0.0001)$, $\hat{\beta}_1 = -0.12(SE = 0.13, P = 0.36)$, $\hat{\beta}_2 = 0.94(SE = 0.03, P < 0.0001)$, and $\hat{\beta}_3 = 0.02(SE = 0.04, P = 0.60)$. θ_1 and α_1 are not statistically significant, and the Prentice criteria are therefore unmet. The posterior estimates for the causal quantities of interest using Beta priors on the unidentified parameters gives $\gamma_0 = -0.05(-0.12, 0.03)$, and $\gamma_1 = 0.93(0.81, 1.07)$. The 95% credible interval for γ_0 contains 0 while the 95% credible interval for γ_1 does not and ρ_{ST}^{-} was 0.92 (0.85, 0.96). We therefore conclude that progression free survival time is a marginally valid principal surrogate for overall survival. This agrees with the findings of Buyse, *et al.* (2000). Figure 2.3(b) provides a plot of the CEP curve and 95% credible interval at each $S(1) - S(0) = s$, for the fourth roots of S and T. The middle and two outer dashed lines indicate the posterior mean and 95% credible interval of $\mu_{S_1} - \mu_{S_0}$, respectively. The plot shows that when there is no treatment effect on S, there is little or no expected treatment effect on T, and as the treatment effect on S increases, the treatment effect on T is also expected to increase.

2.7 Discussion

In this chapter, we develop a method for the assessment of surrogate markers within the principal surrogate framework. We assume a multivariate normal distribution for the potential surrogate outcomes and potential final outcomes and derive quantities that may be useful in determining the validity of a surrogate marker. Through our model setup, context specific assumptions can be incorporated into the prior distributions of unidentified parameters to aid in estimation. The estimation



(a) early change in visual acuity as a surrogate for late change in visual acuity

(b) PFS time as a surrogate for OS time

Figure 2.3: CEP curves for data examples
KEY: $-E[T(1)-T(0)|S(1)-S(0)=s]$ and 95% CI, $---(\mu_{S_1}-\mu_{S_0})$, $\cdots(\mu_{S_1}-\mu_{S_0})\pm 2SD(S(1)-S(0))$

procedure can be extended to scenarios where T is partially missing, or to the multiple trial setting.

We compare some of the proposed quantities for surrogate validation to the original validation criteria put forth by Prentice and show that, in many settings, we might expect the Prentice and principal surrogacy criteria to be in agreement. Based on our simulation study, it appears that when principal surrogacy is present, it is most accurately determined in cases where the Prentice criteria would also correctly identify surrogacy. When principal surrogacy is not present, it can be determined both when the Prentice criteria are able to correctly identify S as invalid and when the Prentice criteria incorrectly deem S to be valid. We note that even with the use of informative priors to aid in the estimation of the partially identified parameters, the coverage rates in many cases are not ideal. Due to the nonidentifiability of some parameters

in our model, certain assumptions on the relationships between nonidentifiable associations were made and informative priors were used for unidentified parameters to aid in estimation. The use of other priors or other context specific assumptions about parameters could be made. Zigler and Belin (2012) also explore the effects of various model assumptions in a principal surrogacy estimation procedure. They use a Bayesian estimation approach for the CEP surface when S is continuous and T is binary. In their procedure, priors are placed on the regression coefficients of the CEP surface, and an independence assumption is made for $T(1)$ and $T(0)$ conditional on the surrogate and other baseline covariates.

Each of the proposed quantities have merits and drawbacks in terms of their ability to characterize surrogacy. The proposed γ_0 and γ_1 quantities are easily interpretable, but proving that γ_0 is equal to 0, a necessary condition for a valid surrogate, is difficult to do in practice. The correlation measure, ρ_{ST}, captures the causal correlation between the treatment effect on the surrogate and the treatment effect on the outcome, but fails to capture the concept of ACN. The CEP graph provides a way to estimate expected treatment effects on T when treatment effects on S are at relevant clinical values, but does not offer a single summary of the value of S as a surrogate. Finally, the Φ_{10} quantity provides information about the entire conditional distribution, as opposed to just the expectation, but is more difficult to estimate and seems to have poor properties. While no single parameter estimate can completely assess principal surrogacy, a variety of measures that consider the distribution of the causal effect of treatment on the outcome conditional on the causal effect of treatment on the surrogate can be used in combination to provide evidence as to whether or not S is a valid surrogate for T.

Table 2.1: MVN model simulation results under different prior specifications

		S Valid PS, S Invalid Prentice				S Invalid PS, S Valid Prentice				S Invalid PS & Prentice				S Valid PS & Prentice			
Parameter	Prior Scenario	True Value	Mean (SD)	$\tilde{PSD}$	95% Coverage	True Value	Mean (SD)	$\tilde{PSD}$	95% Coverage	True Value	Mean (SD)	$\tilde{PSD}$	95% Coverage	True Value	Mean (SD)	$\tilde{PSD}$	95% Coverage
Identified Parameters																	
μ_{s_0}	4	4	3.99(0.08)	0.08	0.96	4	4.01(0.08)	0.08	0.95	4	3.99(0.08)	0.08	0.95	4	4.00(0.08)	0.08	0.98
μ_{s_1}	4	6	6.01(0.08)	0.08	0.95	6	5.99(0.08)	0.08	0.95	6	5.99(0.08)	0.08	0.95	6	5.99(0.09)	0.08	0.91
μ_{t_0}	4	7.8	7.78(0.08)	0.08	0.95	9	9.00(0.08)	0.08	0.96	8.5	8.49(0.07)	0.08	0.94	8.5	8.40(0.08)	0.08	0.95
μ_{t_1}	4	10	10.01(0.09)	0.08	0.91	10	10.00(0.08)	0.08	0.94	10	10.00(0.08)	0.08	0.95	10	10.00(0.09)	0.08	0.92
σ_{s_0}	1	1	0.99(0.06)	0.06	0.93	1	1.01(0.06)	0.06	0.96	1	1.00(0.06)	0.06	0.94	1	0.99(0.06)	0.06	0.96
σ_{s_1}	4	1	1.00(0.06)	0.06	0.93	1	1.01(0.05)	0.06	0.96	1	1.01(0.05)	0.06	0.98	1	0.99(0.06)	0.07	0.95
σ_{t_0}	4	1	1.00(0.06)	0.06	0.95	1	1.01(0.06)	0.06	0.97	1	1.01(0.06)	0.06	0.94	1	0.99(0.05)	0.06	0.96
σ_{t_1}	4	1	1.01(0.06)	0.06	0.94	1	1.00(0.06)	0.06	0.96	1	1.01(0.06)	0.06	0.96	1	0.99(0.06)	0.06	0.96
ρ_{00}	4	0.7	0.68(0.04)	0.04	0.99	0.5	0.48(0.06)	0.06	0.94	0.2	0.20(0.07)	0.07	0.95	0.8	0.79(0.03)	0.03	0.95
ρ_{11}	4	0.7	0.69(0.04)	0.04	0.97	0.5	0.49(0.06)	0.06	0.95	0.2	0.20(0.06)	0.07	0.97	0.8	0.78(0.03)	0.03	0.95
Unidentified Parameters																	
ρ_s	1	0.5	-0.35(0.23)	0.33	0.20	0.5	-0.22(0.22)	0.35	0.48	0.2	-0.15(0.23)	0.37	0.93	0.4	-0.35(0.24)	0.34	0.38
	2		0.32(0.08)	0.19	0.95		0.39(0.07)	0.22	1		0.37(0.08)	0.22	1		0.24(0.07)	0.15	0.91
	3		0.34(0.06)	0.13	0.82		0.45(0.05)	0.16	1		0.46(0.06)	0.21	1		0.22(0.04)	0.10	0.54
	4		0.47(0.07)	0.18	1		0.43(0.06)	0.20	1		0.34(0.06)	0.21	1		0.43(0.08)	0.16	0.995
ρ_{01}	1	0.15	-0.45(0.21)	0.29	0.51	0.45	-0.28(0.21)	0.33	0.42	0.04	-0.18(0.23)	0.35	0.97	0.32	-0.48(0.22)	0.29	0.29
	2		0.32(0.08)	0.19	0.99		0.39(0.07)	0.22	1		0.37(0.06)	0.22	0.97		0.24(0.07)	0.15	0.995
	3		0.14(0.04)	0.11	1		0.16(0.03)	0.10	0.04		0.06(0.02)	0.04	1		0.09(0.03)	0.07	0.25
	4		0.40(0.08)	0.20	0.95		0.28(0.07)	0.19	1		0.14(0.04)	0.15	1		0.40(0.09)	0.18	1
ρ_{10}	1	0.15	-0.37(0.21)	0.32	0.63	0.45	-0.28(0.24)	0.34	0.40	0.04	-0.16(0.23)	0.35	0.97	0.32	-0.39(0.21)	0.33	0.44
	2		0.34(0.08)	0.19	0.995		0.39(0.07)	0.22	1		0.37(0.07)	0.22	0.94		0.24(0.07)	0.15	1
	3		0.15(0.04)	0.11	1		0.16(0.02)	0.11	0.05		0.06(0.02)	0.04	1		0.10(0.03)	0.08	0.25
	4		0.42(0.08)	0.19	0.86		0.28(0.07)	0.19	0.995		0.14(0.03)	0.14	1		0.42(0.08)	0.17	0.99
ρ_t	1	0.18	-0.47(0.19)	0.27	0.43	0.5	-0.32(0.23)	0.32	0.27	0.3	-0.18(0.22)	0.36	0.82	0.4	-0.53(0.18)	0.28	0.14
	2		0.31(0.08)	0.19	1		0.37(0.07)	0.22	1		0.37(0.07)	0.22	1		0.24(0.06)	0.16	0.91
	3		0.32(0.05)	0.13	0.99		0.44(0.05)	0.16	1		0.46(0.06)	0.21	1		0.21(0.04)	0.09	0.42
	4		0.45(0.07)	0.19	0.91		0.42(0.06)	0.20	1		0.34(0.05)	0.21	1		0.42(0.08)	0.17	1
Simulation results: Bias, variability and coverage rate of surrogacy parameters																	
β_1	1	0.8	0.81(0.42)	0.51	0.99	0	0.06(0.60)	0.56	0.93	1.1	1.15(0.59)	0.59	0.96	0	0.05(0.36)	0.47	0.99
	2		0.82(0.45)	0.44	0.96		0.02(0.51)	0.52	0.96		1.18(0.54)	0.54	0.96		0.03(0.37)	0.37	0.95
	3		0.84(0.44)	0.43	0.94		0.04(0.56)	0.52	0.93		1.07(0.51)	0.55	0.98		-0.01(0.35)	0.07	0.96
	4		0.78(0.43)	0.44	0.94		0.004(0.51)	0.52	0.95		1.10(0.54)	0.54	0.96		0.04(0.37)	0.38	0.95
β_2	1	0.7	0.68(0.06)	0.07	0.96	0.5	0.49(0.08)	0.08	0.94	0.2	0.19(0.08)	0.08	0.96	0.8	0.78(0.05)	0.07	0.97
	2		0.69(0.06)	0.06	0.97		0.49(0.07)	0.07	0.96		0.21(0.07)	0.08	0.97		0.79(0.05)	0.05	0.96
	3		0.69(0.06)	0.06	0.96		0.49(0.08)	0.07	0.93		0.20(0.07)	0.07	0.94		0.78(0.05)	0.05	0.93
	4		0.68(0.06)	0.06	0.97		0.49(0.07)	0.07	0.94		0.20(0.07)	0.07	0.95		0.79(0.05)	0.05	0.97
β_3	1	0	0.004(0.08)	0.10	0.99	0	-0.008(0.12)	0.11	0.95	0	-0.0007(0.11)	0.11	0.96	0	-0.004(0.07)	0.09	0.98
	2		0.001(0.08)	0.09	0.96		0.002(0.10)	0.10	0.95		-0.01(0.10)	0.10	0.96		0.001(0.07)	0.07	0.94
	3		-0.002(0.08)	0.08	0.95		-0.001(0.11)	0.10	0.92		0.004(0.10)	0.11	0.97		0.007(0.07)	0.07	0.96
	4		0.010(0.08)	0.08	0.96		0.006(0.10)	0.10	0.95		0.003(0.10)	0.10	0.96		-0.002(0.07)	0.07	0.96
γ_0	1	0	0.54(0.23)	0.38	0.72	0.8	-0.33(0.33)	0.54	0.27	1.1	0.81(0.41)	0.66	0.97	0	-0.25(0.21)	0.33	0.96
	2		1.09(0.26)	0.52	0.49		0.64(0.28)	0.70	1		2.07(0.28)	0.65	0.69		0.12(0.17)	0.37	1
	3		0.54(0.15)	0.31	0.70		-0.30(0.23)	0.50	0.01		0.82(0.22)	0.45	1		-0.19(0.11)	0.20	0.92
	4		1.10(0.26)	0.63	0.60		0.18(0.26)	0.67	0.96		1.30(0.20)	0.48	0.99		0.20(0.22)	0.51	1
γ_1	1	1.1	0.83(0.11)	0.19	0.68	0.1	0.66(0.16)	0.26	0.27	0.2	0.34(0.20)	0.32	0.97	0.8	0.92(0.10)	0.16	0.94
	2		0.55(0.12)	0.26	0.48		0.28(0.13)	0.35	1		-0.29(0.12)	0.32	0.64		0.74(0.08)	0.18	1
	3		0.83(0.06)	0.15	0.63		0.65(0.10)	0.24	0		0.34(0.10)	0.22	1		0.90(0.04)	0.09	0.91
	4		0.55(0.12)	0.31	0.60		0.41(0.12)	0.33	0.97		0.11(0.08)	0.23	1		0.70(0.11)	0.25	1
ρ_{ST}	1	0.86	0.77 (0.06)	0.11	0.97	0.1	0.60 (0.11)	0.18	0.26	0.21	0.31 (0.16)	0.26	0.99	0.8	0.85 (0.04)	0.08	0.95
	2		0.53(0.10)	0.20	0.54		0.17 (0.10)	0.29	1		-0.27 (0.10)	0.27	0.63		0.73(0.06)	0.12	1
	3		0.81(0.04)	0.09	1		0.62 (0.06)	0.15	0		0.31 (0.07)	0.14	1		0.89 (0.02)	0.05	0.73
	4		0.52(0.09)	0.23	0.68		0.38(0.09)	0.26	0.97		0.10(0.07)	0.20	1		0.66(0.08)	0.17	0.99
CEP_{+2SD}	1	4.4	4.86 (0.39)	0.57	0.90	1.2	2.97 (0.49)	0.67	0.27	2.0	2.47 (0.52)	0.80	0.96	3.35	4.59 (0.37)	0.57	0.31
	2		3.45 (0.28)	0.50	0.46		1.37 (0.26)	0.64	1		0.89 (0.25)	0.61	0.56		3.40 (0.24)	0.38	0.99
	3		4.08 (0.21)	0.32	0.93		2.31 (0.20)	0.39	0		2.10 (0.18)	0.28	1		3.81 (0.18)	0.24	0.49
	4		3.31 (0.26)	0.55	0.36		1.82 (0.23)	0.58	0.96		1.74 (0.19)	0.45	1		3.02 (0.27)	0.45	0.96
CEP_{-2SD}	1	0	-0.47 (0.38)	0.57	0.87	0.8	-1.00 (0.47)	0.67	0.25	0.99	0.51 (0.52)	0.79	0.95	-0.15	-1.39 (0.40)	0.57	0.31
	2		0.96 (0.30)	0.50	0.44		0.63 (0.27)	0.64	1		2.11 (0.26)	0.61	0.58		-0.19 (0.24)	0.38	0.99
	3		0.31 (0.22)	0.32	0.91		-0.30 (0.23)	0.39	0.02		0.88 (0.18)	0.28	1		-0.62 (0.17)	0.24	0.45
	4		1.14 (0.26)	0.54	0.32		0.18 (0.25)	0.58	0.96		1.29 (0.20)	0.45	0.97		0.18 (0.25)	0.45	0.96
$\Phi_{10}(0)$	1	0.5	0.69 (0.08)	0.13	0.72	0.79	0.40 (0.09)	0.16	0.24	0.83	0.70 (0.10)	0.16	0.95	0.5	0.39 (0.08)	0.14	0.96
	2		0.84 (0.06)	0.13	0.49		0.70 (0.08)	0.20	1		0.95 (0.02)	0.07	0.69		0.55 (0.08)	0.18	1
	3		0.77 (0.07)	0.15	0.70		0.38 (0.08)	0.19	0.16		0.78 (0.06)	0.15	1		0.36 (0.07)	0.14	0.92
	4		0.85 (0.07)	0.16	0.60		0.56 (0.09)	0.23	0.96		0.86 (0.04)	0.10	1		0.58 (0.10)	0.23	1

1: No restrictions on ρ
2: $\rho \geq 0$
3: $\rho \geq 0$ and $\rho_{10}, \rho_{01} < \rho_s, \rho_t, \rho_{00}, \rho_{11}$
4: Beta priors

Table 2.2: MVN model simulation results: principal surrogacy assessment

Model	1	2	3	4
Truth				
PS satisfied	Yes	No	No	Yes
Prentice satisfied	No	Yes	No	Yes
Estimation Results				
γ_0 Not Rejected, Reject $\gamma_1 = 0$	0.37	0.15	0.01	0.94
$\gamma_0 = 0$ Not Rejected	0.60	1	0.20	1
Reject $\gamma_1 = 0$	0.57	0.15	0.01	0.94
Reject $\rho_{ST} = 0$	0.57	0.17	0.01	0.94
$CEP^U_{-2SD} < CEP^L_{+2SD}$	0.55	0.15	0.01	0.93
$\Phi_{10}(0) = 0.5$ Not Rejected	0.60	1	0.20	1
Prentice Criteria Not Rejected	0.52	0.92	0.26	0.95

CHAPTER III

Surrogacy Assessment Using Principal Stratification and a Gaussian Copula Model

3.1 Introduction

In Chapter II we described an approach for assessing surrogacy for multivariate normal distributions. Here, we extend these ideas by relaxing the multivariate normality assumption and consider scenarios in which the surrogate marker, S and the final outcome, T arise from non-normal distributions. A surrogate endpoint (S) is an intermediate outcome variable occurring in between the treatment (Z) and the outcome of interest (T). Surrogate markers offer the potential to run trials more cheaply and quickly by extracting information regarding the treatment effect on T through the earlier measured S, however, demonstrating the validity of a given surrogate for the outcome of interest can be difficult. Prentice (1989) proposed a formal definition of surrogacy along with a validation strategy, requiring that S and T be correlated and the treatment effect on T be fully captured by S. Other methods for surrogacy evaluation include the proportion of treatment effect explained by S (Freedman, *et al.* 1992), and individual-level and trial-level surrogacy association measures in meta-analyses (Buyse, *et al.* 2000). As these methods rely on estimating treatment effects

by adjusting for a variable measured after randomization, there may be unmeasured confounders in the pathway between the surrogate and final outcome. Thus, the resulting estimates may not have a causal interpretation (Rosenbaum, 1984). Therefore, much recent work has been done on the evaluation of surrogate endpoints using the "principal surrogacy" (PS) framework introduced by Frangakis and Rubin (2002) (henceforth FR). In this framework, each subject has two potential outcomes for each of the surrogate and final endpoints corresponding to each treatment, denoted by $S(Z)$ and $T(Z)$, for $Z = \{0, 1\}$. The principal surrogacy approach looks at the distribution of the potential outcomes of T conditional on principal strata based on the joint distribution of $S(0)$ and $S(1)$. The principal strata are unaffected by treatment, and are thus pre-randomization variables. Treatment effect estimates that condition on these principal strata are therefore causal estimates when treatments are randomly assigned.

Existing literature on methods for surrogacy assessment using the principal stratification approach has examined settings in which both S and T are binary (Li, *et al.* 2010), or in which S is continuous with binary T (Gilbert and Hudgens, 2008; Zigler and Belin, 2011). Work in the PS framework when both S and T are continuous has been discussed in the application to partial compliance (Schwartz, *et al.* 2011). Bartolucci and Grilli (2011) used a Plackett copula to model the joint counterfactual distribution of partial compliances, and then proposed separate models for the conterfacutal outcomes, conditional on compliance. In Chapter II, we explored the scenario where the joint distribution of the counterfactual observations of S and T is multivariate normal and proposed quantities derived from this distribution to assess surrogacy. We extend this work on surrogacy validation in the multivariate normal setting by relaxing the multivariate normality assumption and considering the sce-

nario in which S and T arise from non-normal distributions. Given the marginal distributions of $S(1)$, $S(0)$, $T(1)$, and $T(0)$, a Gaussian copula model can be used to obtain the joint distribution. In our proposed model, we explore the use of the Gaussian copula in the setting where S is a discrete ordinal random variable and T is a continuous time-to-event random variable. The values of $S(0)$ and $S(1)$ are assumed to arise from separate underlying latent normal random variables, denoted $\tilde{S}(0)$ and $\tilde{S}(1)$.

In our data example, we consider the use of the ordinal variable "tumor response" as a surrogate for overall survival in advanced colorectal cancer. Tumor response and overall survival are common endpoints of interest in cancer clinical trials, and there is a large literature on the use of tumor response as a surrogate marker for overall survival (Ellenberg and Hamilton, 1989; Torri, *et al.* 1992; Buyse and Piedbois, 1996). In this setting of mixed discrete and continuous outcomes, surrogacy validation methods have been explored by Molenberghs, *et al.* (2001), where a joint model for the underlying continuous latent variable of the observed discrete surrogate marker and the observed continuous final outcome was developed. The use of copula models in this setting has been explored in by Burzykowski, *et al.* (2004), who proposed a bivariate Plackett copula to jointly model tumor response and survival in advanced colorectal cancer, and assessed surrogacy using a meta-analytic approach. The application of a Gaussian copula model to jointly model bivariate discrete and continuous outcomes was examined by de Leon and Wu (2011). Here, we extend the use of this Gaussian copula model to a four dimensional model, two for the potential surrogate marker values under each treatment arm and two for the potential outcomes under each treatment arm, and derive quantities from it to determine surrogacy. Unlike in the multivariate normal setting explored in Chapter II, these principal

surrogacy measures will no longer be analytically estimable, but can be obtained from the posterior predictive distributions of the potential markers and outcomes under the Gaussian copula model. Because some parameters of the joint distribution are not fully identifiable from the data, we use a Bayesian estimation procedure with plausible prior distributions and some reasonable constraints on model parameters to reduce the non-identifiability problem of modeling counterfactual observations and to aid in estimation of the quantities of interest. In Section 3.2, we describe the model. Section 3.3 outlines the proposed surrogacy measures and Section 3.4 outlines the estimation strategy. Simulation results are presented in Section 3.5, and in Section 3.6 the estimation procedure is applied to data from a meta-analysis in advanced colorectal cancer. Section 3.7 concludes with a discussion.

3.2 The Model

3.2.1 Potential Outcomes

For a randomized trial with treatment assignment Z ($Z = 1$ or 0), surrogate marker S and true endpoint T, each subject i, $i = 1, ..., n$, has two potential outcomes for each of S_i and T_i, denoted by $S_i(Z_i)$ and $T_i(Z_i)$. Only one outcome, corresponding to the received treatment for subject i in each of the pairs $(S_i(0), S_i(1))$ and $(T_i(0), T_i(1))$ can be observed. The joint distribution of $(S_i(0), S_i(1), T_i(0), T_i(1))$ describes the causal associations between Z, S and T. We denote the marginal cumulative distributions of $S_i(0)$, $S_i(1)$, $T_i(0)$ and $T_i(1)$ by $F_{S_i(0)}$, $F_{S_i(1)}$, $F_{T_i(0)}$ and $F_{T_i(1)}$, respectively. We make the standard assumptions of ignorable treatment assignments (Rubin, 1978) and the stable unit treatment value assumption (SUTVA).

3.2.2 Copulas

Sklar (1959) provided the basis for multivariate modeling using copulas. A multivariate function $C = C(u_1, ..., u_k)$ is a copula if it is a continuous distribution function and each marginal is a uniform distribution function. That is, C is a mapping of $(0,1)^k \rightarrow (0,1)$, with $C(u) = p(U_i \leq u_1, ..., U_k \leq u_k)$, where each $U_i \sim Unif(0,1)$. Using known marginal distributions $F_1(y_1), ..., F_k(y_k)$, the function $C(F_1(y_1), ..., F_k(y_k)) = G(y)$ defines a joint distribution for $y_1, ..., y_k$ (Nelsen, 2006).

In this paper, we focus on the Gaussian copula, denoted as

$$C_\Phi(u|\Gamma) = \Phi_k\{\Phi^{-1}(u_1), ..., \Phi^{-1}(u_k)|\Gamma\}$$

where Φ is the standard normal cumulative distribution function and $\Phi_k(x|\Gamma)$ is a k-variate normal cumulative distribution function with covariance matrix Γ. The density of the Gaussian copula is given by

$$|\Gamma|^{-\frac{1}{2}} exp\{\tfrac{1}{2}\mathbf{q^T}(\mathbf{I_k} - \mathbf{\Gamma^{-1}})\mathbf{q}\}$$

where $\mathbf{q} = (q_1, ..., q_k)^T$ with $q_j = \Phi^{-1}(u_j)$, $\mathbf{q} \sim N_k(0, \Gamma)$ and Γ is a correlation matrix (Song, 2000). The copula framework can then be used to obtain a multivariate distribution with specified marginals.

3.2.3 Gaussian Copula Regression Model

In the setting of a single surrogate and single outcome, each measured at one time point, we have n observations each of dimension four, corresponding to the four potential outcomes for each subject. Let $\mathbf{y_i} = (S_i(0), S_i(1), T_i(0), T_i(1))$ represent the set of observations from subject i. For continuous S and T, the Gaussian copula

regression model can be obtained by taking $q_{ij} = h_{ij}(y_{ij};\theta_j) = \Phi^{-1}\{F_j(y_{ij};\theta_j)\}$ where θ_j is the parameter vector for marginal distribution j, where $j = 1, ..., 4$ corresponds to the four marginal distributions for $S_i(0)$, $S_i(1)$, $T_i(0)$, and $T_i(1)$, respectively. By this construction, we have:

$\mathbf{q_i} = [\Phi^{-1}\{F_1(S_i(0))\}, \Phi^{-1}\{F_2(S_i(1))\}, \Phi^{-1}\{F_3(T_i(0))\}, \Phi^{-1}\{F_4(T_i(1))\}] \sim N_4(\mathbf{0}, \Gamma)$

and the density of y_{ij} is given by:

$f(y_{ij}|\theta, \Gamma) = \prod_{i=1}^{n} \left[|\Gamma|^{-\frac{1}{2}} exp\{\frac{1}{2}\mathbf{q}_i(I_4 - \Gamma^{-1})\mathbf{q}_i^T\} f_1(S_i(0);\theta_1) f_2(S_i(1);\theta_2) f_3(T_i(0);\theta_3) f_4(T_i(1);\theta_4) \right]$

When the marginal distributions of S and T are integer valued, the density can be found by taking Radon-Nikodym derivative of $C_\Phi(u|\Gamma)$ with respect to counting measure (Song, 2000), so that

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{l_1=1}^{2}\sum_{l_2=1}^{2}\sum_{l_3=1}^{2}\sum_{l_4=1}^{2}(-1)^{l_1+l_2+l_3+l_4} C_\Phi(F_1(S(0))_{l_1}, F_2(S(1))_{l_2}, F_3(T(0))_{l_3}, F_4(T(1))_{l_4}|\Gamma)$$

where $F_j(y_j)_1 = F_j(y_j)$ and $F_j(y_j)_2 = F_j(y_j - 1)$.

In the setting that we consider, S is an ordinal categorical variable and T is a failure-time random variable. Let $V_i(Z) = \min(T_i(Z), W_i(Z))$ be the minimum of the observed failure time, $T_i(Z)$, and censoring time, $W_i(Z)$, and let $\Delta_i(Z) = I(W_i(Z) > T_i(Z))$ be the censoring indicator. Then, for each subject we have the observed data $S_i(Z)$, $V_i(Z)$, and $\Delta_i(Z)$. We make the ignorable censoring assumption, $T_i(Z) \perp W_i(Z)$, reasonable in cases where censoring is administrative and enrollment times are simultaneous or otherwise unrelated to the outcome. In our estimation procedure, we iteratively impute survival times for censored subjects so that each subject has the vector of outcomes $\mathbf{y}_i = (S_i(0), S_i(1), T_i(0), T_i(1))$. Let $\tilde{S}_i(Z)$ be a latent, Gaussian continuous random variable underlying the surrogate endpoint $S_i(Z)$ such that

$$S_i(Z) = \begin{cases} 1_Z, & \text{if } \tilde{S}_i(Z) \in (-\infty, \alpha_{1_Z}) \\ 2_Z, & \text{if } \tilde{S}_i(Z) \in (\alpha_{1_Z}, \alpha_{2_Z}) \\ \vdots & \\ M_Z, & \text{if } \tilde{S}_i(Z) \in (\alpha_{(M-1)_Z}, \infty) \end{cases}$$

where $\alpha_{1_Z} < \alpha_{2_Z} < \ldots < \alpha_{(M-1)_Z}$ are unknown cutpoints with $\alpha_{0_Z} = -\infty$ and $\alpha_{M_Z} = \infty$. We assume a cumulative probit model for the cutpoints of the underlying continuous random variables of $S(0)$ and $S(1)$ and a proportional hazards model with a Weibull baseline hazard function for the marginal distributions of $T(0)$ and $T(1)$. These models are given by:

$$\Phi^{-1}\{P(S_i(Z) \leq k_Z)\} = \alpha_{k_Z}$$
$$\lambda(T_i(Z)) = \left(\frac{\gamma_{T_Z}}{\lambda_{T_Z}}\right)\left(\frac{T_i(Z)}{\lambda_{T_Z}}\right)^{\gamma_{T_Z}-1}$$

where $\lambda(T_i(Z))$ is a hazard function for a Weibull distribution with scale parameter λ_{T_Z} and shape parameter γ_{T_Z}. Let $\tilde{\mathbf{y}}_\mathbf{i} = (\tilde{S}_i(0), \tilde{S}_i(1), T_i(0), T_i(1))$ represent the set of counterfactual latent surrogate and final outcomes for subject i. We assume that the joint cumulative distribution of $\tilde{S}_i(0)$, $\tilde{S}_i(1)$, $T_i(0)$, $T_i(1)$ is generated by the Gaussian copula function:

$F_{\tilde{\mathbf{y}}_\mathbf{i}}(\tilde{\mathbf{y}}_\mathbf{i}) = \Phi_4\{\Phi^{-1}(F_{\tilde{1}}(\tilde{S}_i(0))), \Phi^{-1}(F_{\tilde{2}}(\tilde{S}_i(1))), \Phi^{-1}(F_3(T_i(0))), \Phi^{-1}(F_4(T_i(1))) \mid \Gamma\}$,

where the subscripts $\tilde{1}$ and $\tilde{2}$ correspond to the CDF of the underlying latent variables of $S(0)$ and $S(1)$, respectively, Φ is the standard normal distribution and Φ_4 is the standard four-variate normal distribution with correlation matrix:

$$\Gamma = \begin{pmatrix} 1 & \rho_s & \rho_{00} & \rho_{01} \\ & 1 & \rho_{10} & \rho_{11} \\ & & 1 & \rho_t \\ & & & 1 \end{pmatrix}.$$

As $\tilde{S}_i(Z)$ is assumed to be Gaussian, the terms $\Phi^{-1}(F_{\tilde{1}}(\tilde{S}_i(0)))$ and $\Phi^{-1}(F_{\tilde{2}}(\tilde{S}_i(1)))$ are simply $\tilde{S}_i(0)$ and $\tilde{S}_i(1)$, respectively. The joint distribution of $S_i(0), S_i(1), T_i(0)$, and $T_i(1)$ under these distributional assumptions is then given by:

$$P(S_i(0) = k_0, S_i(1) = k_1, T_i(0) \leq t_0, T_i(1) \leq t_1)$$
$$= [F_{\tilde{\mathbf{y}}_\mathbf{i}}(\alpha_{k_0}, \alpha_{k_1}, t_0, t_1) - F_{\tilde{\mathbf{y}}_\mathbf{i}}(\alpha_{(k_0-1)}, \alpha_{k_1}, t_0, t_1) - F_{\tilde{\mathbf{y}}_\mathbf{i}}(\alpha_{k_0}, \alpha_{(k_1-1)}, t_0, t_1) + F_{\tilde{\mathbf{y}}_\mathbf{i}}(\alpha_{(k_0-1)}, \alpha_{(k_1-1)}, t_0, t_1)].$$

When both $T_i(0)$ and $T_i(1)$ are uncensored observations, the joint density of $S_i(0), S_i(1), T_i(0)$, and $T_i(1)$ is given by: $f_{\mathbf{y}_\mathbf{i}}(k_0, k_1, t_0, t_1) = \frac{\partial^2}{\partial t_0 \partial t_1} P(S_i(0) = k_0, S_i(1) = k_1, T_i(0) \leq t_0, T_i(1) \leq t_1)$, where the derivative of $F_{\tilde{\mathbf{y}}_\mathbf{i}}(\alpha_{k_0}, \alpha_{k_1}, t_0, t_1)$ with respect to t_1 and t_0 is given by:

$$\frac{\partial^2 F_{\tilde{\mathbf{y}}_\mathbf{i}}(\alpha_{k_0}, \alpha_{k_1}, t_0, t_1)}{\partial t_0 \partial t_1} = \frac{\phi\left(\frac{\Phi^{-1}(F_3(t_0)) - \rho_t \Phi^{-1}(F_4(t_1))}{\sqrt{1-\rho_t^2}}\right)}{\sqrt{1-\rho_t^2}} \Phi_2[S_{0|0:1}, S_{1|0:1}|\Gamma_{|0:1}] \frac{f_4(t_1) f_3(t_0)}{\phi(\Phi^{-1}(F_3(t_0)))} \text{ for}$$

$$S_{0|0:1} = \frac{\left(\frac{\alpha_{k_0} - \rho_{01}\Phi^{-1}(F_4(t_1))}{\sqrt{1-\rho_{01}^2}}\right) - \left(\frac{\rho_{00} - \rho_{01}\rho_t}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_t^2}}\right)\left(\frac{\Phi^{-1}(F_3(t_0)) - \rho_t \Phi^{-1}(F_4(t_1))}{\sqrt{1-\rho_t^2}}\right)}{\sqrt{1 - \left(\frac{\rho_{00} - \rho_{01}\rho_t}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_t^2}}\right)^2}}$$

$$S_{1|0:1} = \frac{\left(\frac{\alpha_{k_1} - \rho_{11}\Phi^{-1}(F_4(t_1))}{\sqrt{1-\rho_{11}^2}}\right) - \left(\frac{\rho_{10} - \rho_{11}\rho_t}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_t^2}}\right)\left(\frac{\Phi^{-1}(F_3(t_0)) - \rho_t \Phi^{-1}(F_4(t_1))}{\sqrt{1-\rho_t^2}}\right)}{\sqrt{1 - \left(\frac{\rho_{10} - \rho_{11}\rho_t}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_t^2}}\right)^2}}$$

where:

$$f_3(t_0) = \left(\frac{\gamma_{T_0}}{\lambda_{T_0}}\right)\left(\frac{t_0}{\lambda_{T_0}}\right)^{\gamma_{T_0}-1} exp\left(-\left(\frac{t_0}{\lambda_{T_0}}\right)^{\gamma_{T_0}}\right),$$
$$f_4(t_1) = \left(\frac{\gamma_{T_1}}{\lambda_{T_1}}\right)\left(\frac{t_1}{\lambda_{T_1}}\right)^{\gamma_{T_1}-1} exp\left(-\left(\frac{t_1}{\lambda_{T_1}}\right)^{\gamma_{T_1}}\right),$$
$$F_3(t_0) = 1 - exp\left(-\left(\frac{t_0}{\lambda_{T_0}}\right)^{\gamma_{T_0}}\right),$$
$$F_4(t_1) = 1 - exp\left(-\left(\frac{t_1}{\lambda_{T_1}}\right)^{\gamma_{T_1}}\right) \text{ and}$$

$$\Gamma_{|0:1} = \begin{pmatrix} 1 & \frac{\left(\frac{\rho_s - \rho_{11}\rho_{01}}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_{01}^2}}\right) - \left(\frac{\rho_{00} - \rho_{01}\rho_t}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_t^2}}\right)\left(\frac{\rho_{10} - \rho_t\rho_{11}}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_t^2}}\right)}{\sqrt{1 - \left(\frac{\rho_{00} - \rho_{01}\rho_t}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_t^2}}\right)^2}\sqrt{1 - \left(\frac{\rho_{10} - \rho_{11}\rho_t}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_t^2}}\right)^2}} \\ \frac{\left(\frac{\rho_s - \rho_{11}\rho_{01}}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_{01}^2}}\right) - \left(\frac{\rho_{00} - \rho_{01}\rho_t}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_t^2}}\right)\left(\frac{\rho_{10} - \rho_t\rho_{11}}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_t^2}}\right)}{\sqrt{1 - \left(\frac{\rho_{00} - \rho_{01}\rho_t}{\sqrt{1-\rho_{01}^2}\sqrt{1-\rho_t^2}}\right)^2}\sqrt{1 - \left(\frac{\rho_{10} - \rho_{11}\rho_t}{\sqrt{1-\rho_{11}^2}\sqrt{1-\rho_t^2}}\right)^2}} & 1 \end{pmatrix}.$$

The scale parameters, λ_{T_0} and λ_{T_1}, and shape parameters, γ_{T_0} and γ_{T_1} from the Weibull models as well as the cutpoints of the latent distributions for $S(0)$ and $S(1)$ are identifiable from the data. The correlation coefficients ρ_{00} and ρ_{11} are the Pearson

correlation coefficients between $\tilde{S}(0)$ and the normally transformed $T(0)$ and between $\tilde{S}(1)$ and the normally transformed $T(1)$, respectively, and can be seen as a proxy for the polyserial correlations between $T(0)$ and $S(0)$ and between $T(1)$ and $S(1)$ (de Leon and Wu, 2011). These polyserial correlations are estimable from the data (Olsson, *et al.* 1982). Because only one of the counterfactual pairs of outcomes is observed for each subject, ρ_s, ρ_t, ρ_{01}, and ρ_{10} are not identifiable. However, the identifiable correlation parameters together with the requirement that the correlation matrix be positive definite place boundary constraints on these non-identified parameters, which, along with other plausible assumptions that we can make, aids in their estimation.

3.2.4 Prior Distributional Assumptions

We place non-informative priors on the fully identified parameters λ_{T_Z}, γ_{T_Z}, and α_{k_Z}'s. Specifically, the priors for $\log(\lambda_{T_Z})$ and the α_{k_Z}'s are $N(0, 10^2)$ and the priors for γ_{T_Z} are gamma distributions with mean 1 and standard deviation 3. We place marginal priors on each of the correlation parameters in Γ and following Chapter II, we consider four different sets of prior assumptions. For each of these there is the

additional assumption that Γ must be positive definite. The four priors are

(a) Jointly uniform prior such that for each of the six correlations $p(\rho) \sim Unif(-1, 1)$

(b) Jointly uniform prior such that for each of the six correlations $p(\rho) \sim Unif(0, 1)$

(c) All $\rho's \geq 0, \rho_{01} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$, and $\rho_{10} < min(\rho_{00}, \rho_{11}, \rho_s, \rho_t)$

(d) Beta priors such that:

- $p(\rho_{11}) \sim Unif(0, 1)$
- $p(\rho_{00}) \sim Unif(0, 1)$
- $p(\rho_{10})$ and $p(\rho_{01}) \sim Beta(3\alpha_0, 3 - 3\alpha_0)$ such that $P(\rho_{01}, \rho_{10} \leq min(\tilde{\rho_{00}}, \tilde{\rho_{11}})) = 0.80$
- $p(\rho_s)$ and $p(\rho_t) \sim Beta(3\alpha_1, 3 - 3\alpha_1)$ such that $P(\rho_s, \rho_t \geq E(\rho_{10})) = 0.80$

where $\tilde{\rho}_{00}$ and $\tilde{\rho}_{11}$ are the polyserial correlation coefficients for $T(0)$, $S(0)$ and $T(1)$, $S(1)$, respectively, estimated from the observed data using the "polyserial" function in R and $E(\rho_{10}) = E(\rho_{01})$ is the mean under the $Beta(3\alpha_0, 3-3\alpha_0)$ distribution. Prior assumption (a) is a non-informative prior on all of the correlations. Under scenario (b), all correlations are constrained to be positive, a plausible assumption especially when $\tilde{\rho}_{00}$ and $\tilde{\rho}_{11}$ are positive. In scenario (c), in addition to the positivity assumption, we restrict ρ_{01} and ρ_{10} to be smaller than the other four correlation parameters. This seems reasonable as ρ_{01} and ρ_{10} are measures of the correlation between the surrogate response and final outcome response in opposite treatment arms, which is unlikely to be larger than the correlation between the surrogate response and final outcome response within the same treatment arm, or the correlation between the surrogate responses or final treatment responses across treatment arms. Finally, prior

assumption (d) places similar restrictions on the correlations as assumption (c), but is a little bit more flexible.

3.3 Measures of Surrogacy from Gaussian Copula Models

To determine the validity of S as a surrogate marker for T, we work within the principal surrogacy (PS) framework proposed by FR which uses a principal stratification approach to assess the validity of a surrogate marker. This framework focuses on the distribution of $p(T(0), T(1)|S(0), S(1))$. Since $S(1)$ and $S(0)$ are unaffected by treatment assignment, they can be treated as baseline covariates and quantities estimated by conditioning on them will always have a causal interpretation. This framework therefore avoids the potentially noncausal estimates that can result from surrogacy measures that condition on the observed post-randomization variable S. FR proposed two measures of surrogacy, the "associative effect" and the "dissociative effect". In our setting, a measure of the dissociative effect is given by $E(log(T_i(1)/T_i(0))|S_i(1) = S_i(0))$ and a measure of the associative effect is given by $E(log(T_i(1)/T_i(0))|S_i(1) \neq S_i(0))$. Values of the dissociative effect near zero indicate that the causal effect of treatment on the final outcome is near zero when the causal effect of treatment on the surrogate is near zero, a characteristic that a good principal surrogate should possess. When the dissociative effect is large, there can be a causal effect of the treatment on the final outcome even if there is no causal effect of the treatment on the surrogate. The value of the associative effect provides information on how the causal treatment effect on the outcome changes as the causal effect of the treatment on the surrogate changes. A good principal surrogate should result in a large associative effect, which would occur if as the treatment effect on the surrogate

increases, the treatment effect on the final outcome increases as well. Also of interest is the "causal effect predictiveness' (CEP) surface proposed by Gilbert and Hudgens (2008), which considers the entire curve of $\mathrm{E}[log(T(1))/log(T(0)) \mid S(1) - S(0)]$ and provides a measures of the treatment effect on T within subgroups defined by the treatment effect on the surrogate.

An additional useful measure to assess surrogacy is the correlation between the difference in the normal variables, $\Phi^{-1}(F_4(T(1))) - \Phi^{-1}(F_3(T(0)))$ and $\tilde{S}(1) - \tilde{S}(0)$. Measures of the associative effect, dissociative effect and correlation are not analytically estimable, but can be obtained from the posterior predictive distributions.

3.4 Estimation Procedure

A Bayesian approach is used to estimate parameters, using the prior assumptions detailed in Section 3.2.4. Unobserved potential outcomes are treated as missing data and imputed from the appropriate posterior distribution at each iteration of the Markov chain. Posterior estimates of the unobserved potential outcomes and parameter values are obtained using a Metropolis Hastings algorithm. When drawing each element of Γ, the range of possible values must first be determined in order to satisfy the positive definite requirement, given that the other correlations are held fixed. The range of values corresponding to a positive definite matrix are those in the interval determined by the roots of the quadratic equation that result from solving $|\Gamma| = 0$. Each iteration of the Markov chain is done as follows:

- Let $\alpha = (\alpha_{1_0}, \alpha_{2_0}, \alpha_{3_0}, \alpha_{1_1}, \alpha_{2_1}, \alpha_{3_1})$, $\theta = (\lambda_{T_0}, \lambda_{T_1}, \gamma_{T_0}, \gamma_{T_1})$,

  $\mathbf{q_i} = [\Phi^{-1}(F_{\tilde{S}_i(0)}(\tilde{S}(0))), \Phi^{-1}(F_{\tilde{S}_i(1)}(\tilde{S}(1))), \Phi^{-1}(F_{T_i(0)}(T(0))), \Phi^{-1}(F_{T_i(1)}(T(1)))]$

  $= [\tilde{S}(0), \tilde{S}(1), \Phi^{-1}(F_{T_i(0)}(T(0))), \Phi^{-1}(F_{T_i(1)}(T(1)))]$

$= [\tilde{S}(0), \tilde{S}(1), q_i(T_i(0)), q_i(T_i(1))]$

- Impute death times for censored subjects by drawing from

$P(q_i(T(Z)) > q_i(T(Z))+v \mid q_i(T(Z)) > q_i(T(Z)), q_i(T(1-Z)), \tilde{S}(0), \tilde{S}(1), \alpha, \theta) =$

$\frac{P(q_i(T(Z))>q_i(T(Z)))+v, q_i(T(1-Z)), \tilde{S}(0), \tilde{S}(1), \alpha, \theta)}{P(q_i(T(Z))>q_i(T(Z)), q_i(T(1-Z)), \tilde{S}(0), \tilde{S}(1), \alpha, \theta)}$

$= \frac{P(q_i(T(Z))>q_i(T(Z))+v | q_i(T(1-Z)), \tilde{S}(0), \tilde{S}(1), \alpha, \theta)}{P(q_i(T(Z))>q_i(V(Z)) | q_i(T(1-Z)), \tilde{S}(0), \tilde{S}(1), \alpha, \theta)}$.

For each censored subject, we draw a uniform random variable $u \sim Unif(0,1)$ and solve $g^{-1}(u) = v$, where $g(v) = \frac{P(q_i(T(Z))>q_i(V(Z))+v | q_i(T(1-Z)), \tilde{S}(0), \tilde{S}(1), \alpha, \theta)}{P(q_i(T(Z))>q_i(V(Z)) | q_i(T(1-Z)), \tilde{S}(0), \tilde{S}(1), \alpha, \theta)}$. This can be solved analytically for v by $v = F^{-1}[1 - u(1 - F\{q_i(T(Z))\})] - q_i(T(Z))$ where F is the CDF of the conditional distribution with mean given by:

$$\begin{pmatrix} \rho_{0Z} & \rho_{1Z} & \rho_t \end{pmatrix} \begin{pmatrix} 1 & \rho_s & \rho_{0(1-Z)} \\ \rho_s & 1 & \rho_{1(1-Z)} \\ \rho_{0(1-Z)} & \rho_{1(1-Z)} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{S}_i(0) \\ \tilde{S}_i(1) \\ q_i(T_i(1-Z)) \end{pmatrix}$$ and variance given by:

$$1 - \begin{pmatrix} \rho_{0Z} & \rho_{1Z} & \rho_t \end{pmatrix} \begin{pmatrix} 1 & \rho_s & \rho_{0(1-Z)} \\ \rho_s & 1 & \rho_{1(1-Z)} \\ \rho_{0(1-Z)} & \rho_{1(1-Z)} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{0Z} \\ \rho_{1Z} \\ \rho_t \end{pmatrix}.$$

- Draw missing counterfactual observations of $\tilde{S}_i(Z)$ and $q_i(T(Z))$ from their conditional distributions:

$$\begin{pmatrix} \tilde{S}_i(1) \\ q_i(T_i(1)) \end{pmatrix} \mid \tilde{S}_i(0),\, q_i(T_i(0)),\, \Gamma$$

$$\sim N\left(\begin{pmatrix} \rho_s & \rho_{10} \\ \rho_{01} & \rho_t \end{pmatrix} \begin{pmatrix} 1 & \rho_{00} \\ \rho_{00} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{S}_i(0) \\ q_i(T_i(0)) \end{pmatrix}, \begin{pmatrix} 1 & \rho_{11} \\ \rho_{11} & 1 \end{pmatrix} - \begin{pmatrix} \rho_s & \rho_{10} \\ \rho_{01} & \rho_t \end{pmatrix} \begin{pmatrix} 1 & \rho_{00} \\ \rho_{00} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_s & \rho_{01} \\ \rho_{10} & \rho_t \end{pmatrix} \right)$$

$$\begin{pmatrix} \tilde{S}_i(0) \\ q_i(T_i(0)) \end{pmatrix} \mid \tilde{S}_i(1),\, q_i(T_i(1)),\, \Gamma$$

$$\sim N\left(\begin{pmatrix} \rho_s & \rho_{01} \\ \rho_{10} & \rho_t \end{pmatrix} \begin{pmatrix} 1 & \rho_{11} \\ \rho_{11} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{S}_i(1) \\ q_i(T_i(1)) \end{pmatrix}, \begin{pmatrix} 1 & \rho_{00} \\ \rho_{00} & 1 \end{pmatrix} - \begin{pmatrix} \rho_s & \rho_{01} \\ \rho_{10} & \rho_t \end{pmatrix} \begin{pmatrix} 1 & \rho_{11} \\ \rho_{11} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_s & \rho_{10} \\ \rho_{01} & \rho_t \end{pmatrix} \right)$$

- For unobserved counterfactuals, transform draws to $S_i(Z)$, and $T_i(Z)$:

$[\tilde{S}_i(Z) \leq \alpha_{1_Z}] \rightarrow S_i(Z) = 1$

$[\alpha_{1_Z} < \tilde{S}_i(Z) \leq \alpha_{2_Z}] \rightarrow S_i(Z) = 2$

$[\alpha_{2_Z} < \tilde{S}_i(Z) \leq \alpha_{3_Z}] \rightarrow S_i(Z) = 3$

$[\tilde{S}_i(Z) > \alpha_{3_Z}] \rightarrow S_i(Z) = 4$

$T_i(Z) = \lambda_{T_Z}[-\log(1 - \Phi\{q_i(T_i(Z))\}]^{1/\gamma_{T_Z}}$

- Draw α's from posterior distribution using a Metropolis Hastings algorithm.

- For observed $S_i(Z)$'s, draw $\tilde{S}_i(Z)$'s from a truncated normal distribution, where:

$$[\tilde{S}_i(0)|\tilde{S}_i(1), q_i(T_i(0)), q_i(T_i(1)), \Gamma, \alpha, Z_i = 0] \sim$$
$$\mathrm{N}\left(\begin{pmatrix}\rho_s & \rho_{00} & \rho_{01}\end{pmatrix}\begin{pmatrix}1 & \rho_{10} & \rho_{11}\\ \rho_{10} & 1 & \rho_t\\ \rho_{11} & \rho_t & 1\end{pmatrix}^{-1}\begin{pmatrix}\tilde{S}_i(1)\\ q_i(T_i(0))\\ q_i(T_i(1))\end{pmatrix}, 1 - \begin{pmatrix}\rho_s & \rho_{00} & \rho_{01}\end{pmatrix}\begin{pmatrix}1 & \rho_{10} & \rho_{11}\\ \rho_{10} & 1 & \rho_t\\ \rho_{11} & \rho_t & 1\end{pmatrix}^{-1}\begin{pmatrix}\rho_s\\ \rho_{00}\\ \rho_{01}\end{pmatrix}\right) I(\alpha_{(k_0-1)} < \tilde{S}_i(0) \leq \alpha_{k_0})$$

$$[\tilde{S}_i(1)|\tilde{S}_i(0), q_i(T_i(0)), q_i(T_i(1)), \Gamma, \alpha, Z_i = 1] \sim$$
$$\mathrm{N}\left(\begin{pmatrix}\rho_s & \rho_{10} & \rho_{11}\end{pmatrix}\begin{pmatrix}1 & \rho_{00} & \rho_{01}\\ \rho_{00} & 1 & \rho_t\\ \rho_{01} & \rho_t & 1\end{pmatrix}^{-1}\begin{pmatrix}\tilde{S}_i(0)\\ q_i(T_i(0))\\ q_i(T_i(1))\end{pmatrix}, 1 - \begin{pmatrix}\rho_s & \rho_{10} & \rho_{11}\end{pmatrix}\begin{pmatrix}1 & \rho_{00} & \rho_{01}\\ \rho_{00} & 1 & \rho_t\\ \rho_{01} & \rho_t & 1\end{pmatrix}^{-1}\begin{pmatrix}\rho_s\\ \rho_{10}\\ \rho_{11}\end{pmatrix}\right) I(\alpha_{(k_1-1)} < \tilde{S}_i(1) \leq \alpha_{k_1})$$

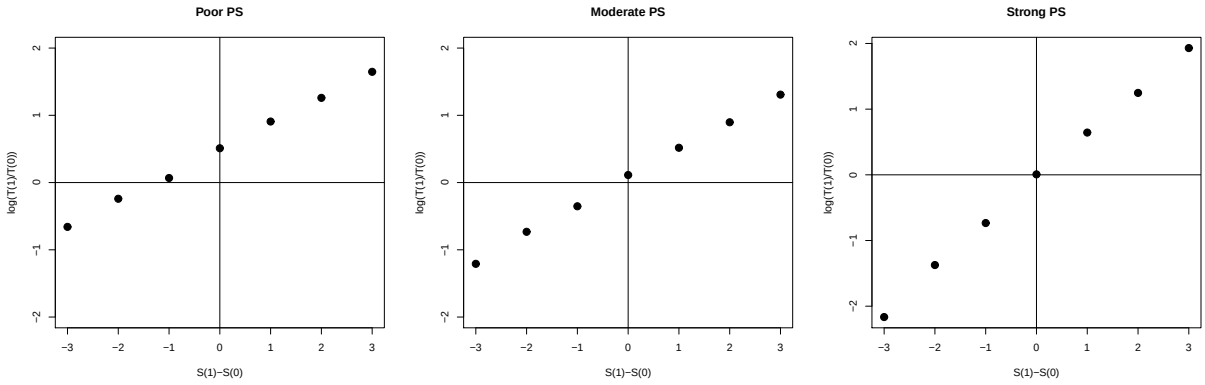
- Use the Metropolis Hastings algorithm to draw $\lambda_{T_Z}, \gamma_{T_Z}, \rho_s, \rho_{00}, \rho_{01}, \rho_{10}, \rho_{11}, \rho_t$ from their posterior distributions. The posterior distributions for all of the parameters can be obtained from the product of the observed data likelihood, detailed in Section 3.2.3 and the prior distributions, described in Section 3.2.4. The chain is run for a 3,000 iteration burn-in period, and then 2,000 draws from the posterior distribution of each parameter are saved. All of the proposal distributions are normal and centered at the most recent parameter draw. The proposal distribution for γ_{T_Z} is truncated at 0 and proposal distribution for each α_{kZ} is truncated by $\alpha_{(k-1)Z}$ and $\alpha_{(k+1)Z}$. The proposal distribution for each of the correlation parameters is truncated by the bounds which results in

a positive definite matrix. For each parameter, the variance of the proposal distribution is adjusted so that the resulting acceptance rates are close to 40%.

3.5 Simulations

We conduct simulations to evaluate the performance of the above methods of surrogacy assessment. We consider three scenarios: one where S is a good principal surrogate for T, one where it is a moderately good principal surrogate, and one where it is a poor principal surrogate. In each scenario, a sample size of 300 is used with 150 subjects in each treatment arm and approximately 30% of the survival outcomes are censored. We first explore the sensitivity of the estimation to the plausible prior restrictions on Γ that we might make. Figure 3.1 provides plots of the true relationship between $E[log(T(1))/log(T(0)) \mid S(1) - S(0) = s]$ for the three surrogacy scenarios considered. In the case of a poor principal surrogate, the plot shows that $E[log(T(1))/log(T(0)) \mid S(1) - S(0) = 0]$ is greater than 0, indicating the average causal necessity is not met. In the moderate principal surrogate case, average causal necessity is close to being met and there is a moderate increasing trend in $E[log(T(1))/log(T(0)) \mid S(1) - S(0) = s]$ as $S(1) - S(0)$ increases. For the strong principal surrogate case, average causal necessity is met and there is a strong increasing trend in $E[log(T(1))/log(T(0)) \mid S(1) - S(0) = s]$ as $S(1) - S(0)$ increases. For each of the three different surrogacy scenarios we perform four simulations, with the estimation procedure done using each of the priors outlined in Section 3.2.4. This results in a total of 12 simulations, each with 200 simulated data sets. Tables 3.1 and 3.2 provide the posterior means and standard deviations of the Bayesian estimates and means of the posterior standard deviations ($P\bar{S}D$). The

Figure 3.1: Plots of $E[log(T(1))/log(T(0)) \mid S(1) - S(0) = s]$ for the three simulation scenarios



identified parameters are not sensitive to changes in the prior specifications, while the unidentified parameters are quite sensitive to prior assumptions. The standard deviation of the Bayesian estimates is generally smaller than $P\bar{S}D$ for the unidentified parameters, leading to overcoverage of some of the unidentified quantities. Table 3.3 provides the means and standard deviations of the Bayesian estimates and $P\bar{S}D$ for the causal quantities of interest, $E(log(T_i(1)/T_i(0)) \mid S_i(1) - S_i(0) = s)$ and $\text{cor}(\Phi^{-1}(F_4(T(1))) - \Phi^{-1}(F_3(T(0))), \tilde{S}(1) - \tilde{S}(0))$. There is some bias in estimating these quantities, as these depend on the unidentified parameters. Prior scenarios 3 and 4 appear to perform better than scenarios 1 and 2 in terms of bias and coverage rates, generally maintaining conservative coverage and small to moderate biases across all surrogate scenarios. Under both of these priors, the estimation procedure does reasonably well at distinguishing the validity of S as a principal surrogate. The estimates of $E(log(T_i(1)/T_i(0)) \mid S_i(1) - S_i(0) = 0)$ are near 0 in the case of a moderate or strong PS, and larger when S is a poor principal surrogate. The estimated correlation of the causal treatment effects on S and T is largest when S is a strong principal surrogate and smallest when it is a poor principal surrogate.

Table 3.1: Copula model simulation results under different prior specifications-indentified parameters

		Moderate PS			Poor PS			Strong PS		
Parameter	Prior Scenario	True Value	Mean (SD)	$P\bar{S}D$	True Value	Mean (SD)	$P\bar{S}D$	True Value	Mean (SD)	$P\bar{S}D$
$\log(\lambda_{t_0})$	1[1]	2	2.01 (0.08)	0.08	2	2.01 (0.08)	0.08	2	2.00 (0.08)	0.08
	2[2]		2.00 (0.08)	0.08		2.01 (0.09)	0.08		2.01 (0.09)	0.08
	3[3]		2.00 (0.09)	0.08		2.01 (0.08)	0.08		2.01 (0.08)	0.08
	4[4]		2.02 (0.08)	0.08		2.01 (0.08)	0.08		2.00 (0.08)	0.08
γ_{t_0}	1	1.2	1.19 (0.09)	0.10	1.2	1.21 (0.10)	0.10	1.2	1.22 (0.11)	0.11
	2		1.21 (0.10)	0.10		1.18 (0.10)	0.10		1.20 (0.11)	0.10
	3		1.20 (0.11)	0.10		1.20 (0.09)	0.10		1.21 (0.10)	0.10
	4		1.22 (0.10)	0.10		1.20 (0.10)	0.10		1.21 (0.11)	0.10
$\log(\lambda_{t_1})$	1	2.3	2.32 (0.10)	0.10	2.5	2.51 (0.11)	0.11	2.3	2.31 (0.09)	0.09
	2		2.32 (0.10)	0.10		2.51 (0.11)	0.11		2.31 (0.09)	0.09
	3		2.31 (0.10)	0.09		2.52 (0.11)	0.11		2.31 (0.09)	0.10
	4		2.31 (0.09)	0.09		2.52 (0.12)	0.11		2.31 (0.09)	0.10
γ_{t_1}	1	1.2	1.21 (0.12)	0.12	1.2	1.20 (0.14)	0.13	1.2	1.23 (0.13)	0.12
	2		1.20 (0.12)	0.12		1.20 (0.13)	0.13		1.23 (0.12)	0.12
	3		1.22 (0.12)	0.12		1.20 (0.12)	0.13		1.22 (0.11)	0.12
	4		1.21 (0.11)	0.12		1.20 (0.13)	0.13		1.21 (0.11)	0.12
α_{01}	1	-0.67	-0.70 (0.11)	0.11	-0.67	-0.68 (0.12)	0.11	-0.67	-0.69 (0.10)	0.11
	2		-0.70 (0.12)	0.11		-0.68 (0.11)	0.11		-0.69 (0.11)	0.11
	3		-0.68 (0.10)	0.11		-0.69 (0.11)	0.11		-0.70 (0.11)	0.11
	4		-0.70 (0.11)	0.11		-0.68 (0.11)	0.11		-0.69 (0.12)	0.11
α_{02}	1	0	-0.008 (0.10)	0.10	0	0.004 (0.10)	0.10	0	0.003 (0.09)	0.10
	2		0.001 (0.10)	0.10		-0.002 (0.10)	0.10		-0.0008 (0.10)	0.10
	3		0.009 (0.11)	0.10		-0.01 (0.10)	0.10		-0.009 (0.10)	0.10
	4		-0.006 (0.10)	0.10		0.007 (0.10)	0.10		0.009 (0.11)	0.10
α_{03}	1	0.67	0.69 (0.11)	0.11	0.67	0.69 (0.11)	0.11	0.67	0.69 (0.11)	0.11
	2		0.69 (0.13)	0.11		0.68 (0.12)	0.11		0.69 (0.12)	0.11
	3		0.70 (0.11)	0.11		0.69 (0.11)	0.11		0.68 (0.12)	0.11
	4		0.68 (0.11)	0.11		0.68 (0.12)	0.11		0.69 (0.12)	0.11
α_{11}	1	-1.28	-1.32 (0.13)	0.14	-0.67	-0.69 (0.11)	0.11	-1.28	-1.34 (0.16)	0.14
	2		-1.32 (0.14)	0.14		-0.69 (0.11)	0.11		-1.32 (0.13)	0.14
	3		-1.33 (0.15)	0.14		-0.69 (0.10)	0.11		-1.34 (0.14)	0.14
	4		-1.31 (0.14)	0.14		-0.68 (0.10)	0.11		-1.31 (0.15)	0.14
α_{12}	1	-0.52	-0.53 (0.11)	0.11	0	0.003 (0.10)	0.10	-0.52	-0.53 (0.10)	0.10
	2		-0.53 (0.10)	0.11		0.006 (0.10)	0.10		-0.54 (0.11)	0.10
	3		-0.54 (0.11)	0.11		-0.02 (0.10)	0.10		-0.54 (0.11)	0.10
	4		-0.53 (0.11)	0.11		-0.003 (0.10)	0.10		-0.54 (0.11)	0.11
α_{13}	1	0.25	0.26 (0.10)	0.10	0.67	0.69 (0.11)	0.11	0.25	0.26 (0.10)	0.10
	2		0.27 (0.10)	0.10		0.68 (0.11)	0.11		0.26 (0.10)	0.10
	3		0.25 (0.10)	0.10		0.67 (0.12)	0.11		0.26 (0.10)	0.10
	4		0.26 (0.11)	0.10		0.67 (0.11)	0.11		0.25 (0.10)	0.10
ρ_{00}	1	0.6	0.58 (0.06)	0.06	0.2	0.18 (0.09)	0.08	0.8	0.78 (0.04)	0.04
	2		0.58 (0.07)	0.06		0.20 (0.08)	0.08		0.78 (0.04)	0.04
	3		0.57 (0.07)	0.06		0.20 (0.08)	0.09		0.78 (0.04)	0.04
	4		0.58 (0.06)	0.06		0.21 (0.08)	0.09		0.79 (0.04)	0.04
ρ_{11}	1	0.6	0.57 (0.07)	0.07	0.2	0.18 (0.09)	0.09	0.8	0.78 (0.04)	0.04
	2		0.58 (0.06)	0.06		0.20 (0.09)	0.09		0.78 (0.04)	0.04
	3		0.57 (0.07)	0.06		0.20 (0.09)	0.09		0.78 (0.04)	0.04
	4		0.58 (0.06)	0.06		0.20 (0.09)	0.09		0.78 (0.04)	0.04

1: No restrictions on ρ
2: $\rho \geq 0$
3: $\rho \geq 0$ and $\rho_{10}, \rho_{01} < \rho_s, \rho_t, \rho_{00}, \rho_{11}$
4: Beta priors

3.6 Application

We apply our estimation method to data from six clinical trials in advanced colorectal cancer (Meta-analysis Group in Cancer, 1998) to determine whether cancer

Table 3.2: Copula model simulation results under different prior specifications-unindentified parameters

		Moderate PS			Poor PS			Strong PS		
Parameter	Prior Scenario	True Value	Mean (SD)	$P\bar{S}D$	True Value	Mean (SD)	$P\bar{S}D$	True Value	Mean (SD)	$P\bar{S}D$
ρ_s	1	0.4	0.0006 (0.35)	0.28	0.2	0.004 (0.24)	0.35	0.4	0.08 (0.37)	0.19
	2		0.48 (0.12)	0.21		0.41 (0.12)	0.23		0.47 (0.16)	0.17
	3		0.51 (0.08)	0.15		0.48 (0.09)	0.22		0.49 (0.14)	0.12
	4		0.48 (0.09)	0.20		0.22 (0.07)	0.22		0.57 (0.13)	0.15
ρ_{01}	1	0.3	0.005 (0.34)	0.27	0.04	-0.04 (0.30)	0.33	0.32	0.09 (0.37)	0.18
	2		0.46 (0.15)	0.20		0.41 (0.15)	0.21		0.47 (0.18)	0.16
	3		0.23 (0.06)	0.12		0.05 (0.03)	0.06		0.33 (0.14)	0.11
	4		0.35 (0.12)	0.19		0.12 (0.06)	0.14		0.53 (0.14)	0.15
ρ_{10}	1	0.3	-0.003 (0.34)	0.26	0.04	0.03 (0.31)	0.32	0.32	0.07 (0.37)	0.18
	2		0.47 (0.14)	0.20		0.39 (0.13)	0.21		0.49 (0.17)	0.16
	3		0.24 (0.06)	0.12		0.05 (0.03)	0.06		0.33 (0.14)	0.11
	4		0.35 (0.11)	0.19		0.12 (0.05)	0.14		0.54 (0.15)	0.15
ρ_t	1	0.4	-0.002 (0.36)	0.26	0.3	-0.01 (0.34)	0.30	0.4	0.08 (0.39)	0.17
	2		0.46 (0.14)	0.19		0.41 (0.16)	0.20		0.48 (0.18)	0.15
	3		0.52 (0.09)	0.15		0.49 (0.11)	0.21		0.48 (0.15)	0.11
	4		0.47 (0.12)	0.18		0.34 (0.09)	0.21		0.57 (0.15)	0.13

1: No restrictions on ρ
2: $\rho \geq 0$
3: $\rho \geq 0$ and $\rho_{10}, \rho_{01} < \rho_s, \rho_t, \rho_{00}, \rho_{11}$
4: Beta priors

Table 3.3: Copula model simulation results: bias, variability and coverage rate of surrogacy parameters

		Moderate PS				Poor PS				Strong PS			
Quantity	Prior Scenario	True Value	Mean (SD)	$P\bar{S}D$	95% Coverage	True Value	Mean (SD)	$P\bar{S}D$	95% Coverage	True Value	Mean (SD)	$P\bar{S}D$	95% Coverage
$E[\log(T(1)/T(0))\|S(1)-S(0)=-3]$	1	-1.18	-1.60 (0.82)	0.96	0.86	-0.02	-0.02 (0.74)	0.90	0.92	-2.10	-2.18 (0.77)	0.78	0.91
	2		-0.37 (0.60)	1.08	0.94		1.32 (0.45)	0.84	0.65		-1.40 (0.69)	0.93	0.87
	3		-1.66 (0.34)	0.94	1		-0.17 (0.24)	0.65	1		-2.28 (0.40)	0.79	0.995
	4		-0.99 (0.41)	1.07	0.99		0.18 (0.29)	0.75	1		-1.33 (0.62)	0.95	0.88
$E[\log(T(1)/T(0))\|S(1)-S(0)=-2]$	1	-0.73	-0.98 (0.54)	0.66	0.89	0.19	0.19 (0.48)	0.58	0.94	-1.39	-1.33 (0.51)	0.48	0.92
	2		-0.17 (0.47)	0.81	0.94		1.04 (0.32)	0.56	0.64		-0.91 (0.50)	0.63	0.88
	3		-1.16 (0.25)	0.60	0.99		-0.02 (0.22)	0.44	0.99		-1.47 (0.26)	0.44	0.995
	4		-0.68 (0.32)	0.77	0.995		0.27 (0.23)	0.47	0.995		-0.88 (0.44)	0.67	0.90
$E[\log(T(1)/T(0))\|S(1)-S(0)=-1]$	1	-0.35	-0.47 (0.35)	0.41	0.93	0.33	0.34 (0.28)	0.34	0.96	-0.73	-0.68 (0.32)	0.29	0.91
	2		0.006 (0.32)	0.49	0.91		0.80 (0.22)	0.32	0.73		-0.44 (0.33)	0.36	0.87
	3		-0.63 (0.18)	0.34	0.98		0.22 (0.17)	0.24	0.98		-0.79 (0.17)	0.23	0.98
	4		-0.33 (0.22)	0.46	0.995		0.37 (0.16)	0.27	0.995		-0.43 (0.28)	0.38	0.88
$E[\log(T(1)/T(0))\|S(1)-S(0)=0]$	1	0.11	0.07 (0.15)	0.18	0.97	0.49	0.49 (0.14)	0.19	0.99	0.01	-0.001 (0.14)	0.14	0.96
	2		0.23 (0.14)	0.17	0.95		0.51 (0.16)	0.16	0.94		0.10 (0.14)	0.14	0.88
	3		0.06 (0.11)	0.14	0.98		0.51 (0.13)	0.15	0.98		0.004 (0.10)	0.12	0.96
	4		0.13 (0.11)	0.16	0.995		0.50 (0.13)	0.15	0.98		0.11 (0.12)	0.14	0.90
$E[\log(T(1)/T(0))\|S(1)-S(0)=1]$	1	0.52	0.55 (0.14)	0.21	0.99	0.69	0.65 (0.28)	0.34	0.95	0.64	0.61 (0.14)	0.16	0.98
	2		0.41 (0.16)	0.20	0.94		0.22 (0.23)	0.32	0.73		0.55 (0.14)	0.16	0.92
	3		0.60 (0.13)	0.16	0.96		0.79 (0.16)	0.24	0.995		0.66 (0.10)	0.13	0.98
	4		0.49 (0.13)	0.19	0.99		0.62 (0.16)	0.28	0.995		0.53 (0.14)	0.16	0.91
$E[\log(T(1)/T(0))\|S(1)-S(0)=2]$	1	0.89	1.03 (0.30)	0.37	0.96	0.83	0.80 (0.48)	0.57	0.94	1.24	1.22 (0.28)	0.26	0.91
	2		0.58 (0.31)	0.43	0.90		-0.03 (0.33)	0.57	0.66		0.99 (0.29)	0.32	0.87
	3		1.13 (0.17)	0.29	0.96		1.03 (0.21)	0.44	0.995		1.30 (0.17)	0.22	0.98
	4		0.85 (0.21)	0.40	0.99		0.73 (0.22)	0.47	1		0.95 (0.26)	0.33	0.87
$E[\log(T(1)/T(0))\|S(1)-S(0)=3]$	1	1.31	1.69 (0.57)	0.63	0.81	1.01	0.97 (0.74)	0.86	0.94	1.95	2.06 (0.55)	0.45	0.80
	2		0.78 (0.48)	0.75	0.95		-0.30 (0.44)	0.83	0.66		1.49 (0.47)	0.58	0.88
	3		1.72 (0.25)	0.54	0.98		1.18 (0.22)	0.66	1		2.03 (0.30)	0.42	0.98
	4		1.25 (0.33)	0.73	0.99		0.83 (0.28)	0.74	1		1.42 (0.42)	0.62	0.88
$\text{cor}(q(T(1))-q(T(0)), \bar{S}(1)-\bar{S}(0))$	1	0.50	0.55 (0.20)	0.20	0.81	0.22	0.18 (0.26)	0.28	0.91	0.80	0.74 (0.14)	0.13	0.96
	2		0.22 (0.21)	0.31	0.90		-0.35 (0.16)	0.28	0.50		0.57 (0.19)	0.21	0.91
	3		0.72 (0.05)	0.16	0.92		0.35 (0.10)	0.17	0.99		0.86 (0.06)	0.10	0.94
	4		0.47 (0.13)	0.28	1		0.14 (0.10)	0.23	1		0.57 (0.16)	0.23	0.95

1: No restrictions on ρ
2: $\rho \geq 0$
3: $\rho \geq 0$ and $\rho_{10}, \rho_{01} < \rho_s, \rho_t, \rho_{00}, \rho_{11}$
4: Beta priors

progression is a valid surrogate for overall survival. These data, along with 21 additional trials comprising four separate meta-analyses, were previously analyzed by Burzykowski, *et al.* (2004) where a meta-analytic surrogacy validation method was

used. All six of the trials considered compared the administration of fluorouracil (5-FU) by continuous intravenous infusion (CI) to bolus administration of 5-FU. As these trials all compared the same two treatments and there were no notable differences in patient characteristics among the trials, we pool the data from these nine trials in the application of our method. All together, there were 1,216 patients with 609 randomized to the 5-FU by CI arm and 607 randomized to the bolus 5-FU arm. Patients were followed with tumor response and survival time recorded. Tumor response was defined by one of four categories: complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD). In our analysis, the true endpoint T is survival time, defined as the time from randomization to death from any cause and the surrogate end point S is tumor response, defined as a categorical variable with $S = 1, 2, 3, 4$ for PD, SD, PR, and CR, respectively. The binary treatment indicator for treatment Z is set to 0 for 5-FU by CI and 1 for bolus 5-FU. Tumor response was measured after approximately 3 to six months of follow-up in advance of the recorded survival time.

The observed tumor response frequencies were 52% for PD, 32% for SD, 14% for PR and 3% for CR. The response rate (combined percentage of CR and PR) was higher in the treatment arm, with 20.8% responding compared to 12.8% responding in the control arm. The odds ratio for response in the treatment vs. control arm was 1.84 (95% CI: 1.35, 2.51). The median survival time was longer for those in the treatment group (12.1 months) than for those in the control group (11.3 months), with an estimated hazard ratio of 0.89 (95% CI: 0.79, 1.00) for the treatment vs. control group. Table 3.4 provides the means and standard deviations of $\mathrm{E}[\log(T(1)/T(0))|S(1) - S(0) = s]$ and of the correlation between $q(T(1)) - q(T(0))$ and $\tilde{S}(1) - \tilde{S}(0)$ for each of the four prior scenarios described in the simulation section. We focus on the

estimation done using priors 3 and 4, as these priors performed better in the simulation settings. Under these two priors, $\mathrm{E}[\log(T(1)/T(0))|S(1)-S(1)=0]$ is approximately 0, indicative of a good principal surrogate, with a fairly large correlation between the causal standardized treatment effects. We would therefore conclude that tumor response appears to be a moderately good principal surrogate for overall survival. Appendix A provides plots of the Kaplan-Meier survival curves for the observed data and the mean and 95% credible interval of the posterior predictive distribution from the model for each of the tumor response categories and for each treatment group. The plots suggest that the model appears to provide an appropriate fit to the data.

We compare the results obtained under the Gaussian copula model to those that would have been obtained had the data been analyzed as multivariate normal. To approximately normalize T, we take the third root, and to approximately normalize S, for each $S=s$ we draw a uniform $(s-1,s)$ random variable. Appendix B provides histograms and normal QQ plots of this data. Under this model, we assume that the joint distribution of the counterfactuals is multivariate normal, such that:

$$\begin{pmatrix} S_i(0) \\ S_i(1) \\ T_i(0)^{1/3} \\ T_i(1)^{1/3} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{S_0} \\ \mu_{S_1} \\ \mu_{T_0} \\ \mu_{T_1} \end{pmatrix}, \begin{pmatrix} \sigma^2_{S_0} & \rho_s\sigma_{S_0}\sigma_{S_1} & \rho_{00}\sigma_{S_0}\sigma_{T_0} & \rho_{01}\sigma_{S_0}\sigma_{T_1} \\ & \sigma^2_{S_1} & \rho_{10}\sigma_{S_1}\sigma_{T_0} & \rho_{11}\sigma_{S_1}\sigma_{T_1} \\ & & \sigma^2_{T_0} & \rho_t\sigma_{T_1}\sigma_{T_0} \\ & & & \sigma^2_{T_1} \end{pmatrix} \right)$$

To obtain parameter estimates from this distribution, a Bayesian estimation approach is used as outlined in Chapter II. Additionally, at the beginning of each iteration of the chain, a new value for each observed $S=s$ is drawn from a Uniform $(s-1,s)$ distribution and death times are imputed for censored subjects. We draw a death time from the residual survival distribution, $P(T(Z)^{1/3} > t_Z^{1/3}+b \mid T(Z)^{1/3} > t_Z^{1/3}, T(1-Z)^{1/3}, S(0), S(1), \mu, \Sigma)$, for observed censoring time $t_Z^{1/3}$. Non

informative priors are placed on the observed parameters. Specifically, the prior for each μ is $N(0, 10^6)$, the prior for each σ is $\propto 1$, and the priors for ρ_{00} and ρ_{11} are $Unif(-1, 1)$. We place mildly informative Beta priors on the remaining partially identified parameters, ρ_s, ρ_{10}, ρ_{01}, and ρ_t. Table 3.5 provides the results from analyzing the data as multivariate normal. This method of analysis would identify S as a weaker principal surrogate than under the copula model, as the estimate of $\mathrm{E}[\log(T(1)/T(0))|S(1) - S(0) = 0]$ is slightly larger. The results also show a slightly attenuated increasing trend of $\mathrm{E}[\log(T(1)/T(0))|S(1) - S(0) = s]$ in s and larger standard deviations as compared to the estimates obtained from the copula model. The copula model, therefore provides more efficient estimation of the principal surrogacy quantities when multivariate normality does not hold. Figure 3.2 provides a plot of $\mathrm{E}[\log(T(1)/T(0))|S(1) - S(0) = s]$ vs. s estimated from the Guassian copula model (using the Beta priors) and from the multivariate normal model. The curve estimated by the copula model is less linear than that estimated by the multivariate normal model.

Table 3.4: Application of Gaussian copula to colorectal cancer data

	ρ's unrestricted Mean (SD)	$\rho \geq 0$ Mean (SD)	$\rho \geq 0$ and $\rho_{10}, \rho_{01} < \rho_s, \rho_t, \rho_{00}, \rho_{11}$ Mean (SD)	Beta priors Mean (SD)
$\mathrm{E}[\log(T(1)/T(0))\|S(1) - S(0) = -3]$	-1.09 (0.49)	-0.47 (0.74)	-1.02 (0.53)	-1.08 (0.46)
$\mathrm{E}[\log(T(1)/T(0))\|S(1) - S(0) = -2]$	-0.91 (0.30)	-0.42 (0.39)	-0.79 (0.26)	-0.79 (0.20)
$\mathrm{E}[\log(T(1)/T(0))\|S(1) - S(0) = -1]$	-0.67 (0.16)	-0.21 (0.21)	-0.46 (0.13)	-0.49 (0.12)
$\mathrm{E}[\log(T(1)/T(0))\|S(1) - S(0) = 0]$	-0.28 (0.08)	0.01 (0.10)	-0.005 (0.06)	-0.03 (0.07)
$\mathrm{E}[\log(T(1)/T(0))\|S(1) - S(0) = 1]$	0.84 (0.12)	0.43 (0.14)	0.54 (0.12)	0.66 (0.09)
$\mathrm{E}[\log(T(1)/T(0))\|S(1) - S(0) = 2]$	1.74 (0.22)	0.72 (0.26)	0.96 (0.23)	1.22 (0.19)
$\mathrm{E}[\log(T(1)/T(0))\|S(1) - S(0) = 3]$	2.82 (0.38)	1.08 (0.50)	1.35 (0.50)	1.75 (0.44)
$\mathrm{cor}(q(T(1)) - q(T(0)), \tilde{S}(1) - \tilde{S}(0))$	0.78 (0.10)	0.31 (0.15)	0.63 (0.14)	0.84 (0.10)

Figure 3.2: Plot of $\mathrm{E}[\log(T(1)/T(0))|S(1)-S(0)=s]$ vs. s for colorectal cancer data: estimates from the copula model and from the multivariate normal model

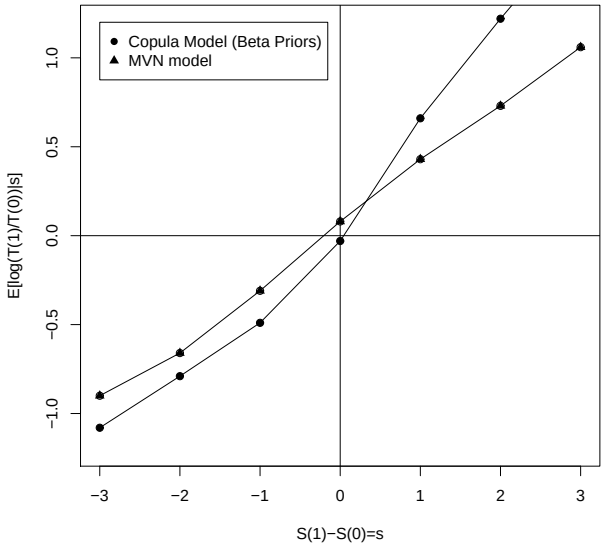


Table 3.5: Surrogacy assessment of colorectal cancer data, analyzed as normal

Parameter	Posterior Mean (SD)
$\mathrm{E}[\log(T(1)/T(0))\|S(1)-S(0)=$-3]	-0.90 (0.77)
$\mathrm{E}[\log(T(1)/T(0))\|S(1)-S(0)=$-2]	-0.66 (0.30)
$\mathrm{E}[\log(T(1)/T(0))\|S(1)-S(0)=$-1]	-0.31 (0.16)
$\mathrm{E}[\log(T(1)/T(0))\|S(1)-S(0)=0]$	0.08 (0.08)
$\mathrm{E}[\log(T(1)/T(0))\|S(1)-S(0)=1]$	0.43 (0.13)
$\mathrm{E}[\log(T(1)/T(0))\|S(1)-S(0)=2]$	0.73 (0.27)
$\mathrm{E}[\log(T(1)/T(0))\|S(1)-S(0)=3]$	1.06 (0.58)

3.7 Discussion

In this chapter, we develop a method for surrogacy assessment under the principal stratification framework for an ordinal surrogate marker and time to event final outcome. We use a Gaussian copula model to jointly model the potential surrogate outcomes and potential final outcomes, and propose quantities from this model that can be used to determine the validity of S as a surrogate marker for T. A Bayesian estimation strategy is used, allowing the use of context specific prior assumptions on the unidentified parameters to be explored in order to aid in estimation. Our

simulation results suggest that the estimation procedure is able to distinguish valid principal surrogates from invalid ones.

In our data example, we compare the results obtained using the proposed Gaussian copula model to those obtained using a multivariate normal model. The results show that there is some efficiency gained by fitting the more appropriate marginal distributions to the data and using the Gaussian copula than by assuming multivariate normality when it may not hold.

In our model formulation, we assume a cumulative probit model for S and a proportional hazards model with a Weibull baseline hazard function for T. The use of different parametric models for these marginal distributions could be explored. Semi-parametric or non-parametric alternatives for the marginal distributions could also be used to model the marginal distributions of S and T.

CHAPTER IV

Using Multi-state Models With a Cured Fraction to Model Colon Cancer Recurrence and Death

4.1 Introduction

In longitudinal medical studies with a time-to-event final outcome, patients may experience multiple disease progression events prior to the event of interest. Examples include CD4 lymphocyte counts in the progression to HIV infection (Foucher, *et al.* 2005) and cancer progression prior to death in survival studies (Putter, *et al.* 2006). The data we examine come from 12 phase III randomized trials in colon cancer where there is interest in building a joint model for the two event times of interest, time to recurrence and time to death and investigating how treatment and other covariates are associated with the event times. A common way to jointly model these disease progression events is through the use of multi-state models (Anderson and Keiding, 2002; Meira-Machado *et al.*, 2009; Putter *et al.*, 2007), which describe the progressions and transitions over time to the various disease states. Common forms of multi-state models include the "progressive three-state model", in which subjects enter some intermediate disease state prior to entering the absorbing state, and the "illness-death model", where healthy subjects may enter a diseased state prior to death,

or die without disease. In these models, transition intensities, which can include covariates, provide the hazards for moving between states. Each disease state that a subject occupies is either a transient state that can be left, such as cancer recurrence, or an absorbing state, such as death, that can never be left once it is entered.

Different model assumptions can be made about the dependence of the transition intensities and time. One approach is to take $t = 0$ as the start of the study and then all subsequent times t refer to the time since the beginning of the study. Klein *et al.* (1994) make this assumption in their analysis of relapse and death in bone marrow transplant patients. A second approach is to set the clock back to 0 upon entry into a new state. This approach assumes that the hazard for entry into each state depends on the entry time into that state. This type of model, termed a semi-Markov model, has been explored by Dabrowska, *et al.* (1994) and Lagakos, *et al.* (1978). Additionally, in the semi-Markov model the hazard for entry into a new state could depend on the time at which the current state was entered (Anderson, *et al.* 2000). In our data analysis, we use a semi-Markov model with recurrence time as a covariate in the hazard model for the transition from recurrence to death. The hazard for moving between states can be modeled either parametrically or semi-parametrically. Putter, *et al.* (2006) explore the use of the semi-parametric Cox model in their analysis of recurrence and survival in breast cancer. Foucher, *et al.* (2005) use a generalized Weibull model for the hazard of transitioning between states. Here, we use a proportional hazards model with a parametric Weibull baseline hazard for each of the transition rates. There is interest in using these semi-Markov multi-state models to jointly model disease progression events as they can be used to assess how individual covariates affect each of the progression rates, and to estimate overall survival, given the disease history.

We propose a semi-Markov model with an incorporated latent cured state to model colon cancer recurrence and survival. This model structure is motivated by the disease process of colon cancer. Cure models have been used to model many different types of cancer where there is known to be a significant proportion of patients whose tumors are completely eliminated by the treatment, and so will never experience a clinical recurrence. These patients are considered to be cured of the disease. We use the mixture model formulation of the cure model, introduced by Berkson and Gage (1952). This model assumes that a proportion of the population, p will never experience the event of interest and are therefore cured. The mixture cure model has been widely discussed in the literature. Yamaguchi (1992) explored the use of a cure model with a logistic mixture probability model and an accelerated failure time model with a generalized gamma distribution. Taylor (1995) used a logistic model for the cure probability and a completely unspecified failure time process. Estimation for a semi-parametric Cox proportional hazards model for the failure time process has been explored by Sy and Taylor (2000) and Peng and Dear (2000).

One issue that arises with the use of the cure model is identifiability due to censoring before the end of the follow-up period (Farewell, 1986). Therefore, it can be difficult to distinguish models with a large population of uncured individuals and long tails of the failure time process from those with small populations of uncured individuals and short tails of the failure time process. In general, in order to justify use of the cure model, there must be sufficient follow-up and a large number of censored observations after the last event. Problems with identifiability are likely to arise if the Kaplan-Meier survival plot of all data does not show a clear level plateau. In the models we propose, the joint modeling of survival time and recurrence time may aid in the identifiability as subjects with survival times greater than the last observed

recurrence time are likely to be cured of disease. Additionally, the appropriateness of the cure fraction in the multi-state model can be assessed through a goodness of fit comparison with a model that does not assume a cured fraction.

The multi-state model and cure model have each been considered with both non-parametric and parametric assumptions. Here, our proposed model combines aspects of both of these models providing insight into the role of covariates on both the curing of the disease and the disease process, as well as the association of the two endpoints of interest, recurrence and death. We apply our model separately to each of the 12 colon cancer trials in order to explore which covariates have similar effects on certain disease aspects across trials performed in different settings. A Bayesian estimation strategy is used to estimate the parameters of the Weibull model, the covariate effects for each of the transition times, and the covariate effects for the logistic model for each trial. As the cured state is only partially observed, we place informative priors on some model parameters to aid in estimation where there is a scientific rationale for these parameters to be close to zero. The adequacy of the model fit is assessed through the use of Cox-Snell residual plots and deviance residual plots are used to determine the proper functional form of covariates. While these would be natural to consider for multi-state models (Kneib and Hennerfeind, 2008), we are not aware of any literature on using them in cure models. The remainder of the paper is organized as follows: Section 4.2 describes the data and Section 4.3 describes the proposed model. In Section 4.4, the estimation procedure is outlined and results of the application of the model to the data from 12 colon cancer clinical trials are provided. Model checking procedures are described in Section 4.5. Section 4.6 provides simulation results and Section 4.7 concludes with a discussion.

4.2 Data Description

The data we consider consist of a total of 13,983 subjects from 12 randomized phase III adjuvant trials of locally advanced colon cancer. Ten of the trials are included in the Sargent, *et al.* (2005) publication, with two additional new trials. These 12 trials were previously analyzed by Conlon, *et al.* (2011), where a separate cure model and Weibull model were used to model time to recurrence and death, respectively. Trial enrollment spanned from 1977 to 1999. One trial (7) included 210 patients with stage 1 cancer; these subjects were excluded from this analysis. Due to differences in the long term follow-up practices between trials, subjects in all trials except trial 1 were censored at 8 years following the time of the last subject accrual. Subjects in trial 1 were censored 4.3 years from the last subject accrual due to a large number of patients administratively censored at this time. The median follow-up time for subjects alive at their last follow up was 8.2 years. In each trial, subjects were followed with a specific protocol, with cancer recurrences and deaths recorded as they occurred. We therefore have two censored event times of interest, recurrence and death. The censoring times for these two events are not necessarily the same, as ascertaining a recurrence time requires active follow-up usually involving a scan, while obtaining a dead or alive status could be obtained through other means. The average proportion of subjects censored for recurrence prior to their last follow up was 9.5% across all trials, with a maximum of 16.9% in Trial 2 and a minimum of 0% in Trial 3. Of the 4346 observed recurrences, 3448 (79.3%) occurred within three years, 4075 (93.8%) occurred within five years, and 4281 (98.5%) occurred within seven years. This type of event time data where very few events happen after a fixed window of time is characteristic of a cured group, and for which a cure model

is appropriate. Kaplan-Meier plots of time to recurrence for the 12 trials, provided in Appendix I, show a clear leveling off providing a strong empirical rationale for use of a cure model. Baseline covariates of interest include age, cancer stage and treatment arm. Each trial compared a different pair of treatments, with one defined as the control arm and the other as the experimental arm. Five of the trials (1, 2, 3, 6 and 7) compared surgery alone to surgery plus some form of chemotherapy. In the other seven trials, both arms contained surgery plus some form of chemotherapy. The primary goal of all 12 trials was to compare overall survival between the pairs of treatments. Based on a simple log-rank test, three of the trials (3, 8, and 9) showed a statistically significant treatment effect. Table 4.1 provides a summary of stage, age and treatment distributions in each trial, as well as the number of recorded recurrences and deaths and longest follow-up time for each trial. As each of these trials compared different pairs of treatments, we fit the model to each trial separately, and then assess which of the covariate effects on the various diseases processes are similar across the 12 trials.

4.3 Multistate model

We model time to recurrence and time to death using a multi-state model for the semi-Markov process. We also incorporate a latent cured fraction into the model for subjects who will never experience a recurrence. The cure model is applied to the recurrence event and assumes that there is a zero probability that some subjects will recur. In this setting, the treatment may eliminate the cancer, resulting in a cured group of patients. Curing of the cancer happens at the time of treatment, but is not immediately observable. If the cancer is not eliminated, the patient is not cured of

Table 4.1: Data summary of 12 trials of colon cancer

Study	N	Recurrences	Recurrence Without Death	Death Without Recur	Total Deaths	Longest Follow-Up (years)	% with Stage 3 Cancer	% in Treatment Group	Age (mean)
1	247	116	14	13	115	9.9	65.6%	49.0%	60.3
2	408	139	11	44	172	9.1	81.6%	62.5%	61.1
3	926	377	31	76	422	11.4	66.1%	49.4%	60.2
4	914	380	36	106	450	9.9	82.5%	75.2%	62.7
5	878	297	33	74	338	12.6	73.8%	49.8%	61.2
6	724	275	10	132	397	13.2	56.8%	48.2%	59.8
7	683	206	32	129	303	12.9	43.4%	50.1%	63.3
8	1040	356	36	67	387	9.7	72.1%	49.8%	56.1
9	2077	605	57	176	724	9.4	58.7%	66.7%	57.0
10	2128	574	66	192	700	10.3	55.9%	49.8%	58.0
11	1549	394	71	115	438	8	53.5%	50.3%	60.5
12	2409	627	189	106	544	6	71.1%	49.8%	57.9
Total	13983	4346	586	1230	4990	13.2	63.8%	54.3%	59.1

disease and will experience a recurrence when the tumor has regrown to a detectable size. The observed data provides information about whether the patient was cured by the treatment. Patients with observed recurrences are known to be in the uncured group. Patients who do not have an observed recurrence may or may not be cured. For patients censored for recurrence, the model assumes that a proportion of these subjects would have never experienced a recurrence even if they had been followed longer, and are therefore cured. Additionally, some subjects who were censored for recurrence could have experienced a recurrence after their censoring time had there been longer follow-up and are therefore in the uncured group with an unobserved recurrence time (Farewell, 1982). Deaths can occur either without a prior recurrence or following a recurrence. The deaths that occur without a prior recurrence are known not to be directly due to the regrowth of the cancer, while deaths following a recurrence may be due to the cancer or other causes. We do not consider cause of death in our models. We use the multi-state model to model four transition intensities

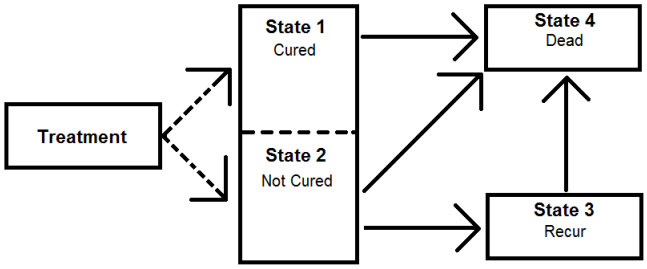
which include the transition from the uncured group to death, the transition from the cured group to death, the transition from the uncured group to recurrence, and the transition from recurrence to death.

4.3.1 Notation and model specifications

Let C_{ir} and T_{ir} be the censoring and event times for recurrence and let C_{id} and T_{id} be the censoring and event times for death for the ith subject, $i = 1, ...n$. Then $Y_{ir} = min(C_{ir}, T_{ir})$ and the event indicator for recurrence, $\delta_{ir} = I(T_{ir} \leq C_{ir})$, and $Y_{id} = min(C_{id}, T_{id})$ and the event indicator for death, $\delta_{id} = I(T_{id} \leq C_{id})$, are observed. Let Z_i, S_i and A_i represent the baseline values of treatment group, cancer stage and age for each subject. Each of these covariates is approximately centered prior to estimation so that $Z_i = -0.5$ or 0.5 for control and treatment, respectively, $S_i = -0.5$ or 0.5 for stage 2 or stage 3 cancer, respectively, and A_i is age, centered at the mean age for a given study in units of 10 years. Let p_i be the probability that the ith subject is cured of disease. We define State 1 to be alive and cured of disease, State 2 to be alive and not cured, State 3 to be alive with recurrence and State 4 to be dead, as illustrated in Figure 4.1. States 1, 2 and 3 are transient states while State 4 is an absorbing state. We model four transition times, $1 \rightarrow 4$, $2 \rightarrow 3$, $2 \rightarrow 4$, and $3 \rightarrow 4$. The true state progressions for many subjects are not fully observed. Specifically, those who were censored for recurrence and alive were either in State 1 or State 2 at the end of their follow-up, and subjects censored for recurrence and dead either made a $1 \rightarrow 4$ or a $2 \rightarrow 4$ transition at their time of death. Subjects with an observed recurrence transitioned from $2 \rightarrow 3$ at their recurrence time and then remained in State 3 if they were censored for death, or made a $3 \rightarrow 4$ transition at their death time.

Figure 4.1: Multi-state cure model structure

Dashed lines represent effect of treatment, solid lines represent transitions between states.



In the standard setting for cure models there is one event time of interest, and the mixture model formulation assumes that a proportion of the population, p will never experience the event of interest, in this case recurrence, and are therefore cured. For the uncured population, the cure model provides information on the estimated time to event from the survival distribution. The marginal survival function for recurrence, $S(t)$, for the entire population is given by $S(t) = p + (1 - p)S_0(t)$, where $S_0(t)$ is the conditional survival function for recurrence for the uncured group. It is common to use a logistic model which includes time independent covariates for the incidence model. Common choices for $S_0(t)$ are the exponential and Weibull distributions. Non-parametric choices for $S_0(t)$ have also been explored. In our more complicated situation with two event times, the structure of the model is more involved, but we retain similar elements, that of a cured fraction described by a logistic model, and distributions of event times given cured status described by proportional hazards models with Weibull baseline hazard functions.

The proposed model can be used to assess how different covariates affect both the probability of being cured and the hazard of transitioning to recurrence or death. Other quantities of interest can also be derived from the model, such as five year survival within each of the treatment arms. Both the models for the time of entry

into each state and p can depend on covariates. The probability of being cured, p_i, is modeled using a logistic link function given by:

$$p_i = \frac{exp(X_i\gamma)}{1+exp(X_i\gamma)}.$$

where X_i is a vector of subject specific covariates that includes the centered covariates of treatment group, stage and age, and γ is a vector of coefficients given by $\gamma = (\gamma_0, \gamma_{Treatment}, \gamma_{Stage}, \gamma_{Age})^T$ The multi-state process is characterized through transition intensities defined as:

$$\lambda_{kj}(t) = \lim_{\Delta t \to 0} p_{kj}(t, t+\Delta t)/\Delta t$$

where $p_{kj}(s,t) = P(M(t) = j | M(s) = k, H_{s-})$ is the probability of being in State j at time t, given that the process was in State k at time s and the history of the process H_{s-}, for states $M(t)$ and $M(s)$ occupied at times t and s, respectively, and $s \leq t$. $\lambda_{kj}(t)$ is then the instantaneous hazard of entering State j, given that the previous state occupied was State k. From this hazard, we can define the survival distributions for each transient state. The survival distributions for remaining in States 1, 2, and 3 are:

$$S_1(t) = exp\left(-\int_0^t \lambda_{14}(u)du\right)$$
$$S_2(t) = exp\left(-\int_0^t \lambda_{23}(u)du - \int_0^t \lambda_{24}(u)du\right)$$
$$S_3(t|t_r) = exp\left(-\int_0^{t-t_r} \lambda_{34}(u)du\right)$$

where t_r is the entry time into state 3.

We use a proportional hazards model with a Weibull baseline hazard function to

model the distribution of waiting times. Specifically, the hazard for transition $k \to j$ for subject i is given by:

$$\lambda_{kj}(t_i; X_i) = \left(\frac{\rho_{kj}}{\alpha_{kj}}\right)\left(\frac{t_i}{\alpha_{kj}}\right)^{\rho_{kj}-1} exp\left(X_i\beta_{kj}\right)$$

For transitions $1 \to 4$ and $2 \to 4$, t_i is a death time. For transition $2 \to 3$, t_i is a recurrence time and for transition $3 \to 4$, t_i is the gap time between entry into the recurred state and death. X_i represents a vector of subject specific covariates which includes the centered covariates of treatment group, stage and age for transitions $1 \to 4$, $2 \to 3$, and $2 \to 4$. For transition $3 \to 4$, X_i includes the centered covaraites of treatment group, stage, age and time to recurrence, centered at the mean time to recurrence for those who recur in a given study in units of years. β_{kj} is a vector of coefficients given by $\beta_{kj} = (\beta_{Treatment_{kj}}, \beta_{Stage_{kj}}, \beta_{Age_{kj}})^T$ for $kj = \{14, 23, 24\}$ and $\beta_{34} = (\beta_{Treatment_{34}}, \beta_{Stage_{34}}, \beta_{Age_{34}}, \beta_{T_r})^T$. ρ_{kj} and α_{kj} are the shape and scale parameters, respectively, for each Weibull model. The covariates can be expected to be associated with each of the model components in differing ways, as each part of the model describes a different aspect of the disease process. The probability of being cured provides information about the tumor and the cell killing effect of the treatment. Transitions $1 \to 4$ and $2 \to 4$ give information about the person as opposed to the cancer and transition $2 \to 3$ provides information about the tumor regrowth. Finally, transition $3 \to 4$ provides information about both the person and the regrowth of the tumor. Six distinct types of people contribute to the likelihood: those who recur and are either dead or alive at a later time, those censored for recurrence at $Y_{ir} = Y_{id}$ and either dead or alive at Y_{id}, and those censored for recurrence at Y_{ir} prior to death or censoring at a later time Y_{id}. These likelihood contributions can be

described by the following three equations:

Recur, dead or alive:

$$(1-p_i)\lambda_{23}(Y_{ir})S_2(Y_{ir})\lambda_{34}(Y_{id}-Y_{ir})^{\delta_{id}}S_3(Y_{id} \mid Y_{ir})$$

Censored for recurrence at $\mathbf{Y_{id}}$, dead or alive:

$$(1-p_i)\lambda_{24}(Y_{id})^{\delta_{id}}S_2(Y_{id}) + p_i\lambda_{14}(Y_{id})^{\delta_{id}}S_1(Y_{id})$$

Censored for recurrence at$\mathbf{Y_{ir}}$ prior to $\mathbf{Y_{id}}$, dead or alive:

$$(1-p_i)\lambda_{24}(Y_{id})^{\delta_{id}}S_2(Y_{id})+$$
$$(1-p_i)\int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id}-u)^{\delta_{id}}S_3(Y_{id} \mid u)du+$$
$$p_i\lambda_{14}(Y_{id})^{\delta_{id}}S_1(Y_{id})$$

In the 12 trials we examine, a small proportion of subjects (0.7% across all trials) had a recurrence at the same date as their time of death. As these subjects likely truly recurred prior to this date but for administrative or other reasons these recurrences were not detected prior to death, their recurrence times were treated as interval censored and their contribution to the likelihood was:

Recur at $\mathbf{Y_{id}}$, dead :

$$(1-p_i)\int_{0}^{Y_{ir}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id}-u)^{\delta_{id}}S_3(Y_{id} \mid u)du$$

4.4 Estimation

We use a Bayesian MCMC technique to estimate the parameters of the multi-state cure model. The parameters for each of the 12 trials are estimated separately, with no mixing of patients across the different trials. There are a total of 25 parameters to estimate for each of the trials which include a shape (ρ) and scale (α) parameter from the Weibull model for each of the hazard rates, covariate effects for each of the hazard models and covariate effects in the logistic model for the probability of cure. As the cured state is only partially observed, we place informative Normal(0, 0.25^2) priors on the treatment and stage coefficients in transition $1 \rightarrow 4$ to aid in estimation. This seems reasonable as treatment group and cancer stage may affect the likelihood of cure, but are unlikely to affect the survival of patients conditional on being cured of disease. Additionally to aid in estimation, we place mildly informative Normal(0, 2^2) priors on the $\log(\alpha)$'s and gamma priors with mean 1 and standard deviation 0.6 on the ρ's. Normal(0, 1) priors are placed on all of the remaining covariate coefficients in the hazard models and in the logistic model. At each iteration of the chain the latent variable representing cured status is simulated. Specifically, subjects without recurrence are placed in either the cured or uncured group by drawing a Bernoulli random variable with probability of cure

$$c_i = \frac{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id})}{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) + (1-p_i)\lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id})}$$

for those censored for recurrence at Y_{id} and

$$c_i = \frac{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Yid)}{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) + (1-p_i)\lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id}) + (1-p_i)\int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u) S_2(u) \lambda_{34}(Y_{id}-u)^{\delta_{id}} S_3(Y_{id}|u) du}$$

for those censored for recurrence at Y_{ir} prior to Y_{id}. As some integrals in the likelihood do not have closed form solutions, numeric integration was used. Specifically,

the following integrals were calculated by adaptive quadrature using the 'integrate' function in R:

$$\int_0^{Y_{ir}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id}-u)^{\delta_{id}}S_3(Y_{id} \mid u)du$$

$$\int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id}-u)^{\delta_{id}}S_3(Y_{id} \mid u)du$$

All covariates are centered prior to estimation, as described in Section 4.3. The posterior distributions for all of the parameters can be obtained from the product of the observed data likelihood and prior distributions. The Metropolis Hastings algorithm is used for parameter estimation. Appendix J provides the full data likelihood and details of the algorithm.

For each parameter, we obtain 5000 draws from its posterior distribution. Table 4.2 provides the posterior mean and standard deviation estimates for all model parameters for each of the 12 trials. Covariate effect estimates that are greater than two times the posterior standard deviation are shown in bold. We estimate the parameters of each of the 12 trials separately, and compare the estimates of each parameter across trials in order to examine which aspects of the disease process have common covariate effects across all of the trials, and which have varying effects. The results show very consistent effects of most covariates on the probability of cure and on each of the transition times across all 12 studies. Stage is seen to have a consistent and strong effect on the probability of being cured and on the time to recurrence and a modest effect on the time to death after recurrence, with higher stage people more likely to recur, recurring earlier and dying sooner after recurrence. The effects of stage on time to death for those who don't recur are smaller in the cured group, with mixed

effects across trials. Age has a strong effect on time to death for both those who are cured and those who are not cured but die before recurrence, with older people dying sooner. There is a mild positive effect of age on time to death after recurrence and mild negative effect of age on time to recurrence. There is also a consistent effect of time to recurrence on time to death, with those recurring quickly dying sooner. The shape parameter, ρ, and scale parameter, α, of the Weibull model are also consistent across trials within each transition. The shape parameter of the Weibull distribution determines the shape of the density curve. When $\rho < 1$, the failure rate decreases with time. Weibull distributions with ρ close to or equal to 1 have a fairly constant failure rate and distributions with $\rho > 1$ have a failure rate that increases with time. For all of the trials, ρ is greater than 1 in the transition from no cure to recur, indicating a short tail in the distributions of recurrence times. This is characteristic of an event with a cured group, as events become unlikely after a certain amount of time. The intercept in the logistic model for the probability of cure indicates that the trials where the control arm was surgery alone tend to have lower cure rates on the control arm than the trials where the control arm included chemotherapy.

As the 12 trials compared different combinations of treatments, the treatment effects vary for one trial to the next. Based on log-rank p-values for overall survival, Trials 3, 8 and 9 had a significant effect of treatment on overall survival, with log-rank p-values equal to 0.004, 0.0003, and 0.04, respectively. Trials 2, 7, and 12 had near significant treatment effects on overall survival with log-rank p-values of 0.09, 0.07, and 0.09, respectively. Additionally, Trials 1, 3, 8, 9, and 12 had a significant treatment effect on time to recurrence, with log-rank p-values of 0.01, < 0.0001, 0.001, 0.04, and 0.03, respectively. The treatment effect estimates from the Markov model show some consistencies with these results, with effects of treatment primarily seen

on the probability of cure and time to recurrence. There is a significant effect of treatment on the probability of cure for Trials 3 and 8. Trial 8 and Trial 1 both have a significant treatment effect on time to recurrence. All trials, except Trial 3 show no effects of treatment on time to death after a recurrence, or on time to death without recurrence for either the cured or uncured group. There is a small adverse effect of treatment on the time to death after a recurrence in Trial 3. Appendix K provides plots of each of the treatment effect estimates across the trials.

4.5 Model Checking and Model Extensions

4.5.1 Checking Goodness of Fit of the Model

Once parameter estimates for the model have been obtained, the model can be used to estimate five year overall survival (OS), traditionally the final endpoint of interest in trials of locally advanced colon cancer. Similarly, three year disease free survival (DFS) which is the minimum of recurrence and death times and has been shown to be an alternative endpoint to five year OS in these types of trials (Sargent, *et al.* 2007), can also be estimated from the model. The point estimates for five year OS and three year DFS for each treatment arm can be compared with the Kaplan-Meier estimate to check the model fit and the standard error estimate can be compared with that of the Kaplan-Meier estimate to assess gains in efficiency through use of the multi-state model. For each subject we can calculate their five year OS probability as:

$$P(T_i > 5 \mid X_i, \theta) = p_i S_1(5) + (1 - p_i) S_2(5) + (1 - p_i) \int_0^5 S_2(u) \lambda_{23}(u) S_3(5 \mid u) du$$

and three year DFS probability as:

$$P(DFS_i > 3 \mid X_i, \theta) = p_i S_1(3) + (1 - p_i) S_2(3)$$

where θ is the vector of parameter values, and the probability is averaged across the age and stage covariate values for each subject and across all of the parameter draws. Table 4.3 provides the Kaplan-Meier and multi-state model five year OS and three year DFS estimates and standard errors. For the 12 trials considered, both the five year OS and three year DFS estimates from the multi-state model are similar to the Kaplan Meier estimates, indicating that the model is an appropriate fit to the data as measured by predictions of OS at five years and DFS at three years. There is also some efficiency gained in estimating these quantities, as seen by the smaller standard errors in the multi-state model estimates as compared to the Kaplan-Meier estimates. Also noticeable from the estimates are the general improvement in five years OS from the chronologically early trials (trials 1,2 and 3) to the more recent trials (10, 11 and 12).

Additional model fit assessments can be made by examining Cox-Snell residual plots. For each subject, we calculate the Cox-Snell residual for time to death. Let $w_i = -\log\hat{S}(Y_{id})$, where:

$$\hat{S}(Y_{id}) = p_i S_1(Y_{id}) + (1 - p_i) S_2(Y_{id}) + (1 - p_i) \int_0^{Y_{id}} S_2(u) \lambda_{23}(u) S_3(Y_{id} \mid u) du.$$

If the model is correct, the pairs (w_i, δ_{id}) should be like a sample from a censored $Exp(1)$ distribution. Therefore, a plot of w_i vs. the Nelson-Aalen estimator for the pairs of (w_i, δ_{id}) should yield a straight line through the origin with slope 1. Figure 4.2 provides the Cox-Snell residual plots for overall survival for the 12 trials. For most of the trials, the proposed model appears to provide an adequate fit to the data.

The above Cox-Snell residuals involve all five model components. Cox-Snell residuals can also be used to check the model fit for some selected aspects of the overall model, specifically for the time to recurrence and for the transition from recur to death. For the time to recurrence, the residual is $w_i = -\log\hat{S}(Y_{ir})$, with the pair ($w_i$, δ_{ir}) compared to the $Exp(1)$ distribution, where:

$$\hat{S}(Y_{ir}) = p_i + (1-p_i)exp\left(-\left(\frac{Y_{ir}}{\alpha_{23}}\right)^{\rho_{23}} exp(\beta_{trt_{23}}Z_i + \beta_{st_{23}}S_i + \beta_{age_{23}}A_i)\right)$$

and for the recur to death transition the residual is $w_i = -\log\hat{S}(Y_{id} - Y_{ir})$, with the pair ($w_i$, δ_{id}) compared to the $Exp(1)$ distribution, where:

$$\hat{S}(Y_{id} - Y_{ir}) = exp\left(-\left(\frac{Y_{id}-Y_{ir}}{\alpha_{34}}\right)^{\rho_{34}} exp(\beta_{trt_{34}}Z_i + \beta_{st_{34}}S_i + \beta_{age_{34}}A_i + \beta_{Tr_{34}}Y_{ir})\right)$$

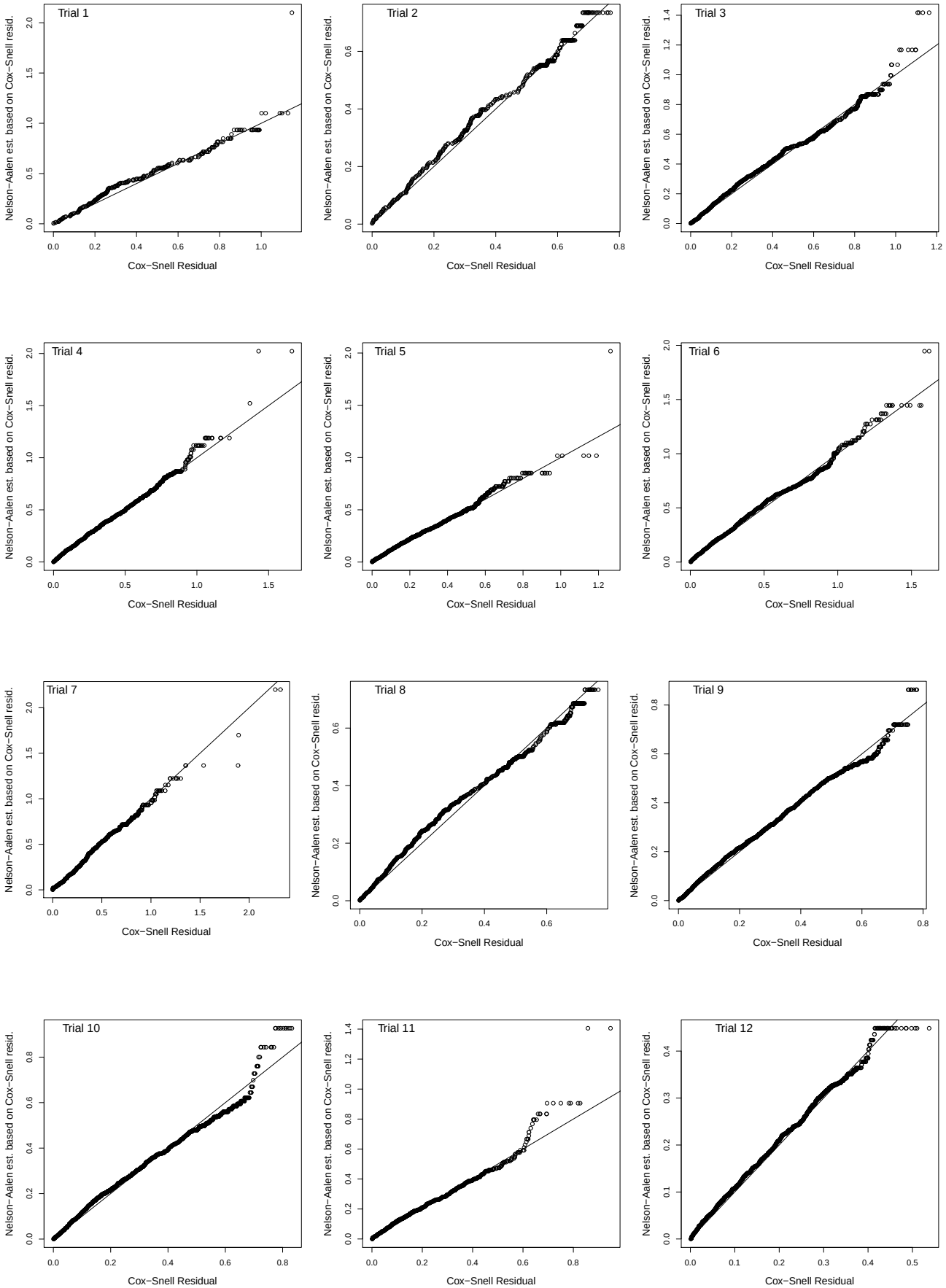
where w_i is calculated only for those who recur. Appendix L provides the Cox-Snell residual plots for the time to recurrence transition and the recurrence to death transition. The residual plots for the individual transitions show that the model fits for the time to recurrence fairly well, but there may be some lack of fit in the transition from recurrence to death in some trials.

Deviance residuals are a standard way of assessing the functional form of covariates in survival analysis models. Here, we adapt them to the multi-state cure model. For each person, the Martingale residual for overall survival is defined as $r_i = \delta_{id} - \hat{\Lambda}(Y_{id})$, where:

$$\hat{\Lambda}(Y_{id}) =$$
$$-log\left(p_iS_1(Y_{id}) + (1-p_i)S_2(Y_{id}) + (1-p_i)\int_0^{Y_{id}} S_2(u)\lambda_{23}(u)exp\left(-\int_0^{Y_{id}-u}\lambda_{34}(v)dv\right)du\right)$$

The Martingale residuals can be viewed as the difference between the observed number of deaths for subject i between time 0 and Y_{id} and the expected number based

Figure 4.2: Cox-Snell residual plots for time to death. Results from 12 trials.



on the model. As the Martingale residuals have an asymmetric range, the deviance residuals are often preferred. The deviance residuals are defined as $D_i = \text{sign}(r_i)\sqrt{-2[r_i + \delta_{id} log(\delta_{id} - r_i)]}$. Plots of the deviance residuals against covariates should be symmetric about a horizontal line at 0.

In addition to checking the Deviance residuals for overall survival, we can also use them to assess the functional forms of covariates for time to recurrence and for time to death after recurrence. In the first case, we have: $r_i = \delta_{ir} - \hat{\Lambda}(Y_{ir})$ where:

$$\hat{\Lambda}(Y_{ir}) = -\log\left(p_i + (1-p_i)exp\left(-\left(\frac{Y_{ir}}{\alpha_{23}}\right)^{\rho_{23}} exp(\beta_{trt_{23}}Z_i + \beta_{st_{23}}S_i + \beta_{age_{23}}A_i)\right)\right)$$
$$D_i = \text{sign}(r_i)\sqrt{-2[r_i + \delta_{ir} log(\delta_{ir} - r_i)]}$$

and in the second case we have: $r_i = \delta_{id} - \hat{\Lambda}(Y_{id})$ where:

$$\hat{\Lambda}(Y_{ir}) = \left(\frac{(Y_{id}-Y_{ir})}{\alpha_{34}}\right)^{\rho_{34}} exp(\beta_{trt_{34}}Z_i + \beta_{st_{34}}S_i + \beta_{age_{34}}A_i + \beta_{Tr_{34}}Y_{ir})$$
$$D_i = \text{sign}(r_i)\sqrt{-2[r_i + \delta_{id} log(\delta_{id} - r_i)]}$$

where r_i is only calculated for those who recur. Results of these deviance residuals are shown in Appendix M. Deviance residuals plots are in general hard to interpret. The plots suggest that a linear function of age is not inadequate for both the recurrence transition and the death after recurrence transition and for the overall model fit. The covariate for recurrence time in the death after recurrence transition also appears to be adequately modeled by a linear function. The graphs suggest that there is no obvious, consistent departure across the 12 trials.

4.5.2 Model Adaptations

While we have shown a good fit of the multi-state cure model with meaningful interpretation of the parameters, a natural question is could we have obtained an

adequate fit with a simpler model. To compare the full multi-state model to simpler models we use the DIC statistic (Spiegelhalter *et al.*, 2002). The DIC value for each study is calculated as $DIC = 2E[D(\theta)] - D(\bar{\theta})$, where $D(\bar{\theta})$ is -2 times the log-likelihood calculated at the mean value of each parameter, and $E[D(\theta)]$ is the mean of -2 times the log-likelihood across all parameter draws. A multi-state model is fit to these data without including a cured group. This results in a simpler model with 16 as opposed to 25 parameters, and may be a better choice if there is uncertainty about the existence of the cured population and may give an adequate fit. In this case the multi-state model would be fit without modeling the cured fraction and there would be only one path from the alive state to the death state for subjects who died without recurrence. Therefore, we only consider transitions $2 \rightarrow 3$, $2 \rightarrow 4$ and $3 \rightarrow 4$. Another simpler model that may provide an adequate fit to the data is one in which all of the parameters in the $1 \rightarrow 4$ transition and in the $2 \rightarrow 4$ transition are forced to be the same. Subjects making a $1 \rightarrow 4$ transition are those who die after being cured of disease and those who make a $2 \rightarrow 4$ transition are those who are not cured of disease but die prior to recurrence. In both of these cases, the subjects are dying from causes other than the cancer, so it is plausible that the parameters of these two hazard rates may be similar. This simpler model has 5 fewer parameters to estimate than the full model. The fit of the models without a cured fraction and with the parameters in transitions $1 \rightarrow 4$ and $2 \rightarrow 4$ constrained to be the same can be compared to the full model through a comparison of DIC values. Table 4.4 provides the DIC values for each of these models for the 12 trials, with the DIC of the best fitting model in bold. In all 12 trials, the model with a cured group and different parameter values for the $1 \rightarrow 4$ and $2 \rightarrow 4$ transitions is preferred over the simpler models, with the model having no cured fraction providing the worst fit. The Cox-Snell residual plots

for time to death can also be compared between the two models to visually assess the adequacy of the model fits. Appendix N provides the Cox-Snell residual plots for the model with no cured fraction. These plots show a larger deviation from the line through the origin than the plots made using the complete model, indicating a poorer fit to the data.

4.5.3 Recurrence only model

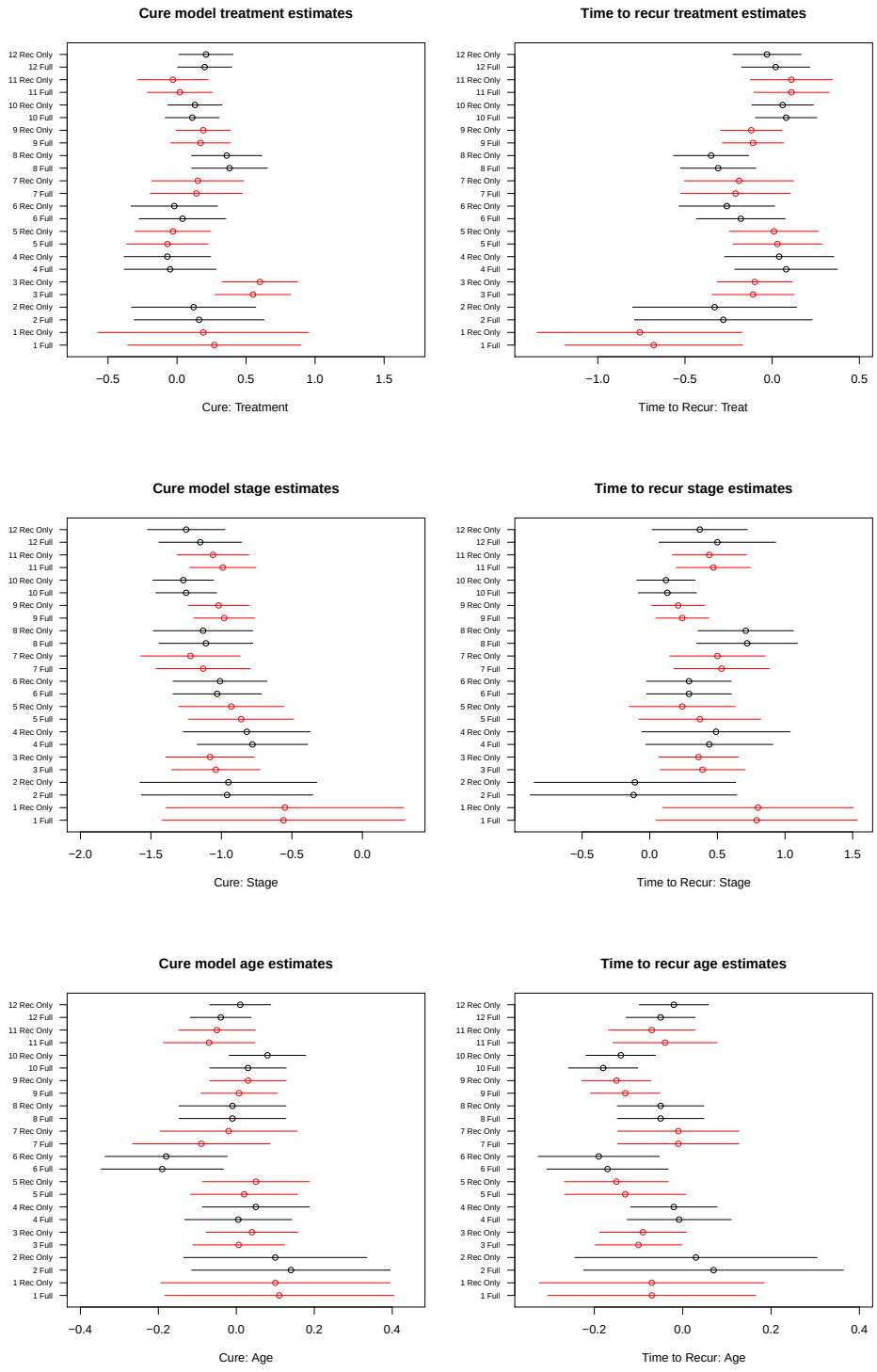
The multi-state Markov model provides a convenient way to deal with the competing risk nature of the recurrence and death events. In colon cancer, recurrence substantially changes the risk of death, and some non-cured patients die from other causes prior to experiencing a recurrence. Therefore, for these patients, their recurrence time is unobservable and some caution must be taken in interpreting the probability of recurrence in the presence of death as a competing risk. If we are only interested in the recurrence event, we can empirically examine the effect of ignoring death by comparing the estimates obtained from the proposed multi-state model to those from a standard cure model for recurrence with a marginal proportional hazard model with Weibull baseline function for the hazard of recurrence. In this model, subjects who die without recurrence are censored for recurrence at their death time. A comparison of the estimates from the multi-state cure model and the simpler cure model provides insight into whether or not it is necessary to build the entire joint multi-state model for recurrence and death if we are only interested in the recurrence event. Figure 4.3 provides plots of the parameter estimates for treatment, stage and age in the logistic model and hazard model for recurrence from both the multi-state model and from a standard cure model for recurrence. The plots show similar estimates of covariate effects with similar standard errors of effect estimates for both

parts of the model, indicating that if we are only interested in the recurrence end point, a simpler model that ignores the time to death endpoint provides adequate estimation of parameters for these data sets. It can be shown algebraically that if we assume that $\lambda_{14}(t) = \lambda_{24}(t)$ then the maximum likelihood estimates from the full multi-state cure model are equivalent to the maximum likelihood estimates from the simple cure model, if the censoring times for recurrence and death do not differ. The deaths in the $1 \to 4$ and $2 \to 4$ transitions are both primarily due to causes other than cancer, so it might be reasonable to expect $\lambda_{14}(t)$ and $\lambda_{24}(t)$ to be similar although in the previous section we showed better DIC values for the full model compared to the model with $\lambda_{14}(t) = \lambda_{24}(t)$. While the points estimates from the data for transitions $1 \to 4$ and $2 \to 4$ are dissimilar, there is wide uncertainty in the estimation of these quantities. Thus it is not too surprising that there is considerable similarity in the estimates for the logistic model and for the $\lambda_{23}(t)$ parameters from the full multi-state cure model and simple cure model.

4.6 Simulations

A small simulation study was conducted to assess the performance of the estimation procedure and the impact of the prior distributions and sample size. The first simulation uses the same prior distributions as were used in the estimation of the colon cancer data, with a sample size of 1000. In the second simulation, the variance of all prior distributions is increased. The third simulation has the same prior distributions as the first simulation with the sample size decreased to 500 to assess how the prior distributions used in our data analysis performed among the varying trial sizes. In each of these three cases, half of the subjects were assigned to be in the treatment

Figure 4.3: Comparison of estimates from full multi-state cure model to recurrence only model.



Results shown for 12 trials. The lower line for each trial is the posterior mean and 95% credible interval for the coefficient from the full multi-state cure model. The upper line is the posterior mean and 95% credible interval for the coefficient from the cure model that does not incorporate death times.

arm and half were assigned to be in the control arm. Two-thirds of subjects were assigned to have stage three disease, and the remaining one-third were assigned to have stage two disease. Subjects were accrued over a four year period with six years of additional follow-up.

The probability of being cured of disease is first generated using $p_i = \frac{exp(\gamma_0+\gamma_{trt}Z_i+\gamma_{stage}S_i)}{1+exp(\gamma_0+\gamma_{trt}Z_i+\gamma_{stage}S_i)}$, where Z_i denotes treatment group and S_i denotes stage. Each of these covariates are centered at 0 so that Z_i is equal to -0.5 for the control group and 0.5 for the treatment group and S_i is equal to -0.75 for stage two disease and 0.25 for stage three diseases. For those who are cured of disease, we then generate a death time using the hazard model for transition $1 \rightarrow 4$ with treatment as the only covariate. For those who are not cured of disease, we generate a recurrence time using the hazard model for transition $2 \rightarrow 3$ and a death time using the hazard model for transition $2 \rightarrow 4$ with treatment as the only covariate in each of these hazard rates. If the death time for uncured subjects is less than the recurrence time, then a $2 \rightarrow 4$ transitions is made at the death time. If the recurrence time is less than the death time, then a $2 \rightarrow 3$ transition is made at that time. For those who recur, the time between their recurrence and death is generated using the hazard model for transition $3 \rightarrow 4$ with treatment and recurrence time as covariates. Subjects are censored six years after the last accrual.

Table 4.5 provides posterior means, standard deviations, average posterior standard deviations ($P\bar{S}D$) and coverage rates from 200 simulations. As the parameters in the hazards for transition for subjects with a recurrence are fully identified, the parameters in transitions $2 \rightarrow 3$, and $3 \rightarrow 4$ are estimated with little bias in all three simulations. Parameters in the logistic model for the probability of being cured are also consistently estimated. When the sample size remains the same, but the vari-

ance on the priors is increased, the parameters in transition $1 \rightarrow 4$ have slightly lower coverage rates and slightly higher standard deviations and $P\bar{S}D$'s. The parameters in transition $2 \rightarrow 4$ also have slightly lower coverage rates and noticeably higher standard deviations and $P\bar{S}D$'s. When the priors are kept the same but the sample size is decreased, the parameter estimates remain similar, but the standard deviations and $P\bar{S}D$'s are increased across all parameters. Simulation 1, which uses the priors that were used in obtaining estimates for our 12 trials and a large sample size, appears to provide the best estimation across all parameters. We note however, that even in this simulation there is some bias and large uncertainty in estimating the parameters for transition $2 \rightarrow 4$ due to the small number of subjects who are uncured of disease but die prior to a recurrence.

4.7 Discussion

In this paper, we have used a multi-state Markov model to formulate a joint model for recurrence and death in colon cancer with an incorporated cured fraction. The proposed model is complex with a large number of parameters to estimate, however it is well motivated by the context of the disease process of colon cancer, where it is likely that a proportion of the population will be cured of disease, and recurrence is known to influence survival time. The parameter estimates obtained from the model provide meaningful interpretations as to how different covariates affect the various disease elements. We presented methods for assessing the adequacy of the model fit and the functional form of covariates, both for the overall model and individual model components, which can aid in choosing an appropriate model. A Bayesian estimation strategy was used to estimate parameters, with informative priors placed on some

parameters to aid in estimation. As our simulation results show, some parts of the model are more sensitive to the choice of the prior due to a lack of identifiability.

Adaptations to this model are possible. For example, we have used fully parametric models in this article. Semi-parametric alternatives for the Weibull model could be explored (Saten and Sternberg, 1999), however the Weibull model appeared to provide an adequate fit to the data in this setting, except, possibly for the transition from recurrence to death. To explore this, we fit the proposed model with a generalized Weibull baseline hazard, as described by Foucher, *et al.* (2007) for the $3 \rightarrow 4$ transition. Based on a comparison of DIC values, this model provided a slightly better overall fit to the data for all 12 trials, and the Cox-Snell residual plots for the $3 \rightarrow 4$ transition appeared to provide a better fit. However, the covariate coefficient estimates and posterior standard deviations for this model were nearly identical to those obtained by the model presented in this paper, indicating that the proposed model is somewhat robust to slight model misspecification.

Maximum likelihood estimation could be used to obtain parameter estimates as opposed to our Bayesian approach; however this is computationally more difficult, and the Bayesian estimation approach facilitates placing informative priors on selected parameters. In our model formulation, recurrence times are treated as known, when more realistically they occur sometime within an interval of scheduled clinic visits, but, the left hand end of this interval is unknown to us. However, if this information were available, the models could be formulated to reflect this by treating recurrence time as interval censored. An additional useful extension to the proposed model for current clinical trial practice would be the ability to return to the disease-free state after recurrence. This now appears to occur for about 10% of patients who recur but then live for a long time after the recurrence, presumably due to subsequent

therapy. The model could also be adapted to consider cause specific mortality if this information were available in the data.

As the models demonstrated, there were common effects of age, stage and the Weibull shape parameter across studies within the logistic model and within the transition rates. Therefore, information from these covariates could be borrowed across trials with estimates shrunk towards common values by fitting a Hierarchical model to the 12 trials examined. Estimates from the Hierarchical model could then be applied to data from a new trial during follow-up to aid in the estimation and analysis of treatment effects on overall survival. This will be described in future research.

The proposed model also has the potential to be used to use recurrence as an auxiliary variable for overall survival. As recurrence time is often an informative marker in predicting a patient's overall survival time, recurrence information along with the parameter estimates from the joint model could be used to impute death times for censored subjects and potentially improve the efficiency in the analysis of overall survival. This strategy could also result in the shortening of the length of the trial, if the information lost due to early censoring could be correctly recovered by the model. The effectiveness of this strategy will be examined in future work.

Table 4.2: Multi-state model parameter estimates and posterior standard deviations for all trials

Regression parameter estimates greater than two times the posterior standard deviation are shown in bold.

	Trial											
	1	2	3	4	5	6	7	8	9	10	11	12
Cure												
Logistic Model												
γ_0	-0.07 (0.23)	0.79 (0.18)	0.48 (0.09)	0.40 (0.11)	0.73 (0.09)	0.38 (0.09)	0.56 (0.09)	0.83 (0.09)	0.85 (0.06)	0.99 (0.05)	1.01 (0.07)	1.17 (0.08)
$\gamma_{Treatment}$	0.27 (0.32)	0.16 (0.24)	**0.55 (0.14)**	-0.05 (0.17)	-0.07 (0.15)	0.04 (0.16)	0.14 (0.17)	**0.38 (0.14)**	0.17 (0.11)	0.11 (0.10)	0.02 (0.12)	**0.20 (0.10)**
γ_{Stage}	-0.56 (0.44)	**-0.96 (0.31)**	**-1.04 (0.16)**	**-0.78 (0.20)**	**-0.86 (0.19)**	**-1.03 (0.16)**	**-1.13 (0.17)**	**-1.11 (0.17)**	**-0.98 (0.11)**	**-1.25 (0.11)**	**-0.99 (0.12)**	**-1.15 (0.15)**
γ_{Age}(10 yrs)	0.11 (0.15)	0.14 (0.13)	0.006 (0.06)	0.005 (0.07)	0.02 (0.07)	**-0.19 (0.08)**	-0.09 (0.09)	-0.01 (0.07)	0.007 (0.05)	0.03 (0.05)	-0.07 (0.06)	-0.04 (0.04)
Time to recur												
$2 \rightarrow 3$ **Transition**												
$\log(\alpha_{23})$	1.11 (0.19)	0.82 (0.22)	0.79 (0.07)	1.09 (0.10)	0.99 (0.08)	0.91 (0.06)	0.97 (0.07)	1.04 (0.07)	0.83 (0.04)	0.93 (0.04)	1.02 (0.05)	0.94 (0.08)
ρ_{23}	1.15 (0.10)	1.13 (0.10)	1.12 (0.05)	1.22 (0.05)	1.39 (0.07)	1.25 (0.06)	1.30 (0.07)	1.50 (0.06)	1.50 (0.05)	1.49 (0.05)	1.49 (0.06)	1.46 (0.05)
$\beta_{Treatment_{23}}$	**-0.68 (0.26)**	-0.28 (0.26)	-0.11 (0.12)	0.08 (0.15)	0.03 (0.13)	-0.18 (0.13)	-0.21 (0.16)	**-0.31 (0.11)**	-0.11 (0.09)	0.08 (0.09)	0.11 (0.11)	0.02 (0.10)
$\beta_{Stage_{23}}$	**0.79 (0.38)**	-0.12 (0.39)	**0.39 (0.16)**	0.44 (0.24)	0.37 (0.23)	0.29 (0.16)	**0.53 (0.18)**	**0.72 (0.19)**	**0.24 (0.10)**	0.13 (0.11)	**0.47 (0.14)**	**0.50 (0.22)**
$\beta_{Age_{23}}$ (10 yrs)	-0.07 (0.12)	0.07 (0.15)	**-0.10 (0.05)**	-0.008 (0.06)	-0.13 (0.07)	**-0.17 (0.07)**	-0.01 (0.07)	-0.05 (0.05)	**-0.13 (0.04)**	**-0.18 (0.04)**	-0.04 (0.06)	-0.05 (0.04)
Cure to death												
$1 \rightarrow 4$ **Transition**												
$\log(\alpha_{14})$	4.26 (0.77)	3.54 (0.34)	3.28 (0.21)	3.00 (0.16)	3.47 (0.21)	3.08 (0.13)	3.16 (0.12)	4.18 (0.52)	3.56 (0.22)	3.42 (0.13)	3.85 (0.37)	3.43 (0.33)
ρ_{14}	1.59 (0.63)	1.58 (0.41)	2.36 (0.47)	2.72 (0.48)	2.46 (0.38)	1.95 (0.33)	1.86 (0.23)	1.45 (0.43)	1.89 (0.29)	2.25 (0.22)	1.80 (0.37)	2.35 (0.52)
$\beta_{Treatment_{14}}$	0.05 (0.24)	-0.18 (0.22)	-0.05 (0.20)	0.05 (0.19)	-0.01 (0.19)	0.01 (0.17)	-0.22 (0.17)	-0.20 (0.20)	-0.21 (0.16)	0.10 (0.14)	0.08 (0.18)	-0.19 (0.20)
$\beta_{Stage_{14}}$	0.04 (0.24)	0.05 (0.22)	0.08 (0.19)	0.04 (0.19)	0.06 (0.19)	-0.02 (0.16)	0.16 (0.16)	-0.01 (0.21)	0.20 (0.16)	-0.13 (0.15)	0.29 (0.18)	-0.10 (0.21)
$\beta_{Age_{14}}$ (10 yrs)	**1.23 (0.47)**	0.53 (0.31)	**0.75 (0.16)**	**1.05 (0.19)**	**1.45 (0.23)**	**0.65 (0.12)**	**0.79 (0.14)**	**0.49 (0.20)**	**0.75 (0.14)**	**0.90 (0.13)**	**1.00 (0.17)**	**0.78 (0.17)**
Noncure to death												
$2 \rightarrow 4$ **Transition**												
$\log(\alpha_{24})$	4.05 (0.90)	3.58 (0.87)	3.96 (0.71)	4.62 (0.78)	5.61 (0.96)	3.70 (0.66)	5.64 (0.91)	4.42 (0.90)	4.42 (0.77)	4.49 (0.59)	4.70 (0.80)	5.16 (0.80)
ρ_{24}	1.73 (0.64)	1.43 (0.54)	1.06 (0.20)	0.74 (0.12)	0.72 (0.13)	1.07 (0.20)	0.52 (0.08)	1.45 (0.64)	0.90 (0.14)	0.88 (0.12)	0.87 (0.27)	0.67 (0.10)
$\beta_{Treatment_{24}}$	0.10 (0.88)	-0.69 (0.83)	0.71 (0.54)	-0.26 (0.44)	0.04 (0.56)	-0.57 (0.51)	0.10 (0.40)	0.22 (0.82)	0.31 (0.48)	0.05 (0.35)	-0.43 (0.57)	-0.02 (0.34)
$\beta_{Stage_{24}}$	0.11 (0.90)	-0.19 (0.85)	-0.33 (0.53)	-0.48 (0.47)	-0.98 (0.56)	0.33 (0.53)	-0.56 (0.38)	0.07 (0.86)	-0.34 (0.44)	-0.48 (0.38)	-0.65 (0.55)	-0.68 (0.40)
$\beta_{Age_{24}}$ (10 yrs)	0.64 (0.78)	-0.13 (0.61)	**0.65 (0.30)**	**0.82 (0.24)**	**0.92 (0.38)**	**0.66 (0.31)**	**0.87 (0.30)**	0.50 (0.61)	0.62 (0.36)	**0.99 (0.24)**	0.64 (0.37)	**0.74 (0.18)**
Recur to death												
$3 \rightarrow 4$ **Transition**												
$\log(\alpha_{34})$	0.85 (0.17)	0.40 (0.27)	1.19 (0.19)	1.45 (0.16)	1.51 (0.23)	0.72 (0.19)	1.26 (0.26)	1.13 (0.20)	1.17 (0.18)	1.26 (0.16)	1.64 (0.21)	1.60 (0.18)
ρ_{34}	1.15 (0.09)	0.99 (0.07)	0.92 (0.04)	1.01 (0.04)	0.81 (0.04)	0.89 (0.04)	0.84 (0.05)	0.91 (0.04)	0.90 (0.03)	0.88 (0.03)	0.87 (0.04)	0.96 (0.04)
$\beta_{Treatment_{34}}$	0.16 (0.21)	0.14 (0.18)	**0.27 (0.11)**	0.16 (0.13)	0.18 (0.13)	0.25 (0.13)	-0.005 (0.15)	-0.06 (0.11)	0.03 (0.09)	-0.07 (0.09)	-0.15 (0.11)	**0.18 (0.09)**
$\beta_{Stage_{34}}$	**1.05 (0.26)**	-0.01 (0.29)	**0.39 (0.14)**	**0.59 (0.18)**	**0.50 (0.18)**	0.13 (0.14)	**0.34 (0.16)**	**0.32 (0.16)**	**0.33 (0.10)**	**0.41 (0.11)**	**0.38 (0.13)**	0.16 (0.15)
$\beta_{Age_{34}}$ (10 yrs)	0.06 (0.09)	0.16 (0.10)	0.07 (0.04)	**0.12 (0.05)**	**0.12 (0.06)**	0.03 (0.07)	**0.30 (0.08)**	0.01 (0.05)	**0.12 (0.04)**	**0.14 (0.04)**	**0.13 (0.05)**	**0.20 (0.04)**
β_{T_r} (yrs)	-0.07 (0.07)	0.04 (0.06)	**-0.11 (0.04)**	**-0.23 (0.04)**	**-0.17 (0.05)**	**-0.10 (0.04)**	-0.07 (0.05)	**-0.11 (0.05)**	**-0.12 (0.04)**	**-0.10 (0.03)**	**-0.17 (0.04)**	**-0.22 (0.05)**

Table 4.3: Five year OS and three year DFS estimates: Kaplan-Meier, and Multi-state model

	$P(T>5\|Z_i=1,X_i,\theta)$		$P(DFS>3\|Z_i=1,X_i\theta)$		$P(T>5\|Z_i=0,X_i,\theta)$		$P(DFS>3\|Z_i=0,X_i,\theta)$	
	Kaplan-Meier Est. (SE)	Multi-state Model Est. (SE)	Kaplan-Meier Est. (SE)	Multi-state Model Est. (SE)	Kaplan-Meier Est. (SE)	Multli-state Model Est. (SE)	Kaplan-Meier Est. (SE)	Multi-state Model Est. (SE)
Trial 1	0.651 (0.044)	0.657 (0.037)	0.669 (0.043)	0.690 (0.036)	0.561 (0.044)	0.552 (0.036)	0.548 (0.044)	0.540 (0.038)
Trial 2	0.695 (0.029)	0.705 (0.025)	0.721 (0.028)	0.727 (0.025)	0.638 (0.039)	0.653 (0.034)	0.665 (0.038)	0.659 (0.035)
Trial 3	0.693 (0.022)	0.694 (0.019)	0.700 (0.021)	0.702 (0.020)	0.620 (0.022)	0.622 (0.020)	0.590 (0.023)	0.589 (0.021)
Trial 4	0.635 (0.018)	0.632 (0.017)	0.641 (0.018)	0.650 (0.017)	0.658 (0.032)	0.656 (0.029)	0.637 (0.032)	0.665 (0.031)
Trial 5	0.693 (0.022)	0.698 (0.019)	0.697 (0.022)	0.705 (0.020)	0.716 (0.022)	0.730 (0.019)	0.712 (0.022)	0.722 (0.020)
Trial 6	0.639 (0.026)	0.626 (0.022)	0.682 (0.025)	0.675 (0.022)	0.601 (0.025)	0.618 (0.021)	0.634 (0.025)	0.642 (0.021)
Trial 7	0.748 (0.024)	0.735 (0.020)	0.731 (0.024)	0.731 (0.021)	0.668 (0.026)	0.697 (0.021)	0.693 (0.025)	0.691 (0.023)
Trial 8	0.762 (0.019)	0.756 (0.016)	0.749 (0.019)	0.759 (0.017)	0.656 (0.021)	0.661 (0.019)	0.665 (0.021)	0.657 (0.019)
Trial 9	0.738 (0.012)	0.746 (0.011)	0.743 (0.012)	0.747 (0.012)	0.704 (0.017)	0.712 (0.016)	0.711 (0.017)	0.707 (0.016)
Trial 10	0.765 (0.013)	0.775 (0.011)	0.767 (0.013)	0.770 (0.012)	0.761 (0.013)	0.762 (0.012)	0.760 (0.013)	0.760 (0.012)
Trial 11	0.788 (0.015)	0.794 (0.013)	0.778 (0.015)	0.779 (0.013)	0.788 (0.015)	0.783 (0.013)	0.783 (0.015)	0.783 (0.014)
Trial 12	0.802 (0.012)	0.797 (0.011)	0.780 (0.012)	0.775 (0.011)	0.783 (0.012)	0.781 (0.011)	0.748 (0.013)	0.743 (0.012)

Table 4.4: Multi-state model comparison by DIC values

	Complete Model	No Cured Fraction	Parameters in $1 \to 4$, $2 \to 4$ same
# of Parameters	25	16	20
		DIC	
Trial 1	**1093.7**	1130.8	1099.4
Trial 2	**1789.4**	2106.3	1793.5
Trial 3	**4206.5**	4375.7	4212.0
Trial 4	**4392.7**	4507.4	4419.6
Trial 5	**3572.5**	3694.5	3591.7
Trial 6	**3607.8**	3740.2	3615.0
Trial 7	**3176.4**	3301.0	3209.4
Trial 8	**4084.6**	4303.0	4096.4
Trial 9	**7677.2**	8083.2	7684.4
Trial 10	**7706.4**	8061.3	7721.2
Trial 11	**5225.7**	5407.6	5238.6
Trial 12	**7172.9**	7349.0	7178.5

Table 4.5: Simulation results from the multi-state cure model

Simulations 1 and 3 use informative prior distributions, simulation 2 uses weakly informative priors.

	Simulation 1 (n=1000)					Simulation 2 (n=1000)					Simulation 3 (n=500)				
Parameter	Prior	True Value	Estimate (SD)	PSD	Coverage	Prior	True Value	Estimate (SD)	PSD	Coverage	Prior	True Value	Estimate (SD)	PSD	Coverage
$\log(\alpha_{14})$	N(0,4)	1.8	1.79 (0.04)	0.04	0.94	N(0,25)	1.8	1.79 (0.04)	0.04	0.92	N(0,4)	1.8	1.79 (0.05)	0.05	0.98
ρ_{14}	G(1,0.4)	2	1.94 (0.16)	0.16	0.94	G(1,1.6)	2	1.94 (0.18)	0.17	0.93	G(1,0.4)	2	1.94 (0.19)	0.20	0.95
$\beta_{trt_{14}}$	N(0,0.06)	0	0.004 (0.10)	0.11	0.97	N(0,4)	0	0.006 (0.12)	0.12	0.94	N(0,0.06)	0	-0.0001 (0.10)	0.12	0.995
$\log(\alpha_{23})$	N(0,4)	0.8	0.80 (0.07)	0.07	0.96	N(0,25)	0.8	0.79 (0.07)	0.07	0.95	N(0,4)	0.8	0.80 (0.09)	0.10	0.94
ρ_{23}	G(1,0.4)	1.5	1.49 (0.08)	0.08	0.93	G(1,1.6)	1.5	1.49 (0.08)	0.08	0.93	G(1,0.4)	1.5	1.47 (0.10)	0.10	0.94
$\beta_{trt_{23}}$	N(0,1)	-0.5	-0.49 (0.16)	0.17	0.95	N(0,4)	-0.5	-0.49 (0.17)	0.17	0.93	N(0,1)	-0.5	-0.50 (0.19)	0.21	0.97
$\log(\alpha_{24})$	N(0,4)	1.8	2.06 (0.35)	0.41	0.95	N(0,25)	1.8	2.25 (0.60)	0.61	0.91	N(0,4)	1.8	2.13 (0.37)	0.55	0.97
ρ_{24}	G(1,0.4)	1	0.98 (0.13)	0.13	0.95	G(1,1.6)	1	0.96 (0.16)	0.16	0.92	G(1,0.4)	1	1.01 (0.15)	0.19	0.98
$\beta_{trt_{24}}$	N(0,1)	0	-0.02 (0.25)	0.30	0.97	N(0,4)	0	-0.08 (0.36)	0.40	0.96	N(0,1)	0	-0.07 (0.34)	0.43	0.98
$\log(\alpha_{34})$	N(0,4)	0.9	0.89 (0.08)	0.08	0.95	N(0,25)	0.9	0.89 (0.08)	0.08	0.95	N(0,4)	0.9	0.92 (0.09)	0.10	0.95
ρ_{34}	G(1,0.4)	0.9	0.91 (0.05)	0.05	0.95	G(1,1.6)	0.9	0.91 (0.05)	0.05	0.94	G(1,0.4)	0.9	0.90 (0.06)	0.06	0.95
$\beta_{trt_{34}}$	N(0,1)	0	0.007 (0.13)	0.14	0.98	N(0,4)	0	0.005 (0.13)	0.14	0.97	N(0,1)	0	0.002 (0.18)	0.18	0.95
$\beta_{Tr_{34}}$	N(0,1)	-0.1	-0.10 (0.06)	0.07	0.95	N(0,4)	-0.1	-0.10 (0.06)	0.07	0.96	N(0,1)	-0.1	-0.11 (0.08)	0.08	0.95
γ_0	N(0,1)	0.5	0.52 (0.13)	0.14	0.94	N(0,4)	0.5	0.54 (0.15)	0.14	0.92	N(0,1)	0.5	0.53 (0.15)	0.18	0.97
γ_{trt}	N(0,1)	0.6	0.60 (0.16)	0.18	0.98	N(0,4)	0.6	0.61 (0.17)	0.19	0.97	N(0,1)	0.6	0.58 (0.22)	0.24	0.97
γ_{stage}	N(0,1)	-0.9	-0.90 (0.16)	0.17	0.97	N(0,4)	-0.9	-0.92 (0.16)	0.17	0.96	N(0,1)	-0.9	-0.87 (0.20)	0.23	0.98

Where N(x,y)==Normal(mean, variance), G(x,y)==Gamma(mean, variance)

Where SD denotes standard deviation of posterior means. PSD denotes average of the posterior standard deviations.

CHAPTER V

Improving Efficiency in Clinical Trials Using Auxiliary Information; Application of a Multi-state Cure Model

5.1 Introduction

There is much interest in the use of intermediate outcome variables as either surrogate endpoints (Alonso and Molenberghs, 2008; Buyse and Molenberghs, 1998; Wang and Taylor, 2003) or auxiliary variables for the true outcome of interest in randomized clinical trials. A surrogate endpoint is one that is intended to replace the true outcome of interest in evaluating therapy and an auxiliary variable is one that is intended to be used to improve the efficiency of the analysis of the true endpoint. For clinical trials in locally advanced colon cancer, overall survival is traditionally considered the definitive endpoint. However, the earlier endpoint of disease-free survival, defined as the time to the first event of either death or cancer recurrence, has been determined to be a good surrogate for overall survival (Chen, *et al.* 1998; Sargent, *et al.* 2007). Therefore, disease free survival is now often used as the outcome in place of overall survival in clinical trials of colon cancer. Here, we explore an alternative use of recurrence in colon cancer trials, that of an auxiliary variable which can be

used to improve the efficiency of analysis, as measured by smaller standard errors on treatment effect estimates of the primary endpoint of interest, overall survival.

A variety of methods have been explored to utilize the information of an intermediate variable to improve the efficiency of the analysis of the final endpoint (Finkelstein and Schoenfeld, 1994; Fleming, *et al* 1994; Kosorok and Fleming, 1993; Lagakos, 1977). Cook and Lawless (2001) used a three-stage model for a time-to-event intermediate marker and true endpoint and showed that substantial gains in efficiency are possible with parametric models that assume a close structural relationship between the intermediate variable and true endpoint. Li, *et al.* (2011) used a parametric model formulation and showed an increase in efficiency gains in the analysis of the true endpoint when plausible prior assumptions are placed on certain model parameters. In particular, they showed that gains in efficiency can be made if the treatment effect on the true outcome, conditional on the intermediate variable, was adaptively shrunk towards zero. Broglio and Berry (2009) partitioned overall survival time into two parts, progression-free survival and survival post-progression and discuss the benefits of considering the treatment effects on each of these endpoints separately. In the situation of an auxiliary longitudinal variable and a censored event time of interest, Faucett, *et al.* (2002) developed an approach for using auxiliary variables to recover information from censored observations in survival analysis using a joint longitudinal and survival model and a multiple imputation procedure for the event times of censored subjects. Conlon, *et al.* (2011) considered the use of recurrence time as an auxiliary variable for overall survival by building separate models for time to recurrence and time to death. A cure model was used to model time to recurrence and a proportional hazards model with a Weibull baseline hazard function that included recurrence as a time dependent covariate was used to model death. The model for

time to death was then used in a multiple imputation procedure to impute death times for censored subjects, and these new data were used in the primary analyses on overall survival. Using the same data as considered in this paper, they showed modest but consistent gains in efficiency obtained by using the auxiliary information in recurrence times. Here, we extend this idea by building a joint multi-state model with an incorporated cured fraction for recurrence and death and use this model to impute death times for censored subjects with the goal of improving the efficiency of the analysis on overall survival. The model proposed here, while more complex and more difficult to estimate than the model used by Conlon, *et al.* (2011), utilizes the full data likelihood rather than a two-step procedure and offers the potential for larger gains in efficiency. The multi-state model also allows for adaptation to the imputation model on any of the individual sub-models that may lead to further efficiency gains.

The model that we propose for the recurrence and death events is a multi-state model with a cured fraction, described in detail in Chapter IV. This model is motivated by the disease process in colon cancer clinical trials. In these trials there are two outcomes of interest, recurrence and death, where death can occur either without prior recurrence or after a recurrence. Additionally, a proportion of subjects censored for recurrence may be cured of disease, and would therefore have never experienced a recurrence even if they had been followed for longer. For other censored subjects, their recurrence time would occur after their censoring time with longer follow-up and is therefore unobserved. To model these data, we use a multi-state model with an incorporated cured fraction that jointly models the probability of being cured of disease and the hazard of transitioning between disease states. A brief description of this model is given here, with full details given in Chapter IV of this dissertation.

Our model includes four hazards for transitioning between the four disease states

which include: alive and cured of disease, alive and uncured of disease, alive with recurrence and death. Transitions between these states are described by the multi-state model. Multi-state models (Anderson and Keiding, 2002; Meira-Machado, *et al.* 2009; Putter, Fiocco and Geskus, 2007) are a common way to jointly model disease progression events by describing the progressions and transitions over time to the various disease states through transition intensities, which can include covariates, and provide the hazards for moving between states. The hazard of each transition is modeled using a proportional hazard model with Weibull baseline hazard function. The cured fraction is modeled using the mixture model formulation of the cure model, introduced by Berkson and Gage (1952).

The proposed parametric model itself can be used to obtain efficiency gains relative to Kaplan-Meier estimates in the estimation of quantities of interest such as the difference in five year survival and the difference in three year disease free survival between treatment arms can be derived. Once parameter estimates from the model are obtained, the estimated five year survival and three year disease free survival can be computed from the model, with the point estimates and standard errors compared to the five year Kaplan-Meier survival estimate and the three year Kaplan-Meier disease free survival estimate, respectively. In addition to gains in efficiency due to the parametric assumptions, the multi-state model incorporates recurrence information which also contributes to efficiency gains in estimating this quantity from the model as compared to the Kaplan-Meier estimate.

In an alternative way to gain efficiency in the estimation of overall survival, the model can be used in a slightly weaker way by utilizing it in a multiple imputation procedure to impute death times for censored subjects. Patients who are alive at the time of their last follow-up are right censored for death, which we consider as a form

of missing data. A patient's recurrence status prior to their censoring time is usually known, and those who experience a recurrence are likely to die sooner than those who are recurrence free. Therefore, the information on a patient's recurrence time and status may be useful in predicting their survival time.

Multiple imputation is a common strategy for dealing with missing data problems (Rubin, 1978) and has been used to impute missing event times for censored observations in survival analysis (Faucett, *et al.* 2002; Hsu, *et al.* 2006). This strategy fills in missing values by drawing from the posterior predictive distribution of the missing data given the observed data. The procedure is then independently repeated M times to produce separate datasets. These completed datasets can then be analyzed separately to get estimates of overall survival, such as Cox model estimates and Kaplan-Meier estimates and their standard errors, and log-rank tests. The results from each of these analyses are then combined following established rules (Rubin, 1987), with potential gains in efficiency obtained as compared to an analysis of the original data.

We explore the use of some plausible restrictions that may be placed on model parameters and the effect of these restrictions on the amount of efficiency gained in the final analysis of the imputed data and in the estimation of model derived quantities relating to survival. Additionally, we can obtain model based survival estimates and apply the imputation procedure using estimates obtained from a hierarchical model that facilitates the sharing of information from covariates with consistent effects across trials and assess efficiency gains. The efficiency gains obtained in the treatment effect estimates offer the potential for the length of trials to be shorter and for the sample size of trials to be smaller.

The remainder of the paper is organized as follows: Section 5.2 describes the data

and Section 5.3 describes the proposed model. In Section 5.4, ways in which efficiency can be gained from the model or the trial can be shortened are explored. Section 5.5 provides details and results of the imputation procedure and simulation results are provided in Section 5.6. Section 5.7 concludes with a discussion.

5.2 Data Description

The data we consider consist of a total of 13,983 subjects from 12 randomized phase III adjuvant trials of locally advanced colon cancer. Ten of the trials are included in the Sargent, *et al.* (2005) publication, with two additional new trials. A detailed description of these data can be found in Chapter IV of this dissertation. These 12 trials were previously analyzed by Conlon, *et al.* (2011), where a separate cure model and Weibull model were used to model recurrence and death, respectively. Of the 4346 observed recurrences, 3448 (79.3%) occurred within three years, 4075 (93.8%) occurred within five years, and 4281 (98.5%) occurred within seven years. This type of event data where very few events happen after a fixed window of time is characteristic of a cured group. Kaplan-Meier plots of time to recurrence show a clear leveling off, indicating that this is data for which a cure model is appropriate. For subjects who experienced a recurrence, 44.5% died within one year, 67.9% died within 2 years and 78.9% died within 3 years. Table 5.1 provides Kaplan-Meier estimates of three year survival after recurrence for each trial. These estimates range from a survival probability of 0.10 in trial 6 and 0.30 in trial 12, demonstrating the high likelihood of dying quickly after recurring. Baseline covariates include age, cancer stage and treatment arm. Each trial compared a different pair of treatments, with one defined as the control arm and the other as the experimental arm. Five of the trials

(1,2,3,6 and 7) compared surgery alone to surgery plus some form of chemotherapy. In the other seven trials, both arms contained surgery plus some form of chemotherapy. The primary goal of all 12 trials was to compare overall survival between the pairs of treatments. A summary of the stage, age and treatment distributions for each trial, as well as the number of recorded recurrences and deaths and longest follow-up time for each trial can be seen in Table 4.1.

Table 5.1: Kaplan-Meier estimates for three year survival after recurrence for 12 trials

	3 Year Survival After Recurrence (95% CI)
Trial 1	0.13 (0.08, 0.21)
Trial 2	0.13 (0.08, 0.20)
Trial 3	0.16 (0.12, 0.20)
Trial 4	0.15 (0.12, 0.19)
Trial 5	0.21 (0.17, 0.26)
Trial 6	0.10 (0.07, 0.15)
Trial 7	0.24 (0.19, 0.31)
Trial 8	0.15 (0.12, 0.19)
Trial 9	0.15 (0.12, 0.18)
Trial 10	0.19 (0.16, 0.22)
Trial 11	0.25 (0.21, 0.30)
Trial 12	0.30 (0.26, 0.34)

5.3 Multistate model

We model the data from the 12 trials using a multi-state model with a cured fraction, as described in Chapter IV. We provide a brief description of this model here. Full details of the model can be found in Chapter IV. The proposed model jointly models the recurrence and death events as well as a latent incorporated cured fraction for the recurrence event. Deaths can occur either without a prior recurrence or following a recurrence. The deaths that occur without a prior recurrence are

known not to be directly due to the regrowth of the cancer, while deaths following a recurrence may be due to the cancer or other causes. Cause of death is not available and not considered in our models. We use the multi-state model to model four transition intensities between four disease states, as described in Chapter IV, and illustrated in Figure 4.1. We define State 1 to be alive and cured of disease, State 2 to be alive and not cured, State 3 to be alive with recurrence and State 4 to be death. We model four transition times which include $1 \rightarrow 4$, $2 \rightarrow 3$, $2 \rightarrow 4$, and $3 \rightarrow 4$.

5.3.1 Notation and model specifications

Let C_{ir} and T_{ir} be the censoring and event times for recurrence and let C_{id} and T_{id} be the censoring and event times for death for the ith subject, $i = 1, ...n$. Then $Y_{ir} = min(C_{ir}, T_{ir})$ and the event indicator for recurrence, $\delta_{ir} = I(T_{ir} \leq C_{ir})$, and $Y_{id} = min(C_{id}, T_{id})$ and the event indicator for death, $\delta_{id} = I(T_{id} \leq C_{id})$, are observed. Let Z_i, S_i and A_i represent the baseline values of treatment group, cancer stage and age for each subject.

Both the models for the time of entry into each state and for the probability of cure, p, can depend on covariates. The multi-state process is characterized through transition intensities defined as:

$$\lambda_{kj}(t) = \lim_{\Delta t \rightarrow 0} p_{kj}(t, t + \Delta t)/\Delta t$$

where $p_{kj}(s,t) = P(X(t) = j | X(s) = k, H_{s-})$, for $s \leq t$ is the probability of being in State j at time t, given that the process was in State k at time s and the history of the process, H_{s-}. $\lambda_{kj}(t)$ is then the instantaneous hazard of entering State j, given that the previous state occupied was State k. From this hazard, we can define the survival distributions for each transient state and their probability density functions.

We use a proportional hazards model with a Weibull baseline hazard function. Specifically, the hazard for transition kj for subject i is given by:

$$\lambda_{kj}(t_i; X_i) = \left(\frac{\rho_{kj}}{\alpha_{kj}}\right)\left(\frac{t_i}{\alpha_{kj}}\right)^{\rho_{kj}-1} exp\left(X_i\beta_{kj}\right)$$

and the probability of being cured, p, is modeled using a logistic link function given by:

$$p_i = \frac{exp(X_i\gamma)}{1+exp(X_i\gamma)}.$$

For transitions $1 \rightarrow 4$ and $2 \rightarrow 4$, t_i is a death time. For transition $2 \rightarrow 3$, t_i is a recurrence time and for transition $3 \rightarrow 4$, t_i is the gap time between entry into the recurred state and death. X_i represents a vector of subject specific covariates. For transitions $1 \rightarrow 4$, $2 \rightarrow 3$, and $2 \rightarrow 4$, and for the probability of cure we include the covariates age, treatment group and stage. For transition $3 \rightarrow 4$, we include these variables as covariates and also include recurrence time as a covariate.

5.3.2 Estimation

We use a Bayesian MCMC technique to estimate the parameters of the multi-state model. There are a total of 25 parameters to estimate for each of the trials which include a shape (ρ) and scale (α) parameter from the Weibull model for each of the hazard rates, covariate effects for each of the hazard models and covariate effects in the logistic model for the probability of cure. We place informative Normal(0,0.25^2) priors on the treatment and stage coefficients in transition $1 \rightarrow 4$ as treatment group and cancer stage are unlikely to have much affect on the hazard of death in patients who are cured of disease. We place Normal(0,2^2) priors on the log(α)'s and gamma priors with mean 1 and standard deviation 0.6 on the ρ's. Normal(0,1) are placed

on all of the remaining covariate coefficients in the hazard models and in the logistic model. The impact of these mildly informative priors is evaluated in Chapter IV of this dissertation. To aid in estimation, at each iteration of the chain, subject's without recurrence are placed in either the cured or uncured group by drawing a Bernoulli(c_i) random variable, where c_i is the probability of being cured of disease, given Y_{id}, Y_{ir}, X_i and the current parameter draws.

A Bayesian estimation scheme using the Metropolis Hastings algorithm is used to obtain parameter draws from the posterior distribution. Appendix J provides the full data likelihood and details of the algorithm. For each parameter, we obtain 5000 draws from its posterior distribution.

5.4 Efficiency gains from the model

The typical analysis of the treatment effect on overall survival would be estimates of hazard ratios using Cox models and estimates of differences in survival at specific times points. Since the multi-state cure model does not result in the proportional hazard being satisfied for time to death, we focus mainly on estimates of overall survival. Once parameter estimates for the model have been obtained, the multi-state cure model can be used to estimate the difference in five year overall survival (OS) between the two treatment arms. The point estimate can be compared with the Kaplan-Meier estimate to check the model fit and the standard error estimate can be compared with that of the Kaplan-Meier estimate to assess gains in efficiency

through use of the multi-state cure model. Let

$$S_1(t) = exp\left(-\int_0^t \lambda_{14}(u)du\right) \text{ and}$$
$$S_2(t) = exp\left(-\int_0^t \lambda_{23}(u)du - \int_0^t \lambda_{24}(u)du\right),$$

which are the survival distributions for remaining in State 1 or State 2, respectively. Then, for each subject we can calculate their five year OS probability as:

$$P(T_i > 5|X_i, \theta) = p_i S_1(5) + (1-p_i)S_2(5) + (1-p_i)\int_0^5 S_2(u)\lambda_{23}(u)exp\left(-\int_0^{5-u}\lambda_{34}(v)dv\right)du$$

where θ is the vector of parameter values. This probability is calculated separately for subjects in the treatment group and in the control group, and then averaged across the stage and age covariate values for each subject and across all of the parameter posterior draws to obtain a population estimate. Similarly, three year disease free survival (DFS), defined as the time to the first event of either death or recurrence and often used as a surrogate for five year OS, can be calculated from the model and compared to the three year Kaplan-Meier DFS estimate to assess efficiency gains from using the proposed model at the earlier time point. For each subject, three year DFS is calculated as:

$$P(DFS_i > 3|X_i, \theta) = p_i S_1(3) + (1-p_i)S_2(3).$$

This probability is then averaged across covariate values for each subject and across all parameter draws. Using the above model derived quantities, we estimate the treatment effect for these two separate endpoints of interest, five year OS and three

year DFS. For trials such as these in locally advanced colon cancer, five year OS is often considered the definitive endpoint. In this setting, three year DFS has been determined to be a valid surrogate marker for five year OS. Therefore, there is interest in the treatment effect estimate at both of these endpoints.

5.4.1 Application of model for efficiency gains and shortening trial length: Model based estimates

The above modeling strategy could be used to shorten the length of a clinical trial. To illustrate this, we artificially censor each of the 12 trials at either two years after the last patient accrual, or at the minimum length of time after the last accrual that provides at least 5.5 years of patient follow-up time. This artificial censoring resulted in an overall 9.8% reduction in the number of recurrences across all trials compared to the original data with a maximum of 15.4% in Trial 3 and a minimum of 4.3% in Trial 9. The overall reduction in the number of deaths was 30.9% with a maximum of 44.2% in Trial 7 and a minimum of 21.0% in Trial 9. Appendix O provides accrual length and maximum follow-up times before and after the artificial censoring for the 12 trials. Estimates of five year OS and three year DFS can then be obtained using parameter estimates from the multi-state cure model on the reduced follow-up data. The point estimates and posterior standard deviations of these quantities can then be compared to the Kaplan-Meier estimates from the full follow-up data to assess gains in efficiency from using the multi-state model and whether these quantities can correctly be estimated using shorter follow-up data.

5.4.2 Model restrictions

Extensions and adaptations could be made to the proposed model that may provide a better fit to the data and provide gains in efficiency in estimating overall survival. Li, *et al.* (2011) showed that when an intermediate variable captures even just a modest amount of the treatment effect on the final outcome, efficiency gains of the estimated treatment effect on the final outcome can be achieved by shrinking the treatment effect estimate in the conditional distribution of the final outcome given the intermediate variable and treatment toward 0. In our setting, it is plausible that much of the treatment effect is captured in the recurrence event by affecting the probability of being cured of disease and the time to recurrence. Therefore, one strategy to potentially improve efficiency gains in the estimation of the treatment effect on overall survival is to fit the multi-state cure model with prior assumptions placed on the treatment effects of some transition times. Specifically, the treatment effect on time to death for those who are cured ($1 \rightarrow 4$ transition) and the treatment effect on time to death for those who are not cured but without recurrence ($2 \rightarrow 4$ transition) are likely close to zero as the treatment may affect the probability of being cured, but after this most likely has little or no effect on the hazard of death from other causes if the person does not die from cancer. The treatment effect on time to death after recurrence ($3 \rightarrow 4$ transition) is also likely near zero, as patients often go off treatment or start new treatment regimens after a recurrence. We fit one restricted model with the above mentioned treatment effects shrunk towards zero with the use of tighter prior distributions and another restricted model with these treatment effects forced to be zero. All other covariates in the logistic model for the probability of cure and in the transition time models are the same as the full model. The fit of

the restricted models can be compared to that of the full model by calculating the DIC values for each model. Table 5.2 provides the DIC values for comparing the full model to the models with the treatment effects for transitions $1 \rightarrow 4$, $2 \rightarrow 4$ and $3 \rightarrow 4$ either shrunk to zero or forced to be zero for the 12 trials for the full follow-up data and the reduced follow-up data, with the DIC of the best fitting model in bold. For a majority of the studies, a restricted model is preferred over the full model in both the full follow-up and reduced follow-up data.

Table 5.2: Model comparison by DIC values- Restrictions on treatment effect parameters β_{14}, β_{24}, β_{34}

	Full Follow-up			Reduced Follow-up		
	Complete Model	β_{14}, β_{24}, β_{34} Shrunk to 0	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	Complete Model	β_{14}, β_{24}, β_{34} Shrunk to 0	$\beta_{14} = \beta_{24} = \beta_{34} = 0$
Trial 1	1093.7	1093.6	**1092.4**	**894.1**	909.4	909.1
Trial 2	1789.4	**1789.2**	1792.0	1441.7	**1440.8**	1443.2
Trial 3	4206.5	**4197.2**	4204.3	2615.7	**2611.0**	2617.3
Trial 4	4392.7	4393.4	**4389.5**	3263.0	3144.5	**3142.7**
Trial 5	3572.5	3573.1	**3571.9**	2674.0	2635.7	**2609.6**
Trial 6	3607.8	3616.8	**3607.3**	**2346.1**	2361.4	2363.9
Trial 7	3176.4	3176.0	**3174.2**	**1992.0**	2010.3	2010.2
Trial 8	**4084.6**	4089.1	4085.2	**3210.6**	3244.2	3247.0
Trial 9	7677.2	7689.3	**7675.2**	6346.7	6338.3	**6292.6**
Trial 10	7706.4	7713.4	**7703.9**	5706.3	5683.1	**5667.2**
Trial 11	**5225.7**	5235.4	5229.9	3885.1	**3823.2**	3850.7
Trial 12	7172.9	7174.0	**7172.6**	6074.1	**6027.2**	6074.2

5.4.3 Hierarchical model

Another way to extend the use of the multi-state model and potentially improve upon the efficiency gains is to borrow information for other trials by use of a hierarchical model. The original multi-state models fit to each individual trial provide evidence for common effects of some covariates on the probability of cure and transition rates. In particular, the coefficients associated with age and stage in all of the sub-models were quite similar. In addition, the coefficients associated with T_r in the $3 \rightarrow 4$ transition and the shape parameters of the Weibull models were

similar across trials. We can therefore use a hierarchical model to borrow information across trials and shrink selected parameters towards common values. To illustrate this, we let $\rho_{skj} \sim N(\rho_{kj},\ \sigma^2_{\rho kj})I(\rho_{skj} \geq 0)$, $\beta_{STskj} \sim N(\beta_{STkj},\ \sigma^2_{\beta_{ST}kj})$, $\beta_{AGEskj} \sim N(\beta_{AGEkj},\ \sigma^2_{\beta_{AGE}kj})$, $\beta_{T_rs34} \sim N(\beta_{T_r34},\ \sigma^2_{\beta_{T_r}34})$, $\gamma_{STskj} \sim N(\gamma_{STkj},\ \sigma^2_{\gamma_{ST}kj})$, and $\gamma_{AGEskj} \sim N(\gamma_{AGEkj},\ \sigma^2_{\gamma_{AGE}kj})$, where $kj = \{12, 23, 24, 34\}$ corresponds to the transition and $s = 1, ..., 12$ represents the study number. We place Gamma hyper-priors with mean 1 and standard deviation 1 on ρ_{kj} and on $\sigma_{\rho kj}$, $\sigma_{\beta_{ST}kj}$, $\sigma_{\beta_{AGE}kj}$, $\sigma_{\beta_{T_r}kj}$, $\sigma_{\gamma_{ST}kj}$, and $\sigma_{\gamma_{AGE}kj}$ and $N(0, 2^2)$ hyper-priors on $\beta_{AGEkj}, \beta_{STkj}$, β_{T_rkj}, γ_{STkj}, and γ_{AGEkj}. The remaining parameters are independent across studies. For the full follow-up data, we borrow information across the trials for the above parameters by fitting the hierarchical model once using all 12 trials. For the reduced follow-up data, we fit the hierarchical model separately 12 times, each time with 1 trial artificially censored and the remaining 11 with their full follow-up data. The parameter estimates obtained from the hierarchical models can then be used in estimating five year OS and three year DFS.

Table 5.3 provides the Kaplan-Meier estimates and standard errors and multi-state model estimates and posterior standard deviations for five year OS and three year DFS for the full follow-up data and the reduced follow-up data using the full multi-state cure model, the multi-state cure model with restrictions on certain treatment effect coefficients, and the hierarchical model. For the 12 trials considered, both the five year OS estimates and the three year DFS estimates from the multi-state model are similar to the Kaplan-Meier estimates, with moderate gains in efficiency obtained by using the multi-state model, as seen by the smaller posterior standard deviations. There is also a small amount of additional efficiency gained for some trials in the estimation of five year OS using the restricted models. Estimating these

quantities using estimates from the hierarchical model does not, in general, result in efficiency gains, likely due to the fact that these are all randomized trials and thus estimates for age and stage are likely to be at most weakly correlated with the estimate for treatment. The point estimates from the reduced follow-up data tend to be near those from the full follow-up data for both five year OS and three year DFS, with posterior standard deviations that are very close to the standard errors of the Kaplan-Meier estimates from the full follow-up data, indicating that similar conclusions about treatment effects on these quantities would be drawn using the reduced follow-up data and the multi-state model estimates as compared to the full follow-up data Kaplan-Meier estimates.

5.5 Efficiency gains through imputation

5.5.1 Imputation strategy

An alternative way that the multi-state model with a cured fraction can be used to improve efficiency in the estimation of overall survival is through a multiple imputation strategy that imputes death times for people who are censored for death. Using the proposed model in a multiple imputation procedure is a less model dependent approach than the estimation procedure in Section 5.4.1 because the model is only used to aid in estimation of the missing data, with the end analysis of the original data augmented by the imputed data. The multiple imputation approach could be used to improve efficiency of the analysis of overall survival or to shorten the length of a clinical trial while still keeping the primary endpoint of overall survival.

The imputation procedure is performed as follows. For each set of parameter draws, θ, from the posterior distribution, we impute a death time from the resid-

Table 5.3: Kaplan-Meier treatment effect estimates (standard errors) and multi-state model estimates (posterior standard deviations) for 12 colon cancer trials

		$\Delta S(5)^*$		$\Delta DFS(3)^{**}$	
		Full Follow-up	Reduced Follow-up	Full Follow-up	Reduced Follow-up
Trial 1	Kaplan-Meier	0.090 (0.062)	0.098 (0.070)	0.122 (0.062)	0.132 (0.071)
	Full Model	0.105 (0.049)	0.107 (0.057)	0.150 (0.051)	0.141 (0.057)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.106 (0.047)	0.112 (0.054)	0.146 (0.052)	0.142 (0.055)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.121 (0.046)	0.116 (0.055)	0.153 (0.052)	0.142 (0.057)
	Hierarchical model	0.098 (0.045)	0.103 (0.055)	0.147 (0.054)	0.144 (0.056)
Trial 2	Kaplan-Meier	0.057 (0.049)	0.039 (0.049)	0.056 (0.047)	0.076 (0.050)
	Full Model	0.051 (0.042)	0.053 (0.043)	0.068 (0.042)	0.070 (0.042)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.051 (0.040)	0.048 (0.043)	0.066 (0.043)	0.066 (0.045)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.050 (0.038)	0.051 (0.041)	0.059 (0.042)	0.065 (0.046)
	Hierarchical model	0.046 (0.041)	0.043 (0.043)	0.064 (0.042)	0.062 (0.044)
Trial 3	Kaplan-Meier	0.074 (0.031)	0.115 (0.080)	0.110 (0.031)	0.210 (0.086)
	Full Model	0.072 (0.026)	0.050 (0.028)	0.113 (0.027)	0.121 (0.030)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.081 (0.026)	0.098 (0.029)	0.119 (0.027)	0.137 (0.030)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.086 (0.022)	0.124 (0.027)	0.118 (0.028)	0.138 (0.029)
	Hierarchical model	0.067 (0.027)	0.070 (0.029)	0.110 (0.029)	0.115 (0.029)
Trial 4	Kaplan-Meier	-0.023 (0.037)	0.027 (0.043)	0.004 (0.037)	-0.003 (0.042)
	Full Model	-0.024 (0.032)	-0.010 (0.027)	-0.015 (0.034)	-0.003 (0.032)
	β_{14}, β_{24}, β_{34} Shrunk to 0	-0.023 (0.031)	-0.009 (0.034)	-0.017 (0.032)	-0.005 (0.035)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	-0.011 (0.024)	-0.005 (0.031)	-0.020 (0.033)	-0.008 (0.033)
	Hierarchical model	-0.023 (0.031)	-0.014 (0.033)	-0.016 (0.032)	-0.007 (0.035)
Trial 5	Kaplan-Meier	-0.023 (0.031)	-0.009 (0.035)	-0.015 (0.031)	-0.026 (0.035)
	Full Model	-0.032 (0.027)	-0.029 (0.024)	-0.017 (0.028)	-0.018 (0.028)
	β_{14}, β_{24}, β_{34} Shrunk to 0	-0.026 (0.026)	-0.029 (0.031)	-0.013 (0.027)	-0.020 (0.032)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	-0.013 (0.020)	-0.017 (0.029)	-0.015 (0.028)	-0.017 (0.031)
	Hierarchical model	-0.029 (0.027)	-0.028 (0.029)	-0.013 (0.029)	-0.018 (0.030)
Trial 6	Kaplan-Meier	0.037 (0.036)	0.026 (0.042)	0.048 (0.035)	0.031 (0.041)
	Full Model	0.009 (0.030)	0.004 (0.032)	0.033 (0.029)	0.032 (0.038)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.009 (0.030)	0.010 (0.034)	0.028 (0.031)	0.032 (0.034)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.018 (0.025)	0.022 (0.031)	0.026 (0.030)	0.032 (0.033)
	Hierarchical model	0.009 (0.030)	0.010 (0.030)	0.034 (0.032)	0.039 (0.030)
Trial 7	Kaplan-Meier	0.080 (0.035)	0.122 (0.045)	0.037 (0.035)	0.082 (0.041)
	Full Model	0.039 (0.029)	0.051 (0.028)	0.040 (0.030)	0.043 (0.028)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.039 (0.026)	0.052 (0.033)	0.041 (0.029)	0.045 (0.033)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.028 (0.020)	0.029 (0.027)	0.038 (0.028)	0.032 (0.032)
	Hierarchical model	0.036 (0.029)	0.045 (0.030)	0.041 (0.031)	0.035 (0.030)
Trial 8	Kaplan-Meier	0.105 (0.028)	0.112 (0.031)	0.084 (0.028)	0.116 (0.031)
	Full Model	0.095 (0.025)	0.085 (0.024)	0.103 (0.026)	0.096 (0.025)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.096 (0.024)	0.105 (0.029)	0.103 (0.025)	0.107 (0.030)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.074 (0.019)	0.102 (0.027)	0.099 (0.024)	0.108 (0.028)
	Hierarchical model	0.092 (0.025)	0.089 (0.024)	0.101 (0.025)	0.093 (0.024)
Trial 9	Kaplan-Meier	0.034 (0.021)	0.050 (0.026)	0.032 (0.021)	0.042 (0.022)
	Full Model	0.034 (0.018)	0.028 (0.018)	0.039 (0.019)	0.042 (0.020)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.035 (0.018)	0.040 (0.021)	0.041 (0.019)	0.050 (0.021)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.028 (0.014)	0.043 (0.019)	0.039 (0.019)	0.049 (0.021)
	Hierarchical model	0.034 (0.018)	0.034 (0.018)	0.039 (0.019)	0.042 (0.019)
Trial 10	Kaplan-Meier	0.004 (0.019)	0.012 (0.021)	0.007 (0.018)	0.018 (0.022)
	Full Model	0.012 (0.016)	0.015 (0.015)	0.010 (0.017)	0.008 (0.017)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.011 (0.016)	0.018 (0.019)	0.009 (0.017)	0.010 (0.020)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.007 (0.013)	0.009 (0.018)	0.010 (0.016)	0.011 (0.020)
	Hierarchical model	0.013 (0.016)	0.016 (0.016)	0.010 (0.017)	0.008 (0.017)
Trial 11	Kaplan-Meier	-0.0001 (0.021)	0.026 (0.030)	-0.005 (0.021)	-0.033 (0.026)
	Full Model	0.011 (0.018)	0.017 (0.017)	-0.004 (0.019)	-0.008 (0.019)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.006 (0.017)	0.011 (0.021)	-0.006 (0.019)	-0.007 (0.022)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	-0.001 (0.012)	-0.008 (0.018)	-0.007 (0.019)	-0.009 (0.021)
	Hierarchical model	0.009 (0.017)	0.016 (0.019)	-0.006 (0.019)	-0.002 (0.020)
Trial 12	Kaplan-Meier	0.018 (0.017)	0.029 (0.019)	0.032 (0.017)	0.048 (0.020)
	Full Model	0.017 (0.015)	0.008 (0.013)	0.033 (0.017)	0.035 (0.016)
	β_{14}, β_{24}, β_{34} Shrunk to 0	0.018 (0.015)	0.020 (0.018)	0.034 (0.016)	0.041 (0.019)
	$\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.021 (0.010)	0.034 (0.015)	0.033 (0.016)	0.041 (0.019)
	Hierarchical model	0.016 (0.014)	0.016 (0.016)	0.032 (0.015)	0.033 (0.017)

$^*\Delta S(5) = P(T > 5|Z_i = 1) - P(T > 5|Z_i = 0)$

$^{**}\Delta DFS(3) = P(DFS > 3|Z_i = 1) - P(DFS > 3|Z_i = 0)$

ual survival distribution, $P(T_{id} > Y_{id} + a_i | T_{id} > Y_{id}, \delta_{id} = 0, Y_{ir}, \delta_{ir}, X_i, \theta)$, for each censored subject. Specifically, we set this function equal to a $u \sim U(0,1)$ random variable and solve for a_i, the imputed time to death after Y_{id} for each censored subject. For subjects with a recurrence prior to their censoring time ($\delta_{ir} = 1$), we solve $u = exp\left(-\int_{Y_{id}-Y_{ir}}^{Y_{id}+a_i-Y_{ir}} \lambda_{34}(u)du\right)$ for a_i. For subjects censored for recurrence ($\delta_{ir} = 0$) we first calculate their probability of being in the cured group by drawing a Bernoulli(c_i) random variable, where c_i is the probability of being cured of disease, given Y_{id}, Y_{ir}, X_i and the current parameter draws. For subjects censored for recurrence at Y_{id}, c_i is given by

$$c_i = \frac{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id})}{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) + (1-p_i)\lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id})}$$

and for those censored for recurrence at Y_{ir} prior to Y_{id}, c_i is given by

$$c_i = \frac{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Yid)}{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) + (1-p_i)\lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id}) + (1-p_i)\int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u) S_2(u) \lambda_{34}(Y_{id}-u)^{\delta_{id}} exp\left(-\int_0^{Y_{id}-u} \lambda_{34}(v)dv\right) du}.$$

For those placed in the cured group, we solve $u = exp\left(-\int_{Y_{id}}^{Y_{id}+a_i} \lambda_{14}(u)du\right)$ for a_i, and for those placed in the uncured group, we solve $u = \frac{g(Y_{id}+a_i)}{g(Y_{id})}$ for a_i, where:

$$\begin{aligned} g(t) &= P(T_{id} > t \mid \delta_{id} = 0, Y_{ir}, \delta_{ir} = 0, X_i, \theta) \\ &= exp\left(-\int_0^t \lambda_{23}(u)du - \int_0^t \lambda_{24}(u)du\right) + \\ &\int_{Y_{ir}}^t exp\left(-\int_0^v \lambda_{23}(u)du - \int_0^v \lambda_{24}(u)du\right) \lambda_{23}(v) exp\left(-\int_0^{t-v} \lambda_{34}(u)du\right) dv \end{aligned}$$

For each subject, we solve the appropriate equation using every 10th draw from the posterior distribution of the parameters, giving a total of 500 data sets with imputed

death times for censored subjects. The imputed death times are censored at the longest follow-up time for a study. With death as the endpoint of interest, these new imputed times are combined with the observed data and compared to analyses of the original data to assess efficiency gains. Specific estimates of interest that are compared include the treatment effect estimates from a Cox model (which also includes stage and age as covariates), the log rank statistics, and the five year Kaplan Meier survival estimates. Parameter estimates and standard errors of the Kaplan Meier estimate and from the Cox model are obtained using the rules for multiple imputation established by Rubin (1987). Log-Rank test Chi-Square statistics are combined using the methods of Li, et al. (1991). Table 5.4 provides results from the analyses on the original data and on the imputed data. The point estimates are consistent across the analyses, suggesting that there was no distortion of the results introduced by the imputation, but there is little gain in efficiency from using the imputed data. This is likely due to the fact that these trials all had good follow-up. In the following section, we demonstrate the potential of the model to shorten the length of a trial by artificially censoring the 12 trials at an earlier time point and demonstrating the recovery of lost information due to censoring through the imputation procedure.

5.5.2 Application of model for efficiency gains and shortening trial length: Multiple Imputation

The artificially censored data described in Section 5.4.1 can be used to illustrate the use of the multiple imputation procedure in shortening the length of a clinical trial. We use the multiple imputation procedure described in Section 5.5.1 on the reduced follow-up data with death as the endpoint of interest. These analyses are then compared to analyses of the original, full follow-up data to assess efficiency

Table 5.4: Analysis of the effect of treatment on survival, from original data, and data with imputation

Study	Data	Log-Rank P-Value	Cox model Log Hazard Ratio (SE)	5 year KM Estimate (SE)
1	Original	0.224	-0.28 (0.188)	0.090 (0.062)
	Imputed	0.149	-0.31 (0.183)	0.092 (0.062)
2	Original	0.094	-0.25 (0.155)	0.057 (0.049)
	Imputed	0.105	-0.24 (0.154)	0.054 (0.048)
3	Original	0.004	-0.31 (0.098)	0.074 (0.031)
	Imputed	0.007	-0.30 (0.098)	0.073 (0.031)
4	Original	0.642	0.05 (0.111)	-0.023 (0.037)
	Imputed	0.704	0.05 (0.109)	-0.021 (0.037)
5	Original	0.352	0.09 (0.109)	-0.023 (0.031)
	Imputed	0.464	0.06 (0.108)	-0.021 (0.031)
6	Original	0.804	-0.04 (0.101)	0.037 (0.036)
	Imputed	0.734	-0.04 (0.100)	0.037 (0.036)
7	Original	0.075	-0.19 (0.115)	0.080 (0.035)
	Imputed	0.070	-0.21 (0.114)	0.077 (0.035)
8	Original	0.0003	-0.37 (0.103)	0.105 (0.028)
	Imputed	0.0004	-0.35 (0.102)	0.105 (0.028)
9	Original	0.037	-0.16 (0.077)	0.034 (0.021)
	Imputed	0.026	-0.17 (0.077)	0.034 (0.021)
10	Original	0.855	-0.02 (0.076)	0.004 (0.019)
	Imputed	0.788	-0.02 (0.076)	0.004 (0.018)
11	Original	0.838	0.008 (0.096)	-0.0001 (0.021)
	Imputed	0.930	0.02 (0.095)	-0.002 (0.021)
12	Original	0.090	-0.14 (0.086)	0.018 (0.017)
	Imputed	0.080	-0.14 (0.086)	0.018 (0.017)

gains. Table 5.5 provides log rank statistics, Cox model log hazard ratios and five year Kaplan-Meier estimates from the original, artificially censored and imputed data. The log rank tests and Cox models were stratified by cancer stage and the Cox models also included age as a covariate. The imputation procedure on the reduced follow-up data is performed using estimates obtained from the full multi-state cure model, the multi-state cure model with restrictions on some treatment effect coefficients, and the hierarchical model. The point estimates from the imputed data tend to be in between those of the original data and the reduced follow-up data, indicating that some of the information lost due to early censoring was correctly recovered using the imputation procedure. Gains in efficiency in the estimation of the log hazard ratio was achieved for some trials, as indicated by the smaller standard errors. The standard errors of the Kaplan-Meier estimates from the imputed data are consistently

smaller than those of the artificially censored data, and in many cases are very close to the standard errors of the original data. The point estimates resulting from the imputation procedure on the restricted models are nearly identical to those obtained using the full multi-state model, and for most trials, the standard errors from the imputation procedure on the restricted models are the same, or only barely smaller than those obtained through the use of the full model. This is likely because for these trials, much of the efficiency lost due to early censoring was recovered through the imputation procedure that uses the recurrence time information with the full multi-state model, leaving little further efficiency to be gained through the use of the more restricted models. There is no gain in efficiency in estimating the treatment effects on overall survival by using hierarchical model, probably due to the fact that these are all randomized trials and thus estimates for age and stage are likely to be at most weakly correlated with the estimate for treatment. Hence improving estimates for age and stage through the use of the hierarchical model could have limited impact on summary measures of the treatment effect. Table 5.5 also provides results using the modeling and imputation procedure of Conlon, *et al.* (2011). For these 12 trials, the simpler method of Conlon, *et al.* (2011) which models overall survival with recurrence as a time dependent covariate and bases the multiple imputation procedure off of this model appears to perform similarly to the more complex multi-state model with a cure fraction.

5.6 Simulations

We conduct simulations to examine the performance of the proposed imputation method using the multi-state cure model under the full and restricted models. We

Table 5.5: Analysis of the effect of treatment on survival, from original data, censored data and censored data with imputation

Study	Data	Log-Rank P-Value	Cox model Log Hazard Ratio (SE)	5 year KM Estimate (SE)
1	Original	0.136	-0.28 (0.188)	0.090 (0.062)
	Censored	0.035	-0.45 (0.214)	0.098 (0.070)
	Imputed Censored	0.047	-0.39 (0.197)	0.104 (0.063)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.045	-0.39 (0.197)	0.105 (0.063)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.040	-0.40 (0.197)	0.105 (0.063)
	Imputed Censored, hierarchical model	0.057	-0.38 (0.203)	0.102 (0.063)
	Conlon, *et al.* (2011) method	0.117	-0.31 (0.199)	0.101 (0.065)
2	Original	0.097	-0.25 (0.155)	0.057 (0.049)
	Censored	0.255	-0.20 (0.179)	0.039 (0.049)
	Imputed Censored	0.187	-0.23 (0.176)	0.051 (0.049)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.199	-0.23 (0.177)	0.051 (0.049)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.179	-0.24 (0.176)	0.051 (0.049)
	Imputed Censored, hierarchical model	0.215	-0.22 (0.178)	0.048 (0.049)
	Conlon *et al.* (2011) method	0.203	-0.22 (0.175)	0.053 (0.050)
3	Original	0.002	-0.31 (0.098)	0.074 (0.031)
	Censored	0.045	-0.27 (0.131)	0.115 (0.080)
	Imputed Censored	0.036	-0.26 (0.124)	0.076 (0.033)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.010	-0.31 (0.118)	0.082 (0.033)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.005	-0.33 (0.117)	0.092 (0.033)
	Imputed Censored, hierarchical model	0.027	-0.27 (0.122)	0.072 (0.033)
	Conlon *et al.* (2011) method	0.039	-0.25 (0.118)	0.068 (0.034)
4	Original	0.719	0.06 (0.111)	-0.023 (0.037)
	Censored	0.912	-0.005 (0.134)	0.027 (0.043)
	Imputed Censored	0.843	0.003 (0.132)	-0.005 (0.038)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.832	0.02 (0.130)	-0.006 (0.038)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.841	0.01 (0.131)	-0.004 (0.038)
	Imputed Censored, hierarchical model	0.823	0.01 (0.132)	-0.002 (0.038)
	Conlon *et al.* (2011) method	0.739	0.05 (0.131)	-0.007 (0.038)
5	Original	0.355	0.09 (0.109)	-0.023 (0.031)
	Censored	0.459	0.10 (0.134)	-0.009 (0.035)
	Imputed Censored	0.374	0.11 (0.129)	-0.020 (0.032)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.405	0.10 (0.125)	-0.019 (0.032)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.459	0.09 (0.124)	-0.017 (0.032)
	Imputed Censored, hierarchical model	0.385	0.11 (0.128)	-0.019 (0.032)
	Conlon *et al.* (2011) method	0.443	0.09 (0.121)	-0.019 (0.034)
6	Original	0.695	-0.04 (0.101)	0.037 (0.036)
	Censored	0.518	-0.08 (0.126)	0.026 (0.042)
	Imputed Censored	0.451	-0.09 (0.122)	0.024 (0.037)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.384	-0.11 (0.121)	0.026 (0.038)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.387	-0.10 (0.119)	0.026 (0.037)
	Imputed Censored, hierarchical model	0.465	-0.09 (0.123)	0.023 (0.038)
	Conlon *et al.* (2011) method	0.578	-0.06 (0.119)	0.019 (0.036)
7	Original	0.053	-0.20 (0.115)	0.080 (0.035)
	Censored	0.027	-0.33 (0.156)	0.122 (0.045)
	Imputed Censored	0.027	-0.31 (0.147)	0.079 (0.036)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.026	-0.31 (0.144)	0.078 (0.036)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.037	-0.28 (0.142)	0.074 (0.036)
	Imputed Censored, hierarchical model	0.026	-0.31 (0.147)	0.077 (0.036)
	Conlon *et al.* (2011) method	0.014	-0.35 (0.146)	0.081 (0.037)
8	Original	0.0004	-0.36 (0.103)	0.105 (0.028)
	Censored	0.0005	-0.41 (0.119)	0.112 (0.031)
	Imputed Censored	0.0005	-0.40 (0.117)	0.103 (0.029)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.0005	-0.40 (0.115)	0.104 (0.029)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.0005	-0.40 (0.115)	0.103 (0.029)
	Imputed Censored, hierarchical model	0.0005	-0.40 (0.117)	0.103 (0.029)
	Conlon *et al.* method	0.0004	-0.41 (0.116)	0.103 (0.030)
9	Original	0.041	-0.16 (0.077)	0.034 (0.021)
	Censored	0.105	-0.14 (0.087)	0.050 (0.026)
	Imputed Censored	0.083	-0.15 (0.086)	0.035 (0.021)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.080	-0.15 (0.086)	0.035 (0.021)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.071	-0.15 (0.085)	0.035 (0.021)
	Imputed Censored, hierarchical model	0.082	-0.15 (0.086)	0.035 (0.021)
	Conlon *et al.* (2011) method	0.113	-0.14 (0.086)	0.035 (0.021)
10	Original	0.827	-0.02 (0.076)	0.004 (0.019)
	Censored	0.398	-0.08 (0.092)	0.012 (0.021)
	Imputed Censored	0.502	-0.06 (0.091)	0.009 (0.019)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.511	-0.06 (0.089)	0.008 (0.019)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.585	-0.05 (0.088)	0.007 (0.019)
	Imputed Censored, hierarchical model	0.488	-0.07 (0.091)	0.009 (0.019)
	Conlon*et al.* (2011) method	0.505	-0.06 (0.088)	0.010 (0.019)
11	Original	0.907	0.007 (0.096)	-0.0001 (0.021)
	Censored	0.456	-0.08 (0.118)	0.026 (0.030)
	Imputed Censored	0.623	-0.05 (0.117)	0.005 (0.022)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.758	-0.01 (0.114)	0.003 (0.022)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.812	0.008 (0.113)	-0.0002 (0.022)
	Imputed Censored, hierarchical model	0.672	-0.03 (0.116)	0.007 (0.022)
	Conlon *et al.* (2011) method	0.622	-0.05 (0.111)	0.007 (0.022)
12	Original	0.083	-0.14 (0.086)	0.018 (0.017)
	Censored	0.273	-0.09 (0.097)	0.029 (0.019)
	Imputed Censored	0.255	-0.09 (0.094)	0.019 (0.018)
	Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.194	-0.12 (0.092)	0.020 (0.018)
	Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.140	-0.12 (0.091)	0.023 (0.018)
	Imputed Censored, hierarchical model	0.226	-0.11 (0.095)	0.019 (0.017)
	Conlon *et al.* (2011) method	0.230	-0.10 (0.091)	0.018 (0.018)

compare the performance of the proposed method to that of Conlon, *et al.* (2011) where the imputation of death times was based on a survival model with a time dependent covariate for recurrence.

Recurrence times and death times were first simulated from the multi-state cure model to give "original data" with long follow up. These times were then censored at an earlier time to give "censored data". The modeling and imputation strategy are then performed on the "censored data" using the full model, restricted models, and the model of Conlon, *et al.* (2011) to give the "imputed censored data". We then assess the treatment effect on overall survival using the log-rank test, the estimated relative hazard from a Cox model, and the five year Kaplan-Meier survival estimate. Four different trial settings are explored, two with a treatment effect and two without a treatment effect. For each setting, we generate 500 data sets, each with 500 subjects per treatment arm, 750 subjects with stage 3 disease, and a five year accrual period with eight years of additional follow-up to provide the "original data". The "censored data" is obtained by censoring these data sets either two years after the last accrual (trials 1 and 3) or one year after the last accrual (trials 2 and 4) to provide a maximum of seven years or six years, respectively, of follow-up time. The probability of being cured of disease was first generated using $p_i = \frac{exp(\gamma_0+\gamma_1 Z_i+\gamma_2 S_i)}{(1+exp(\gamma_0+\gamma_1 Z_i+\gamma_2 S_i))}$, where Z_i denotes treatment group and S_i denotes stage. Each of these covariates are centered at 0 so that Z_i is equal to -0.5 for the control group and 0.5 for the treatment group and S_i is equal to -0.75 for stage 2 disease and 0.25 for stage 3 disease. We set $\gamma_0 = 0.8$, $\gamma_1 = -0.4$ and $\gamma_2 = -1$ in trials 1 and 2 and $\gamma_0 = 0.8$, $\gamma_1 = 0$ and $\gamma_2 = -1$ in trials 3 and 4. For those who are cured of disease, we then generate a death time using hazard model for transition $1 \rightarrow 4$ with $\log(\lambda_{14})= 4$, $\rho_{14} = 1.5$, and the treatment and stage effects set to 0. For those who are not cured of disease, we generate a recurrence time

using the hazard model for transition $2 \rightarrow 3$ with $\log(\lambda_{23})= 1$, $\rho_{23} = 1.5$, $\beta_{st_{23}} = 0.7$, and $\beta_{trt_{23}}$ equal to -0.3 in trials 1 and 2 and 0.0 in trials 3 and 4, and a death time using the hazard model for transition $2 \rightarrow 4$ with $\log(\lambda_{24})= 4$, $\rho_{24} = 1.5$ and the treatment and stage effects set to 0. If the death time for uncured subjects is less than the recurrence time, then a $2 \rightarrow 4$ transition is made at the death time and the recurrence is censored at the death time. If the recurrence time is less than the death time, then a $2 \rightarrow 3$ transition is made at that time. For those who recur, the time between their recurrence and death is generated using the hazard model for transition $3 \rightarrow 4$ with $\log(\lambda_{34})= 1.1$, $\rho_{34} = 0.9$ $\beta_{trt_{34}} = 0$, $\beta_{st_{34}} = 0.3$, and $\beta_{Tr} = -0.1$.

Tables 5.6 and 5.7 provide the size of the log-rank test and the average of the estimated log hazard ratio for the treatment coefficient from a Cox model, both stratified by stage, as well as the average Kaplan-Meier estimate of the difference in five year survival between the treatment and control group. The empirical standard deviations (SD) and average standard errors ($\bar{SE}$) for these estimates is also provided. Additionally, for the null cases (trials 3 and 4) coverage rates for the Cox model log hazard ratio estimate and Kaplan-Meier five year survival difference are given. For each trial, the first row provides estimates for the data with a long follow-up period following the accrual period, which we call the "original data". The second row provides estimates for the data where all subjects could have the maximum amount of follow-up time given in the artificially censored data. These two rows provide a basis of comparison for the estimates obtained from the imputation procedure. Comparison to the first row answers the question of whether or not the imputation procedure performed on the reduced follow-up data results in similar conclusions to those based on the full follow-up data, thus resulting in the potential to shorten the length of the trial. Comparison to the second row answers the statistical question of

the bias in the estimates from the imputation procedure which censors subjects at the maximum follow-up time, as compared to those obtained when all subjects, from the beginning of the study, can be followed for that length of time. We note that the Cox model estimates for the treatment effect differ between the first two rows. This is because the proportional hazards assumption for time to death is not satisfied and the first row is based on a much longer follow-up than the second row.

The results show that there is some efficiency gained by using the imputation procedure, and when there is no treatment effect, the procedure preserves type I error. We note that the size of the log-rank test is slightly over conservative for the multiply imputed data. This is likely related to the issue of uncongeniality discussed by Meng (1994) and Rubin (1996), where the model used in creating the imputed data sets and the model used for analyzing the imputed data differ. Here, the model used to create the imputations was based on the multi-state model and utilized information on recurrence to obtain the imputed survival times. In these settings where the imputation model and analysis model differ due to auxiliary information used in the imputation procedure, the inference with multiple imputation tends to be conservative, but more efficient than inference done without multiple imputation (Meng, 1994). This uncongeniality between the imputation and analysis model is also likely the cause of the slight discrepancy between the empirical standard deviations and average standard errors, where the standard errors tend to be overly conservative.

The simulations demonstrate that some of the information lost due to early censoring can be correctly recovered through the imputation procedure. In the settings where there is a treatment effect on overall survival (trials 1 and 2), the Cox model log hazard ratio estimates from the imputed data are in between that from the "censored data" and from the "original data", and very close to the estimates from the "7

year follow-up data" (in the case of trial 1) and the "6 year follow-up" data (in the case of trial 2). The Kaplan-Meier estimates of the difference in five year survival are estimated within minimal bias in all four trials, with some small gains in efficiency obtained through the imputation procedure, as seen in the smaller standard deviations and smaller average standard errors as compared to the reduced follow-up data. There is a small amount of additional efficiency gained in all four settings by using the restricted multi-state cure model that shrinks the treatment effect estimates β_{14}, β_{24} and β_{34} to zero and some further efficiency gained in all four settings by restricting these treatment effects to be 0. The method of Conlon, *et al.* (2011) has slightly smaller average standard errors for the log hazard ratio estimate than those from the multi-state cure models, but has larger empirical standard deviations, indicating that there is a small amount of efficiency gained through the use of the multi-state cure model.

Table 5.6: Multiple imputation simulation results: treatment effect

Data	Size of Log-Rank	Cox model Log Hazard Ratio (SD)	Log Hazard Ratio $\widehat{SE}$	$\Delta S(5)$ KM Estimate (SD)	$\Delta S(5)$ KM $\widehat{SE}$
		Trial 1: Treatment Effect, 2 year censored			
Original (max 13 year follow-up)	0.772	-0.30 (0.110)	0.111	0.064 (0.025)	0.025
7 year follow-up	0.778	-0.35 (0.126)	0.126	0.064 (0.025)	0.025
Censored (max 7 year follow-up)	0.731	-0.39 (0.155)	0.154	0.064 (0.028)	0.029
Imputed Censored	0.754	-0.37 (0.130)	0.142	0.063 (0.024)	0.026
Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.764	-0.37 (0.128)	0.142	0.063 (0.024)	0.026
Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.764	-0.37 (0.120)	0.140	0.063 (0.023)	0.026
Conlon, *et al.* (2011) method	0.792	-0.38 (0.133)	0.137	0.065 (0.025)	0.026
		Trial 2: Treatment Effect, 1 year censored			
Original (max 13 year follow-up)	0.772	-0.30 (0.110)	0.111	0.064 (0.025)	0.025
6 year follow-up	0.762	-0.36 (0.134)	0.134	0.064 (0.025)	0.025
Censored (max 6 year follow-up)	0.632	-0.41 (0.182)	0.176	0.067 (0.034)	0.033
Imputed Censored	0.678	-0.39 (0.152)	0.163	0.060 (0.025)	0.026
Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.700	-0.39 (0.141)	0.160	0.060 (0.023)	0.025
Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.719	-0.39 (0.132)	0.157	0.060 (0.021)	0.025
Conlon, *et al.* (2011) method	0.714	-0.40 (0.153)	0.156	0.060 (0.024)	0.025

Table 5.7: Multiple imputation simulation results: no treatment effect

Data	Size of Log-Rank	Cox model Log Hazard Ratio (SD)	Log Hazard Ratio $\widehat{SE}$	Coverage	$\Delta S(5)$ KM Estimate (SD)	$\Delta S(5)$ KM $\widehat{SE}$	Coverage
			Trial 3: No Treatment Effect, 2 year censored				
Original (max 13 year follow-up)	0.068	0.00 (0.117)	0.111	0.93	0.000 (0.026)	0.025	0.94
7 year follow-up	0.066	0.00 (0.131)	0.125	0.94	0.000 (0.026)	0.025	0.94
Censored (max 7 year follow-up)	0.052	0.00 (0.155)	0.152	0.95	0.000 (0.030)	0.029	0.93
Imputed Censored	0.040	0.00 (0.132)	0.140	0.96	0.000 (0.025)	0.026	0.96
Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.036	0.00 (0.130)	0.140	0.96	0.000 (0.024)	0.026	0.96
Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.028	0.00 (0.124)	0.138	0.97	0.000 (0.024)	0.026	0.97
Conlon, *et al.* (2011) method	0.062	0.00 (0.143)	0.136	0.93	0.000 (0.026)	0.026	0.94
			Trial 4: No Treatment Effect, 1 year censored				
Original (max 13 year follow-up)	0.068	0.00 (0.117)	0.111	0.93	0.000 (0.026)	0.025	0.94
6 year follow-up	0.064	0.00 (0.142)	0.133	0.93	0.000(0.026)	0.025	0.94
Censored (max 6 year follow-up)	0.046	0.00 (0.171)	0.174	0.93	0.000 (0.034)	0.033	0.96
Imputed Censored	0.018	-0.01 (0.143)	0.158	0.98	0.000 (0.023)	0.026	0.97
Imputed Censored, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk to 0	0.020	-0.01 (0.139)	0.157	0.98	0.000 (0.023)	0.025	0.98
Imputed Censored, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.018	-0.01 (0.131)	0.154	0.98	0.000 (0.022)	0.025	0.98
Conlon, *et al.* (2011) method	0.036	-0.01 (0.147)	0.154	0.96	0.001 (0.023)	0.025	0.93

$\Delta S(5) = P(T > 5 \mid Z_i = 1) - P(T > 5 \mid Z_i = 0)$

5.7 Discussion

In this chapter, we propose a modeling and imputation procedure to assess the use of cancer recurrence as an auxiliary variable that can be used to improve efficiency in the analysis of overall survival. We explore the effects of plausible restrictions on model parameters and explore the use of a hierarchical model to assess the potential for further efficiency gains. The results show modest but consistent gains in efficiency, as measured by smaller standard errors, by using the information from recurrence time, with the potential for further gains by adding more restrictions to the models in certain settings. The methods presented could be useful in shortening the planned length of a trial and in reducing sample sizes. Although the changes in the width of the uncertainty intervals are only modest, sample size requirements are driven by the square of the standard deviation. Hence, if the proposed methodology were adopted, the size of trials could be reduced by 10% to 20%. These methods could also be useful in aiding data safety and monitoring boards in deciding whether or not to end a trial at the time of interim analysis.

We have considered several different ways in which the multi-state cure model

can be used to improve efficiency in the analysis of overall survival. First, we explored analyses of survival using the model itself, with the parametric assumptions of the model and recurrence time contributing to gains in efficiency. Next we explored methods in which the model is utilized in a weak way along with recurrence time information in a multiple imputation procedure to impute death times for censored subjects, with treatment effect estimates of survival obtained by combining the analyses from the multiply imputed data sets. We then placed restrictions on certain model parameters and used this adapted model in the multiple imputation procedure. Lastly, we explored the effects of utilizing external data to obtain parameter estimates for the multi-state model that was then used in the imputation procedure. In the setting explored here, the first method of obtaining estimates directly from the model and the imputation procedure using the multi-state cure model with no restrictions or external data were found to be most effective in obtaining efficiency gains.

The standard error estimates in the simulations are slightly conservative compared to the empirical standard deviations. As we noted, this is likely due to the uncongenialty of the imputation and analysis models. Robins and Wang (2000) have proposed a variance estimator for multiple imputation that is consistent when the imputation and analysis models differ that could be used to obtain a less conservative estimate, however, it is computationally much more difficult to obtain. We have focused on the situation of colon cancer, where there is a strong relationship between recurrence time and death. Cook and Lawless (2001) have noted that gains in efficiency for the estimation of survival distributions are often small when the intermediate variable and survival time are not highly correlated. When the intermediate variable and true endpoint are closely related, the use of parametric models and reasonable assumptions

about the effect of covariates on individual processes of the disease may contribute to further gains in efficiency. For example, there may be settings in which all of the treatment effect is on the probability of being cured, or where all of the treatment effect is on the hazard of recurrence for those who are uncured. In these settings, adding restrictions to the treatment effect coefficients of the full model may play a larger role in contributing to gains in efficiency in the analysis of overall survival.

CHAPTER VI

Discussion

The identification of valid surrogate markers and the use of intermediate outcomes as auxiliary variables has important implications in the clinical trial setting. By utilizing early information, trials could be run faster and more cheaply, and the early outcome information could aid regulatory boards in making preliminary decisions about drug approval. We have considered the use of intermediate variables both as surrogate markers for the true outcome of interest and as auxiliary variables to improve the efficiency of the estimation of the final outcome. For candidate surrogate markers, we proposed modeling and validation methods to assess the surrogate value of S. For auxiliary variables, we proposed a joint model for recurrence and death in colon cancer and demonstrated ways in which the information on recurrence times could be used to improve efficiency in the estimation of overall survival. In this Chapter, we summarize the ideas presented in Chapters II, III, IV and V and discuss potential future work in the area of intermediate variables.

In Chapters II and III, we work under the "principal surrogacy" framework introduced by Frangakis and Rubin (2002) and propose quantities to assess surrogacy from the conditional distribution of the causal treatment effect on T given the causal

treatment effect on S. We first consider the scenario where the potential outcomes of S and the potential outcomes of T follow a multivariate normal distribution. A Bayesian estimation strategy is used to estimate the parameters in this model and, as some parameters in the model are unidentifiable, informative priors consistent with reasonable assumptions in the surrogate marker setting are used to aid in estimation. The assumptions made include restricting the unidentified correlation parameters to be positive, which seems plausible when the identifiable correlation parameters ($cor(S(0), T(0))$ and $cor(S(1), T(1))$) are positive, and assumptions that pertain to the relationship between the correlations of the across treatment arm surrogate and final outcomes ($cor(S(1), T(0))$ and $cor(S(0), T(1))$) and the other pairwise correlations. Specifically, we constrain the across treatment arm surrogate and final outcome correlations to be smaller than the other pairwise correlations with either a probability of 1 or probability of 0.8. It seems reasonable to assume that the across treatment arm correlations of S and T would be smaller than the correlation between S and T within the same treatment arm, or the correlation between $S(0)$ and $S(1)$ or $T(0)$ and $T(1)$. These assumptions, along with the requirement that the covariance matrix be positive definite, restrict the ranges of possible values for the unidentified parameters. A variety of quantities from the conditional distribution of $p(T(1)-T(0) \mid S(1)-S(0))$ are explored. The proposed quantities of γ_0 and γ_1, which are the intercept and slope parameters, respectively, of the causal treatment effect on T conditional on the causal treatment effect on S, are useful measures and easily interpretable, however proving $\gamma_0 = 0$, a necessary condition for a valid surrogate, is difficult to do in practice. The CEP graph is also a useful tool, as it provides a way to estimate the expected treatment effects on T when treatment effects on S are at relevant clinical values. The measures proposed all consider the distribution of the causal effect of treatment on

the outcome conditional on the causal effect of treatment on the surrogate and can be used in combination to provide evidence about the validity of S as a surrogate marker for T.

The ideas explored in Chapter II are extended in Chapter III to settings in which the potential outcomes of S and the potential outcomes of T do not arise from a multivariate normal distribution. Here, we consider an ordinal categorical variable as a surrogate for a censored time-to-event final outcome and use a Gaussian copula to model the joint distribution of the potential outcomes. We again explore the use of different prior distributions for unidentified parameters. The model is applied to data from a trial in advanced colorectal cancer, where disease progression is assessed as a surrogate for overall survival. Using the proposed model, the expected ratio of log survival times within each of the principal strata of $S(1) - S(0)$ can be estimated. The results obtained using the Gaussian copula model are compared to those that would have been obtained using the methods of Chapter II had the model been misspecified and the data analyzed as multivariate normal. The results show some gains in efficiency by fitting the Gaussian copula model using the more appropriate marginal distributions for the data than by assuming multivariate normality when it may not hold. Both the methods of Chapter II and Chapter III could be extended to settings where T is partially missing, and the Gaussian copula model could be applied to data arising from other, non-normal distributions such as the Poisson.

An interesting area of future research concerning principal surrogacy is in the relationship between the principal stratification framework and structural models. In the causal inference literature, there are two types of general approaches. One is based on potential outcomes in the principal stratification framework, and one is based on structural or graphical models. While VanderWeele (2011) has argued that the

principal stratification framework is more appropriate for the surrogacy question, it is also of interest to explore the structural model approach. Surrogacy validation under the principal surrogacy framework considers the distribution of the causal treatment effect on T conditional on the principal strata of S. However, due to the counterfactual nature of the principal surrogacy framework, assumptions must be made to aid in the estimation of unidentified parameters. Consider the following structural models:

$$
\begin{aligned}
S_i(0) &= \alpha_0 + \alpha_2 U_i + \epsilon^{S_i(0)} \\
S_i(1) &= \alpha_0 + \alpha_1 + \alpha_2 U_i + \epsilon^{S_i(1)} \\
T_i(0) &= \beta_0 + \beta_2 S_i + \beta_3 U_i + \epsilon^{T_i(0)} \\
T_i(1) &= \beta_0 + \beta_1 + \beta_2 S_i + \beta_3 U_i + \epsilon^{T_i(1)}
\end{aligned}
$$

where $U_i \sim N(0, \sigma_u^2)$ and is a confounder in the relationship between S and T, and $\epsilon^{S_i(0)} \sim N(0, \sigma_{S0}^2)$, $\epsilon^{S_i(1)} \sim N(0, \sigma_{S1}^2)$, $\epsilon^{T_i(0)} \sim N(0, \sigma_{T0}^2)$, and $\epsilon^{T_i(1)} \sim N(0, \sigma_{T1}^2)$ are uncorrelated errors terms. The parameters from these equations could be related to those of the multivariate normal model described in Chapter II. For the above structural models, we have $(S_i(0), S_i(1), T_i(0), T_i(1))^T$ is normal with mean

$$
\mu = \begin{pmatrix} \alpha_0 \\ \alpha_0 + \alpha_1 \\ \beta_0 + \beta_2 \alpha_0 \\ \beta_0 + \beta_1 + \beta_2(\alpha_0 + \alpha_1) \end{pmatrix} \text{ and variance}
$$

$$
\Sigma = \begin{pmatrix}
\sigma_u^2 + \sigma_{S_0}^2 & \alpha_2^2\sigma_u^2 & \alpha_2^2\beta_2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2 + \beta_2\sigma_{S0}^2 & \alpha_2^2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2 \\
 & \sigma_u^2 + \sigma_{S_1}^2 & \alpha_2^2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2 & \alpha_2^2\beta_2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2 + \beta_2\sigma_{S1}^2 \\
 & & \sigma_u^2(\beta_2^2 + \beta_3^2 + 2\beta_2\beta_3\alpha_2) + \beta_2^2\sigma_{S0}^2 + \sigma_{T0}^2 & 2\beta_2\beta_3\sigma_u^2\alpha_2 + \beta_2^2\alpha_2^2\sigma_u^2 + \beta_3^2\sigma_u^2 \\
 & & & \sigma_u^2(\beta_2^2 + \beta_3^2 + 2\beta_2\beta_3\alpha_2) + \beta_2^2\sigma_{S1}^2 + \sigma_{T1}^2
\end{pmatrix}.
$$

This model relates to the correlation parameters of the multivariate normal model of

Chapter II in following way:

$$\rho_s = \frac{\alpha_2^2\sigma_u^2}{\sqrt{\sigma_u^2+\sigma_{S0}^2}\sqrt{\sigma_u^2+\sigma_{S1}^2}}$$

$$\rho_{00} = \frac{\alpha_2^2\beta_2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2 + \beta_2\sigma_{S0}^2}{\sqrt{\sigma_u^2+\sigma_{S0}^2}\sqrt{\sigma_u^2(\beta_2^2+\beta_3^2+2\beta_2\beta_3\alpha_2)+\beta_2^2\sigma_{S0}^2+\sigma_{T0}^2}}$$

$$\rho_{01} = \frac{\alpha_2^2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2}{\sqrt{\sigma_u^2+\sigma_{S0}^2}\sqrt{\sigma_u^2(\beta_2^2+\beta_3^2+2\beta_2\beta_3\alpha_2)+\beta_2^2\sigma_{S1}^2+\sigma_{T1}^2}}$$

$$\rho_{10} = \frac{\alpha_2^2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2}{\sqrt{\sigma_u^2+\sigma_{S1}^2}\sqrt{\sigma_u^2(\beta_2^2+\beta_3^2+2\beta_2\beta_3\alpha_2)+\beta_2^2\sigma_{S0}^2+\sigma_{T0}^2}}$$

$$\rho_{11} = \frac{\alpha_2^2\beta_2\sigma_u^2 + \alpha_2\beta_3\sigma_u^2 + \beta_2\sigma_{S1}^2}{\sqrt{\sigma_u^2+\sigma_{S1}^2}\sqrt{\sigma_u^2(\beta_2^2+\beta_3^2+2\beta_2\beta_3\alpha_2)+\beta_2^2\sigma_{S1}^2+\sigma_{T1}^2}}$$

$$\rho_t = \frac{2\beta_2\beta_3\sigma_u^2\alpha_2 + \beta_2^2\alpha_2^2\sigma_u^2 + \beta_3^2\sigma_u^2}{\sqrt{\sigma_u^2(\beta_2^2+\beta_3^2+2\beta_2\beta_3\alpha_2)+\beta_2^2\sigma_{S0}^2+\sigma_{T0}^2}\sqrt{\sigma_u^2(\beta_2^2+\beta_3^2+2\beta_2\beta_3\alpha_2)+\beta_2^2\sigma_{S1}^2+\sigma_{T1}^2}}.$$

In the structural model setting, assumptions must be made to estimate the regression coefficients, as only α_0 and α_1 are fully identifiable from the observed data. Various functions of these parameters are identified, however, and some parameter values are restricted by the requirement that Σ be positive definite. We could therefore explore the relationship between the effects of the assumptions on the structural model parameters and the assumptions that were placed on the model in Chapter II to aid in estimation in the principal surrogacy setting, such as constraining certain correlation coefficients to be positive.

In Chapters IV and V, we propose a multi-state model with an incorporated cured fraction to model recurrence and survival in colon cancer. This model is then utilized in a multiple imputation strategy for censored death times that uses recurrence as an auxiliary variable to improve efficiency in the estimation of overall survival. The model and imputation strategy are applied to data from 12 randomized trials in colon cancer. First, in Chapter IV we describe the model and its application. The proposed

multi-state model with a cured fraction is motivated by the disease process of colon cancer, where there is known to be a significant proportion of patients whose tumors are completely eliminated by the treatment and are therefore considered cured of disease. The proposed multi-state model with an incorporated cured fraction can be used to examine the effects of different covariates on all of the various aspects of the disease process, including the probability of being cured of disease, time to recurrence for those who are uncured, time to death for those who are cured, time to death without recurrence for those who are uncured but die before experiencing a recurrence, and time to death after recurrence. We show consistent effects of many covariates across the 12 trials. Once parameter estimates for the model are obtained, quantities of interest such as the differences in five year survival and three year disease free survival between treatment arms can be estimated. We show that the point estimates of these quantities are consistent with the Kaplan-Meier estimates, with some efficiency gained through the use of the multi-state cure model. Additionally, we propose the use of Cox-Snell residual plots and deviance residual plots as ways to visually assess the adequacy of the model fit. In the formulation of this model, recurrence times are treated as known. The model could be easily adapted to reflect the more realistic scenario of recurrence times that are interval censored between clinic visits if this information were available. The model could also be extended to allow the possibility for patients to return to the disease-free state after recurrence for those who live a long time after their recurrence, likely due to subsequent therapy.

In Chapter V, we utilize the multi-state model with a cured fraction and propose a multiple imputation strategy for patients who are censored for death. By using the proposed joint model for recurrence and death, the information on recurrence can be used as an auxiliary variable in predicting the death times of censored subjects, with

the goal of improving efficiency in the estimation of the treatment effect on overall survival. We show the potential of the multiple imputation strategy to shorten the length of a clinical trial by artificially censoring the 12 trials and performing the multiple imputation procedure. The analyses of overall survival from the imputed data sets are combined, and the results are compared to the analyses of overall survival on the original data. We show that some of the information lost due to early censoring can be recovered through the imputation procedure and demonstrate gains in efficiency in the estimates obtained from the imputed data as compared to the artificially censored data. Additionally, we demonstrate ways in which model adaptations and a hierarchical model could be used to further gains in efficiency obtained through the imputation procedure. We show some small gains in efficiency through the imputation procedure using the proposed multi-state model with a cured fraction as compared to the simpler model used by Conlon, *et al.* (2011), where separate models were used for recurrence and death. The proposed model and imputation procedure could therefore be useful to data safety and monitoring boards in deciding whether or not to end a trial at the time of an interim analysis. Additionally, as sample size requirements are driven by the square of the standard deviation, the size of trials could be reduced by 10% to 20% by adopting the proposed methods. Other adaptations to the proposed model, besides those explored here, are possible and may improve efficiency gains obtained through the imputation procedure. For example, semi-parametric alternatives to the Weibull model could be explored. This may be especially useful in the transition between recurrence and death, where there appeared to be a potential lack of fit using the Weibull model. Using the generalized Weibull model described by Foucher, *et al.* (2005) or a semi-parametric alternative for this transition may improve the efficiency gains of the imputation procedure.

There is future work that could be explored in the area of intermediate markers. One concern that arises in the assessment of surrogate makers is the presence of the "surrogate paradox", where S and T are positively correlated, there is a positive treatment effect on S but the treatment effect on T is negative. As we noted in Chapter II, in the principal surrogacy setting if both average causal necessity and average causal sufficiency hold, then the surrogate paradox is avoided. Methods for detecting the presence of the surrogate paradox would be useful in the meta-analytic setting, where information is available from a large number of trials and, given the effect of the treatment on S in a new trial, we are interested in the expected effect of treatment on the outcome in this trial. In this setting, it would be useful to have a measure of the probability that an expected positive effect of the treatment on T given a positive effect of the treatment on S is not due to chance variation. Buyse, *et al.* (2000) proposed the following bivariate mixed model to describe the joint distribution of S and T in the meta-analytic setting:

$$S_{ij} = \alpha_S + \beta_S Z_{ij} + a_{S_i} + b_{S_i} Z_{ij} + \epsilon_{S_{ij}}$$

$$T_{ij} = \alpha_T + \beta_T Z_{ij} + a_{T_i} + b_{T_i} Z_{ij} + \epsilon_{T_{ij}}$$

for subject j in trial i, where

$$\begin{pmatrix} \epsilon_{S_{ij}} \\ \epsilon_{T_{ij}} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma = \begin{pmatrix} \sigma_{ss} & \sigma_{st} \\ & \sigma_{tt} \end{pmatrix} \right) \text{ and}$$

$$\begin{pmatrix} a_{S_i} \\ a_{T_i} \\ b_{S_i} \\ b_{T_i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, D = \begin{pmatrix} d_{ss} & d_{st} & d_{sa} & d_{sb} \\ & d_{tt} & d_{ta} & d_{tb} \\ & & d_{aa} & d_{ab} \\ & & & d_{bb} \end{pmatrix} \right).$$

We could consider the joint distribution of the treatment effect on S, $\beta_S + b_{S_i}$, and the treatment effect on T, $\beta_T + b_{T_i}$, across the trials, and estimate the proportion

of the CDF of this distribution that corresponds to surrogate consistency (where the treatment effect on both S and T is positive and where the treatment effect on both S and T is negative) relative to the proportion of the CDF where the surrogate paradox would occur (opposite treatment effects on S and T). The proportion of the CDF corresponding to the region where the surrogate paradox does not occur is given by $1 - \Phi_1(0 \mid \beta_S, d_{aa}) - \Phi_1(0 \mid \beta_T, d_{bb}) + 2\Phi_2((0,0) \mid (\beta_S, \beta_T), \Gamma)$, where $\Phi_k(x \mid \Theta, \Psi)$ is the CDF of a $k-$variate normal distribution with mean Θ and variance Ψ evaluated at x and $\Gamma = \begin{pmatrix} d_{aa} & d_{ab} \\ & d_{bb} \end{pmatrix}$. This approach could also be used to focus on subgroups of interest, and extended to non-normal data through the use of copulas or kernel density estimators.

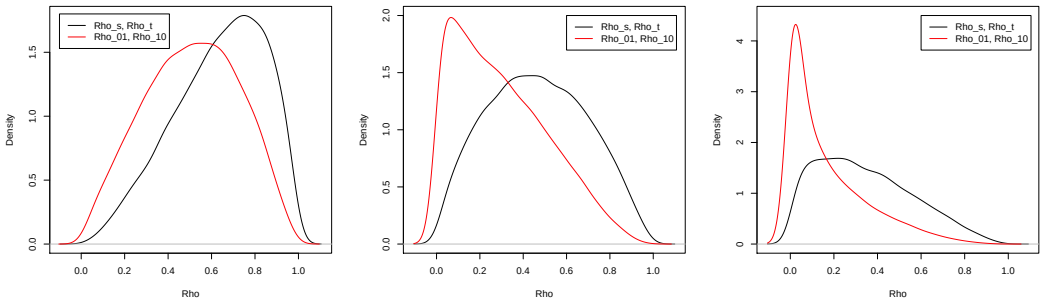
APPENDICES

APPENDIX A

Prior densities for MVN model parameters

Figure A.1: Density plots for MVN model correlation parameters under Beta priors

(a) $\hat{\rho_{00}} = \hat{\rho_{11}} = 0.8$ (b) $\hat{\rho_{00}} = \hat{\rho_{11}} = 0.5$ (c) $\hat{\rho_{00}} = \hat{\rho_{11}} = 0.3$



APPENDIX B

Quadratic equations obtained from $|R| = 0$

Each component of Q and R are drawn one at a time. When drawing each element of R, the range of possible values must first be determined in order to satisfy the positive definite requirement, given that the other correlations are held fixed. The range of values corresponding to a positive definite matrix are those in the interval determined by the roots of the quadratic equation that result from solving $|R| = 0$. For each ρ we have $\rho = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, where a, b, and c for each correlation is given in the following table.

Table B.1: Quadratic Equation Elements for Correlation Ranges

Correlation	a	b	c
ρ_s	$\rho_t^2 - 1$	$2\rho_{00}\rho_{10} + 2\rho_{01}\rho_{11} - 2\rho_{00}\rho_{11}\rho_t - 2\rho_{01}\rho_{10}\rho_t$	$1 - \rho_{00}^2 - \rho_{01}^2 - \rho_{10}^2 - \rho_{11}^2 - \rho_t^2 + \rho_{00}^2\rho_{11}^2 + \rho_{10}^2\rho_{01}^2 + 2\rho_{10}\rho_{11}\rho_t + 2\rho_{00}\rho_{01}\rho_t - 2\rho_{00}\rho_{11}\rho_{01}\rho_{10}$
ρ_{00}	$\rho_{11}^2 - 1$	$2\rho_s\rho_{10} + 2\rho_{01}\rho_t - 2\rho_s\rho_{11}\rho_t - 2\rho_{01}\rho_{10}\rho_{11}$	$1 - \rho_t^2 - \rho_{10}^2 - \rho_{11}^2 - \rho_s^2 - \rho_{01}^2 + \rho_s^2\rho_t^2 + \rho_{01}^2\rho_{10}^2 + 2\rho_{10}\rho_{11}\rho_t + 2\rho_s\rho_{01}\rho_{11} - 2\rho_s\rho_{10}\rho_{01}\rho_t$
ρ_{01}	$\rho_{10}^2 - 1$	$2\rho_s\rho_{11} + 2\rho_{00}\rho_t - 2\rho_s\rho_{10}\rho_{11} - 2\rho_{00}\rho_{11}\rho_{10}$	$1 - \rho_s^2 - \rho_{00}^2 - \rho_{10}^2 - \rho_{11}^2 - \rho_t^2 + \rho_s^2\rho_t^2 + \rho_{00}^2\rho_{11}^2 + 2\rho_{10}\rho_{11}\rho_t + 2\rho_{00}\rho_{10}\rho_s - 2\rho_s\rho_{00}\rho_{11}\rho_t$
ρ_{10}	$\rho_{01}^2 - 1$	$2\rho_t\rho_{11} + 2\rho_{00}\rho_s - 2\rho_s\rho_{01}\rho_t - 2\rho_{00}\rho_{11}\rho_{01}$	$1 - \rho_s^2 - \rho_{00}^2 - \rho_{01}^2 - \rho_{11}^2 - \rho_t^2 + \rho_s^2\rho_t^2 + \rho_{00}^2\rho_{11}^2 + 2\rho_{01}\rho_{11}\rho_s + 2\rho_{00}\rho_{01}\rho_t - 2\rho_s\rho_{00}\rho_{11}\rho_t$
ρ_{11}	$\rho_{00}^2 - 1$	$2\rho_t\rho_{10} + 2\rho_{01}\rho_s - 2\rho_s\rho_{00}\rho_t - 2\rho_{00}\rho_{01}\rho_{10}$	$1 - \rho_t^2 - \rho_{10}^2 - \rho_s^2 - \rho_{00}^2 - \rho_{01}^2 + \rho_s^2\rho_t^2 + \rho_{01}^2\rho_{10}^2 + 2\rho_{10}\rho_{00}\rho_s + 2\rho_t\rho_{01}\rho_{00} - 2\rho_s\rho_{10}\rho_{01}\rho_t$
ρ_t	$\rho_s^2 - 1$	$2\rho_{10}\rho_{11} + 2\rho_{00}\rho_{01} - 2\rho_s\rho_{01}\rho_{10} - 2\rho_s\rho_{00}\rho_{11}$	$1 - \rho_s^2 - \rho_{00}^2 - \rho_{01}^2 - \rho_{10}^2 - \rho_{11}^2 + \rho_{00}^2\rho_{11}^2 + \rho_{10}^2\rho_{01}^2 + 2\rho_s\rho_{00}\rho_{10} + 2\rho_{01}\rho_s\rho_{11} - 2\rho_{00}\rho_{11}\rho_{01}\rho_{10}$

APPENDIX C

Gibbs sampler details for MVN model

As the posterior distributions for the components of Q and R can not be easily sampled from, draws are made using the griddy Gibbs sampler (Ritter and Tanner, 1992).

Let $Y = (S(0), S(1), T(0), T(1))$, and $\theta = (\mu, S, R)$. We impute the unobserved potential outcomes at step l given θ^{l-1} from the following distributions:

$$\begin{pmatrix} S_i(0) \\ T_i(0) \end{pmatrix} \mid S_i(1), T_i(1), \theta^{l-1} \sim N \left(\begin{pmatrix} \mu_{S_0} \\ \mu_{T_0} \end{pmatrix} + \Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} S_i(1) - \mu_{S_1} \\ T_i(1) - \mu_{T_1} \end{pmatrix}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

$$\begin{pmatrix} S_i(1) \\ T_i(1) \end{pmatrix} \mid S_i(0), T_i(0), \theta^{l-1} \sim N \left(\begin{pmatrix} \mu_{S_1} \\ \mu_{T_1} \end{pmatrix} + \Sigma_{12}\Sigma_{11}^{-1} \begin{pmatrix} S_i(0) - \mu_{S_0} \\ T_i(0) - \mu_{T_0} \end{pmatrix}, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \right)$$

Where

$$\Sigma_{11} = \begin{pmatrix} \sigma_{S_0}^2 & \rho_{00}\sigma_{S_0}\sigma_{T_0} \\ \rho_{00}\sigma_{S_0}\sigma_{T_0} & \sigma_{T_0}^2 \end{pmatrix}$$

$$\Sigma_{12} = \begin{pmatrix} \rho_s\sigma_{S_0}\sigma_{S_1} & \rho_{01}\sigma_{S_0}\sigma_{T_1} \\ \rho_{10}\sigma_{S_1}\sigma_{T_0} & \rho_t\sigma_{T_1}\sigma_{T_0} \end{pmatrix}$$

$$\Sigma_{21} = \begin{pmatrix} \rho_s \sigma_{S_0} \sigma_{S_1} & \rho_{10} \sigma_{S_1} \sigma_{T_0} \\ \rho_{01} \sigma_{S_0} \sigma_{T_1} & \rho_t \sigma_{T_1} \sigma_{T_0} \end{pmatrix}$$

$$\Sigma_{22} = \begin{pmatrix} \sigma^2_{S_1} & \rho_{11} \sigma_{S_1} \sigma_{T_1} \\ \rho_{11} \sigma_{S_1} \sigma_{T_1} & \sigma^2_{T_1} \end{pmatrix}$$

$\mu^{(l)}$ and each component of S and R are then drawn from their posterior distributions:

$$\mu^{(l)} | \Sigma^{l-1}, Y \sim N((n\Sigma^{-1(l-1)} + \Sigma_0^{-1})^{-1}(n\Sigma^{-1(l-1)}\bar{Y}), (n\Sigma^{-1(l-1)} + \Sigma_0^{-1})^{-1})$$

For $i = 1, ...4$

$$\sigma_{Y_i}|\cdot \propto \sigma_{Y_i}^{-n} exp\left(-\tfrac{1}{2}\sum_{i=1}^{n}(Y-\mu)\Sigma^{-1}(Y-\mu)'\right)$$

For $j = s, t, 01, 10, 11, 00$

$$\rho_j|\cdot \propto |R|^{-n/2} exp\left(-\tfrac{1}{2}\sum_{i=1}^{n}(Y-\mu)\Sigma^{-1}(Y-\mu)'\right) p(\rho_j),$$

where $p(\rho_j)$ corresponds to the prior distribution for ρ_j. The four different sets of priors considered are detailed in section 2.4.

As the posterior distributions for the σ's and ρ's can not be easily sampled from, the griddy Gibbs algorithm is used as follows:

- Evaluate $p(\sigma_{Y_i}|\cdot)$ over a grid of $m = 200$ points, separated by hundredths to obtain $x_1, x_2, ..., x_m$. The grid for each σ is centered around the estimated standard deviation from the observed data.
- Approximate the inverse cdf using the discrete approximation $p(\sigma_{Y_i j}) = x_j / \sum_{k=1}^{m} x_k$.
- Sample a uniform (0,1) random variable and transform the observation using the approximated cdf.
- Obtain posterior draws for each of σ_{S_0}, σ_{S_1}, σ_{T_0}, and σ_{T_1}
- Posterior draws for each ρ are done similarly, but with the grid of values over

which each posterior distribution is evaluated those which result in a positive definite matrix.

APPENDIX D

Robustness to multivariate normality assumption

Table D.1: Simulation results under MVN model when multivariate normality does not hold

	Normal Marginals			Multivariate Normal			Multivariate t_3			Multivariate Normal			Log Normal			Multivariate Normal		
Parameter	True Value	Mean (SD)	$P\tilde{S}D$	True Value	Mean (SD)	$P\tilde{S}D$	True Value	Mean (SD)	$P\tilde{S}D$	True Value	Mean (SD)	$P\tilde{S}D$	True Value	Mean (SD)	$P\tilde{S}D$	True Value	Mean (SD)	$P\tilde{S}D$
μ_{s_0}	0	0.002(0.08)	0.08	0	0.01 (0.08)	0.08	4	3.99 (0.13)	0.13	4	3.99 (0.13)	0.14	4.48	4.46 (0.50)	0.44	4.48	4.49 (0.49)	0.46
μ_{s_1}	1	1.00(0.08)	0.08	1	1.00 (0.09)	0.08	6	5.99 (0.13)	0.13	6	6.00 (0.14)	0.14	7.39	7.36 (0.81)	0.85	7.39	7.44 (0.88)	0.83
μ_{t_0}	0	0.0001(0.09)	0.08	0	0.007 (0.09)	0.08	8.4	8.39 (0.13)	0.13	8.4	8.39 (0.13)	0.14	4.48	4.43 (0.46)	0.49	4.48	4.51 (0.47)	0.48
μ_{t_1}	2	1.99(0.08)	0.08	2	1.99 (0.08)	0.08	10	9.99 (0.14)	0.13	10	10.00 (0.14)	0.14	6.69	6.65 (0.69)	0.77	6.69	6.73 (0.75)	0.75
σ_{s_0}	1	1.01(0.06)	0.06	1	1.01 (0.06)	0.06	1.7	1.61 (0.25)	0.09	1.7	1.69 (0.11)	0.10	5.87	5.41 (0.46)	0.18	5.87	5.67 (0.15)	0.20
σ_{s_1}	1	1.01(0.06)	0.06	1	1.02 (0.06)	0.06	1.7	1.59 (0.23)	0.09	1.7	1.68 (0.09)	0.10	9.69	10.46 (1.31)	0.35	9.69	10.14 (0.35)	0.44
σ_{t_0}	1	1.01(0.06)	0.06	1	1.00 (0.06)	0.06	1.7	1.62 (0.26)	0.09	1.7	1.69 (0.10)	0.10	5.87	5.99 (0.68)	0.19	5.87	5.88 (0.18)	0.23
σ_{t_1}	1	1.01(0.06)	0.06	1	1.01 (0.06)	0.06	1.7	1.61 (0.22)	0.09	1.7	1.68 (0.10)	0.10	8.76	9.39 (1.06)	0.31	8.76	9.17 (0.27)	0.41
ρ_{00}	0.28	0.28(0.10)	0.07	0.28	0.27 (0.07)	0.07	0.8	0.76 (0.10)	0.03	0.8	0.78 (0.03)	0.04	0.73	0.72 (0.09)	0.05	0.73	0.70 (0.04)	0.05
ρ_{11}	0.28	0.27(0.10)	0.07	0.28	0.28 (0.07)	0.07	0.8	0.76 (0.08)	0.03	0.8	0.78 (0.03)	0.03	0.72	0.76 (0.09)	0.04	0.72	0.72 (0.04)	0.04
ρ_s	0	0.37(0.07)	0.21	0	0.37 (0.06)	0.21	0.4	0.44 (0.09)	0.17	0.4	0.43 (0.08)	0.17	0.29	0.41 (0.09)	0.16	0.29	0.45 (0.08)	0.18
ρ_{01}	0	0.17(0.05)	0.16	0	0.17 (0.04)	0.16	0.3	0.41 (0.10)	0.18	0.3	0.40 (0.08)	0.18	0.22	0.36 (0.10)	0.18	0.22	0.40 (0.08)	0.19
ρ_{10}	0	0.18(0.06)	0.16	0	0.17 (0.04)	0.16	0.3	0.40 (0.09)	0.18	0.3	0.42 (0.08)	0.17	0.22	0.39 (0.10)	0.18	0.22	0.42 (0.08)	0.19
ρ_t	0	0.38(0.07)	0.21	0	0.37 (0.06)	0.21	0.4	0.42 (0.09)	0.17	0.4	0.41 (0.08)	0.17	0.28	0.39 (0.09)	0.18	0.28	0.43 (0.08)	0.18

Table D.2: Simulation results: Bias, variability and coverage rate of surrogacy parameters when multivariate normality does not hold

	Normal Marginals				Multivariate normal			
Parameter	True Value	Mean (SD)	$P\bar{S}D$	95% Coverage	True Value	Mean (SD)	$P\bar{S}D$	95% Coverage
β_1	1.72	1.71(0.14)	0.13	0.94	1.72	1.71 (0.13)	0.13	0.95
β_2	0.28	0.28(0.10)	0.07	0.83	0.28	0.27 (0.07)	0.07	0.95
β_3	0	-0.004(0.15)	0.10	0.80	0	0.005 (0.11)	0.11	0.94
γ_0	1.72	1.79(0.17)	0.30	0.99	1.72	1.80 (0.14)	0.29	0.99
γ_1	0.28	0.20(0.12)	0.27	0.99	0.28	0.19 (0.09)	0.27	1
ρ_{ST}	0.28	0.19 (0.11)	0.23	0.995	0.28	0.19 (0.08)	0.23	1
$\Phi_{10}(0)$	0.90	0.94 (0.02)	0.04	0.89	0.90	0.95 (0.02)	0.04	0.92
	Multivariate t_3				Multivariate Normal			
β_1	0	0.04 (0.86)	0.40	0.72	0	0.01 (0.37)	0.39	0.96
β_2	0.8	0.77 (0.13)	0.05	0.73	0.8	0.79 (0.05)	0.05	0.93
β_3	0	0.003 (0.18)	0.07	0.69	0	0.002 (0.07)	0.07	0.96
γ_0	-0.07	0.27 (0.38)	0.51	0.93	-0.07	0.22 (0.26)	0.53	0.995
γ_1	0.83	0.66 (0.18)	0.25	0.93	0.83	0.69 (0.11)	0.25	0.995
ρ_{ST}	0.83	0.63 (0.14)	0.18	0.90	0.83	0.66 (0.08)	0.18	0.96
$\Phi_{10}(0)$	0.47	0.57 (0.10)	0.15	0.92	0.47	0.55 (0.08)	0.16	0.98
	Log Normal				Multivariate Normal			
β_1	0.67	0.71 (1.00)	0.72	0.87	0.67	0.61 (0.73)	0.75	0.94
β_2	0.73	0.80 (0.14)	0.06	0.52	0.73	0.73 (0.04)	0.06	0.99
β_3	-0.08	-0.11 (0.19)	0.07	0.51	-0.08	-0.07 (0.06)	0.07	0.98
γ_0	0.30	0.48 (0.66)	0.83	0.99	0.30	0.55 (0.74)	0.91	0.99
γ_1	0.65	0.61 (0.16)	0.18	0.93	0.65	0.57 (0.10)	0.21	1
ρ_{ST}	0.70	0.64 (0.13)	0.15	0.96	0.70	0.59 (0.09)	0.18	1
$\Phi_{10}(0)$	0.52	0.53 (0.04)	0.05	0.99	0.52	0.53 (0.05)	0.05	0.98

Table D.3: Simulation results: principal surrogacy assessment when multivariate normality does not hold

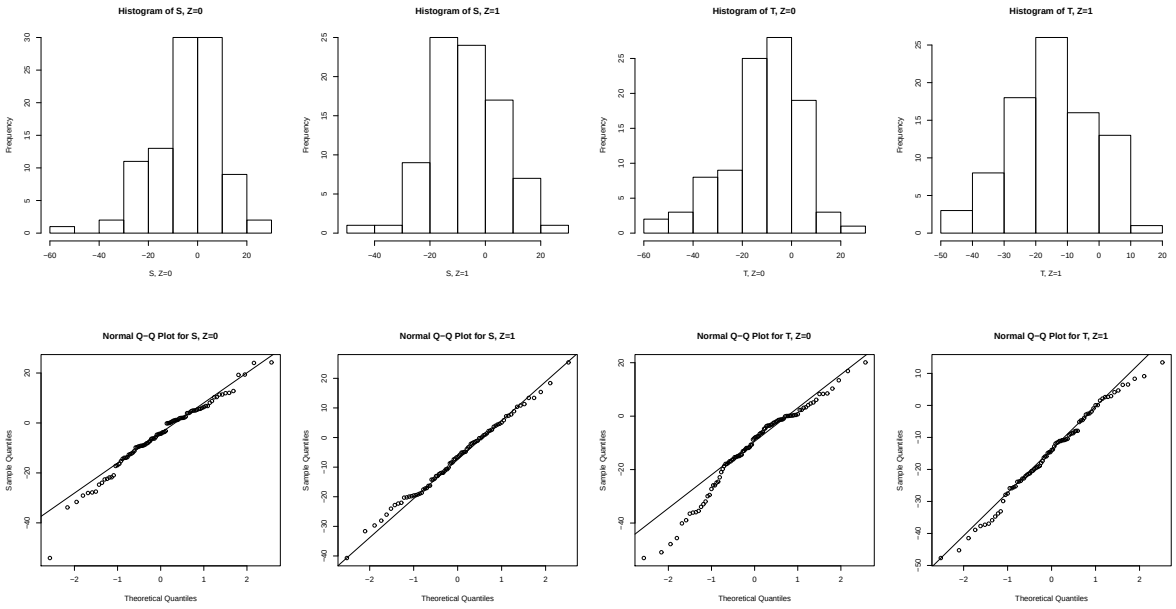
Model	Normal Marginals	MVNorm	Multivariate t_3	MVNorm	Log Normal	MVNorm
Truth						
PS	Invalid Surrogate		Valid Surrogate		Moderate Surrogate	
Prentice	Invalid Surrogate		Valid Surrogate		Invalid Surrogate	
Estimation Results						
$\gamma_0 = 0$ Not Rejected, Reject $\gamma_1 = 0$	0	0	0.82	0.91	0.92	0.84
$\gamma_0 = 0$ Not Rejected	0.01	0.01	0.96	0.995	0.98	0.93
Reject $\gamma_1 = 0$	0.02	0	0.84	0.91	0.92	0.90
Reject $\rho_{ST} = 0$	0.02	0	0.84	0.91	0.92	0.90
$\Phi_{10}(0) = 0.5$ Not Rejected	0.01	0.01	0.96	0.995	0.98	0.93
Prentice Criteria Not Rejected	0	0	0.64	0.95	0.43	0.80

APPENDIX E

Assessment of normality in age-related macular degeneration data

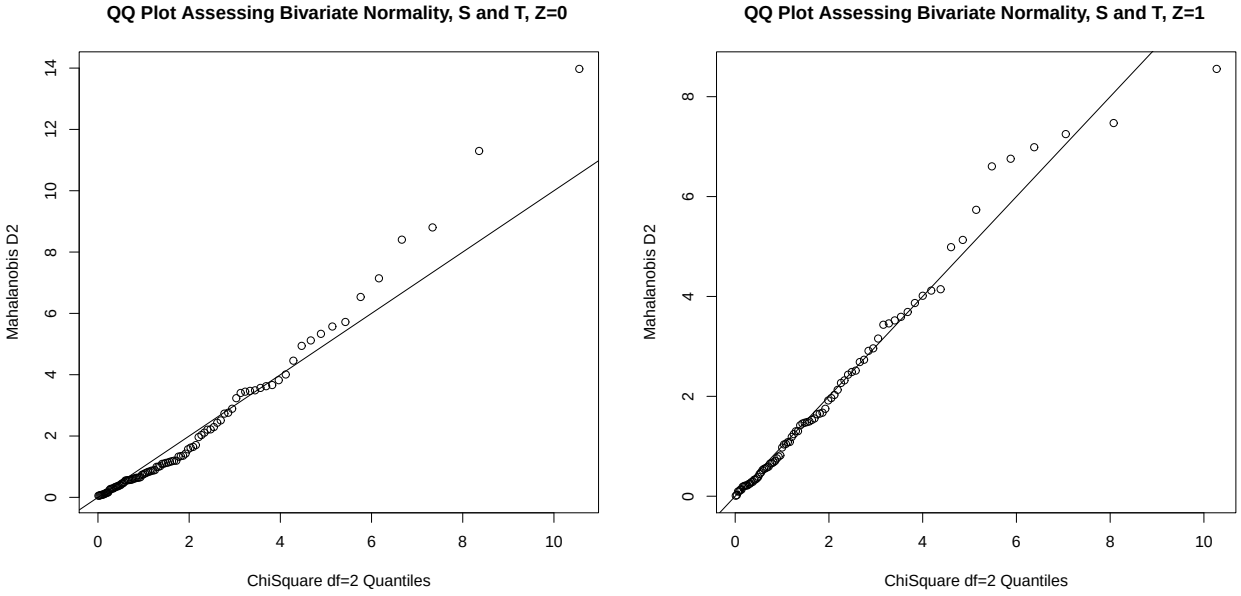
Histograms and normal QQ plots for observed age-related macular degeneration data. S is change in visual acuity at 6 months and T is change in visual acuity at 1 year, both with BLUP estimates subtracted off to account for random center effects.

Figure E.1: Histograms and normal QQ plots for age related macular degeneration data



QQ plot to assess bivariate normality. The plots were obtained by plotting the ordered Mahalanobis d^2 measures ($d_j = (X_j - \bar{X})'S^{-1}(X_j - \bar{X})$), $j = 1, ..., n$, $S = (1/n)\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'$) against the chi-square distribution quantiles $Q_P\left(\frac{j}{n+1}\right)$, where P is the number of columns in X. A graph of $\{Q_P\left(\frac{j}{n+1}\right), d^2_{(j)}\}_{j=1}^{n}$ should be a straight line under normality (Holgersson, 2006).

Figure E.2: QQ plots to assess bivariate normality for age related macular degeneration data

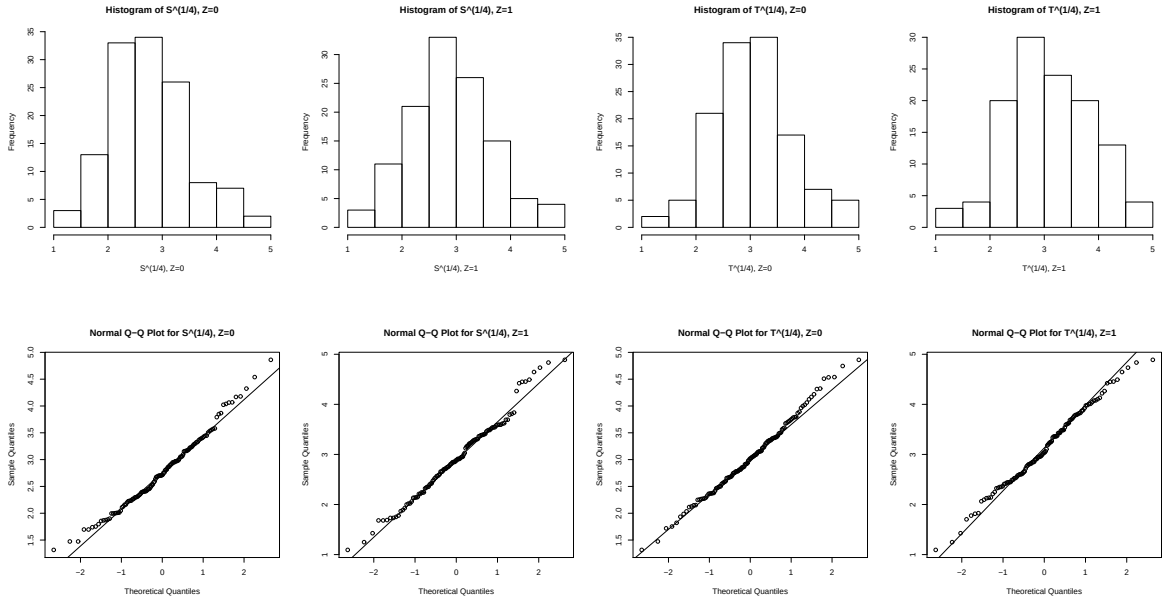


APPENDIX F

Assessment of normality in ovarian cancer data

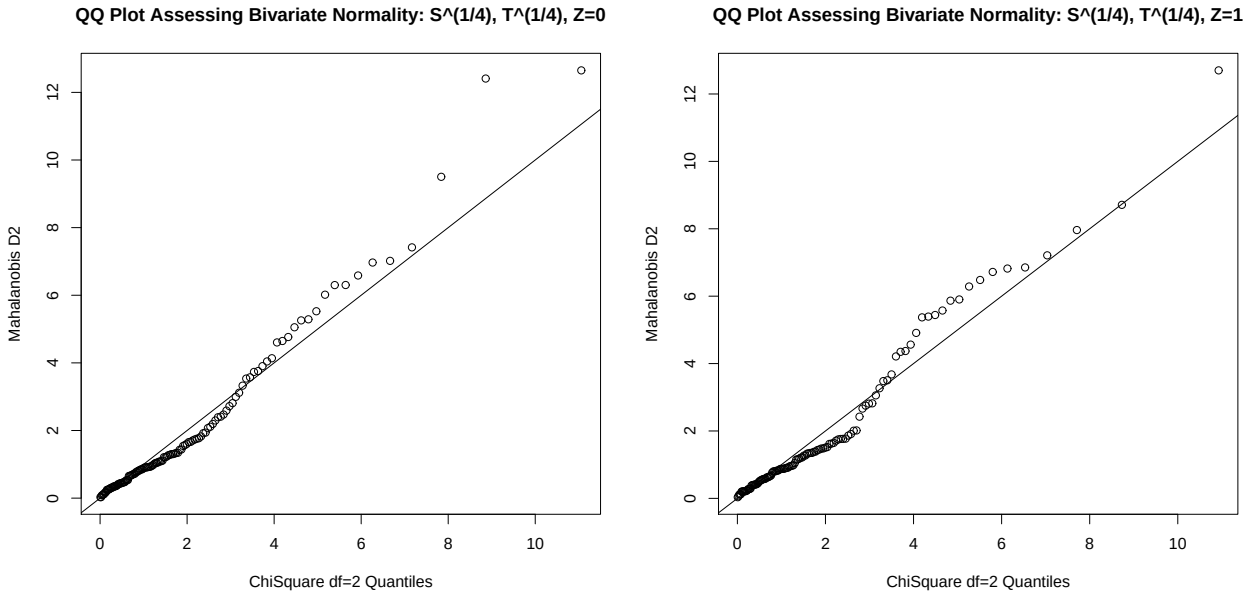
Histograms and normal QQ plots for observed ovarian cancer data. $S^{(1/4)}$ is the fourth root of progression free survival time, in months, and $T^{(1/4)}$ is the fourth root of overall survival time, in months.

Figure F.1: Histograms and normal QQ plots for ovarian cancer data



QQ plot to assess bivariate normality. The plots were obtained by plotting the ordered Mahalanobis d^2 measures ($d_j = (X_j - \bar{X})' S^{-1} (X_j - \bar{X})$), $j = 1, ..., n$, $S = (1/n) \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})'$) against the chi-square distribution quantiles $Q_P\left(\frac{j}{n+1}\right)$, where P is the number of columns in X. A graph of $\{Q_P\left(\frac{j}{n+1}\right), d^2_{(j)}\}_{j=1}^{n}$ should be a straight line under normality (Holgersson, 2006).

Figure F.2: QQ plots to assess bivariate normality for ovarian cancer data



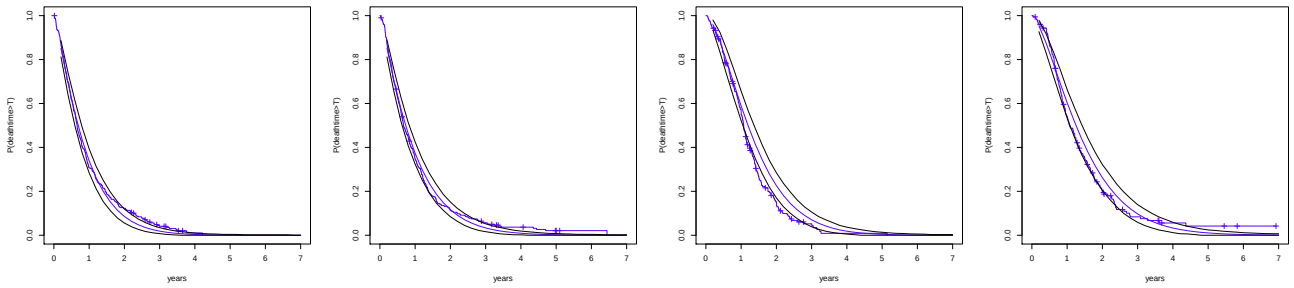
APPENDIX G

Posterior predictive plots for ovarian cancer data

For each of the 4 prior distributions, there are 8 plots, one for each combination of surrogate and treatment value. Each of the plots shows the Kaplan-Meier plot of the observed data at the give S and Z value, with the posterior predictive mean and 95% credible interval obtained from the Gaussian copula model overlayed on top. The plots show that the proposed model appears to provide an adequate fit to the data.

Figure G.1: Kaplan Meier plots for original data and posterior predictive distribution from Gaussian copula–No restriction on ρ's

(a) S=1, Z=0 (b) S=1, Z=1 (c) S=2, Z=0 (d) S=2, Z=1



(e) S=3, Z=0 (f) S=3, Z=1 (g) S=4, Z=0 (h) S=4, Z=1

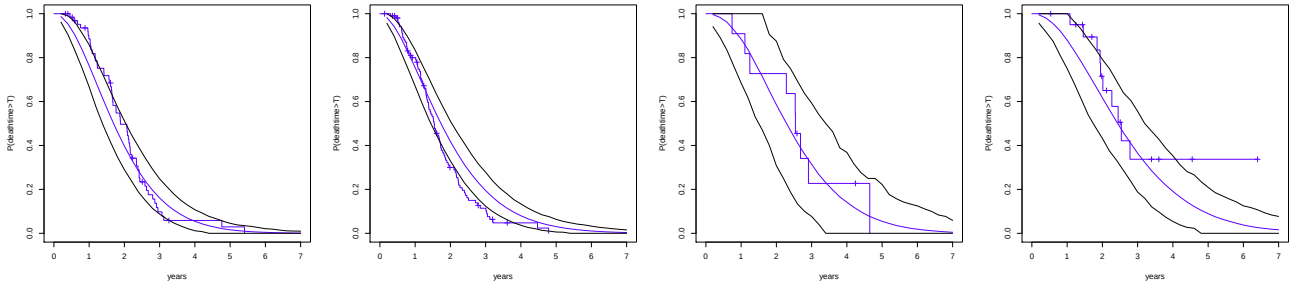
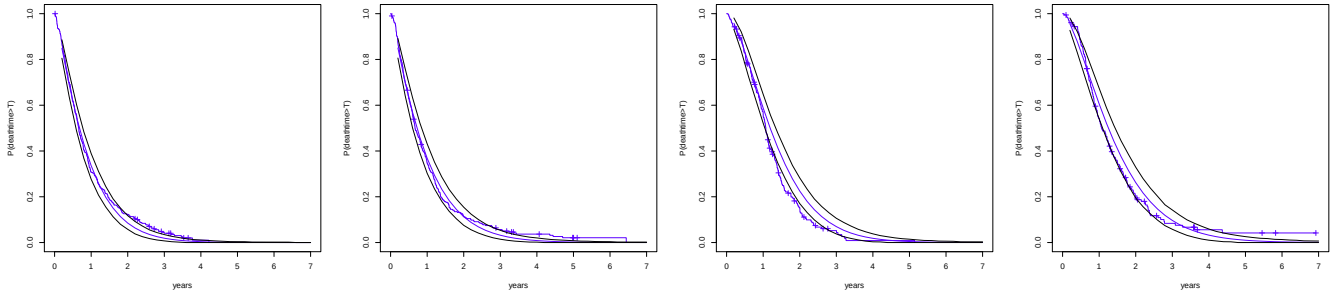


Figure G.2: Kaplan Meier plots for original data and posterior predictive distribution from Gaussian copula-$\rho \geq 0$

(a) S=1, Z=0 (b) S=1, Z=1 (c) S=2, Z=0 (d) S=2, Z=1



(e) S=3, Z=0 (f) S=3, Z=1 (g) S=4, Z=0 (h) S=4, Z=1

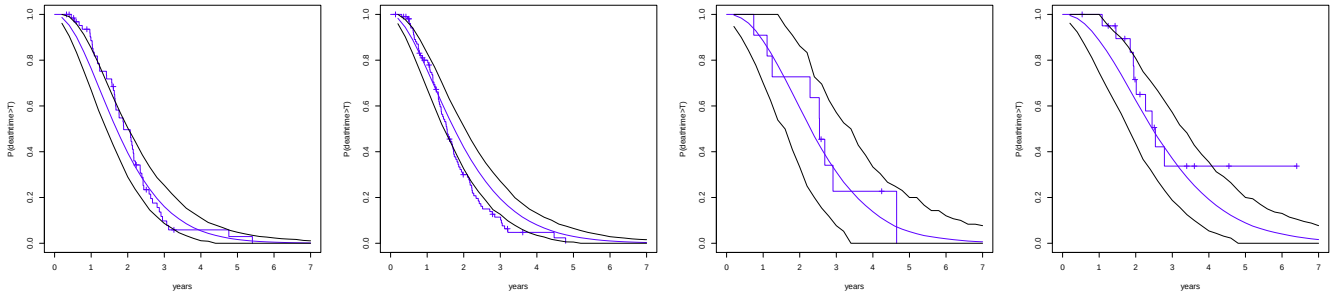
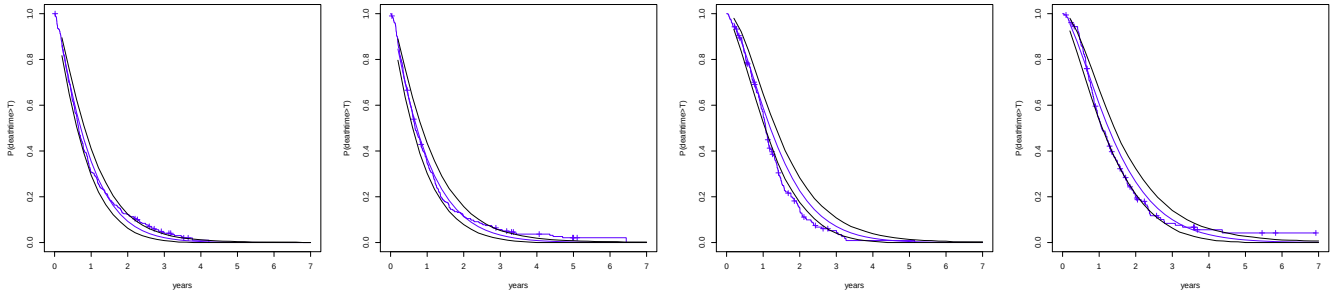


Figure G.3: Kaplan Meier plots from original data and posterior predictive distribution from Gaussian copula-$\rho \geq 0$ and $\rho_{10}, \rho_{01} < \rho_s, \rho_t, \rho_{00}, \rho_{11}$

(a) S=1, Z=0 (b) S=1, Z=1 (c) S=2, Z=0 (d) S=2, Z=1



(e) S=3, Z=0 (f) S=3, Z=1 (g) S=4, Z=0 (h) S=4, Z=1

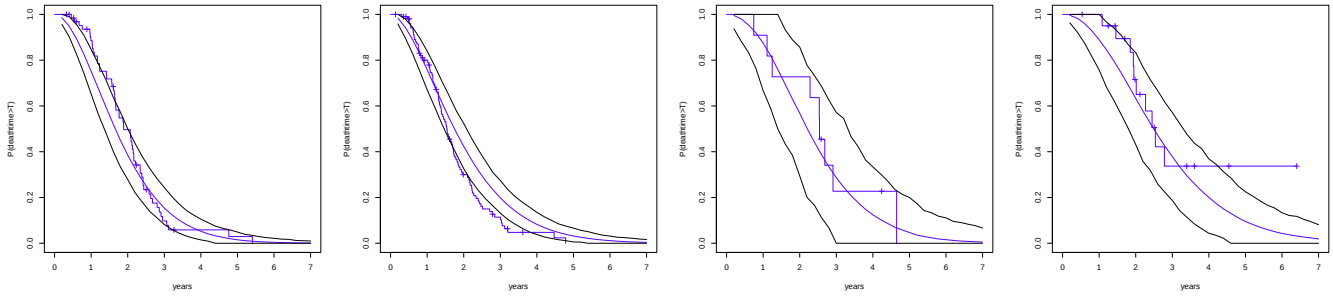
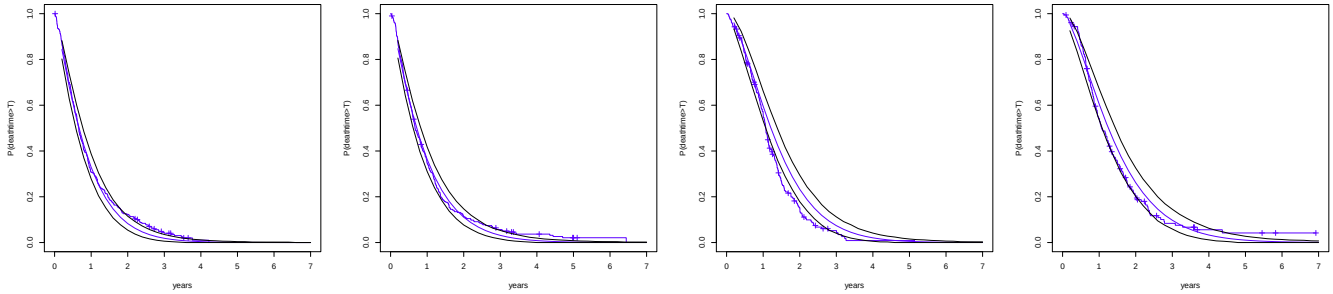
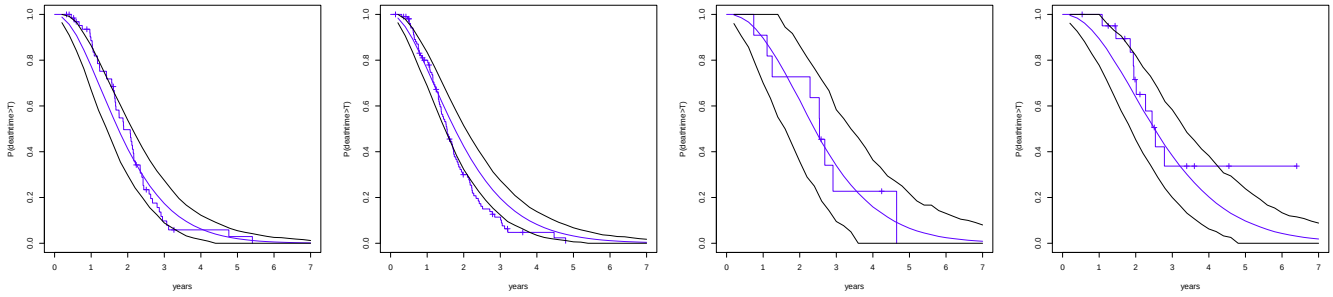


Figure G.4: Kaplan Meier plots, original data and posterior predictive distribution-Beta Priors

(a) S=1, Z=0 (b) S=1, Z=1 (c) S=2, Z=0 (d) S=2, Z=1



(e) S=3, Z=0 (f) S=3, Z=1 (g) S=4, Z=0 (h) S=4, Z=1

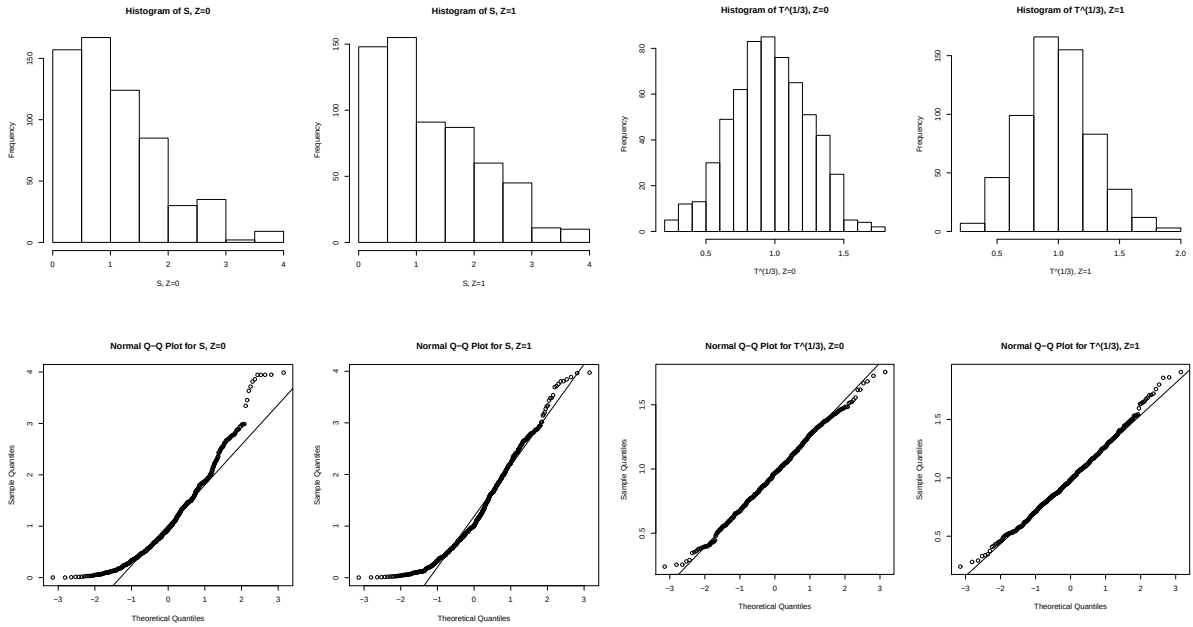


APPENDIX H

Histograms and normal QQ plots of transformed colorectal cancer data.

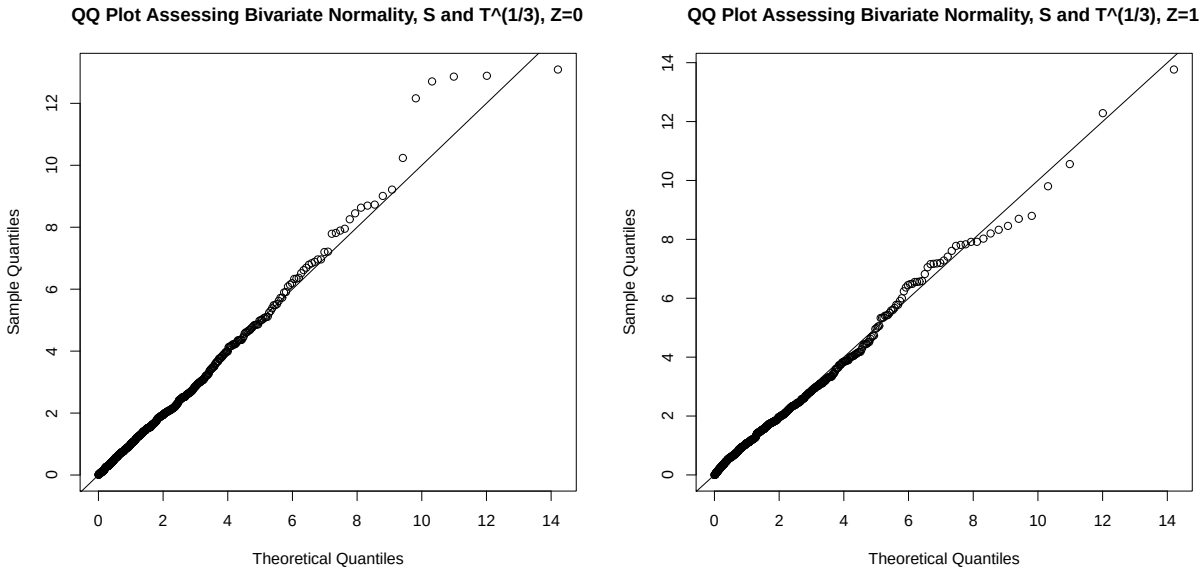
Histograms and QQ plots of the observed S and T after a transformation to approximately normalize them. For each observed $S = s$, a uniform random variable between $s - 1$ and s was draw, and for T, the third root was taken. Marginal normality appears to hold for $T^{1/3}$, while the transformed S is right skewed.

Figure H.1: Histograms and normal QQ plots of transformed colorectal cancer data



QQ plots to assess bivariate normality. The plots were obtained by plotting the ordered Mahalanobis d^2 measures ($d_j = (X_j - \bar{X})' S^{-1} (X_j - \bar{X})$), $j = 1, ..., n$, $S = (1/n) \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})'$) against the chi-square distribution quantiles $Q_P\left(\frac{j}{n+1}\right)$, where P is the number of columns in X. A graph of $\{Q_P\left(\frac{j}{n+1}\right), d^2_{(j)}\}_{j=1}^{n}$ should be a straight line under normality (Holgersson, 2006). In this case, bivariate normally appears to hold.

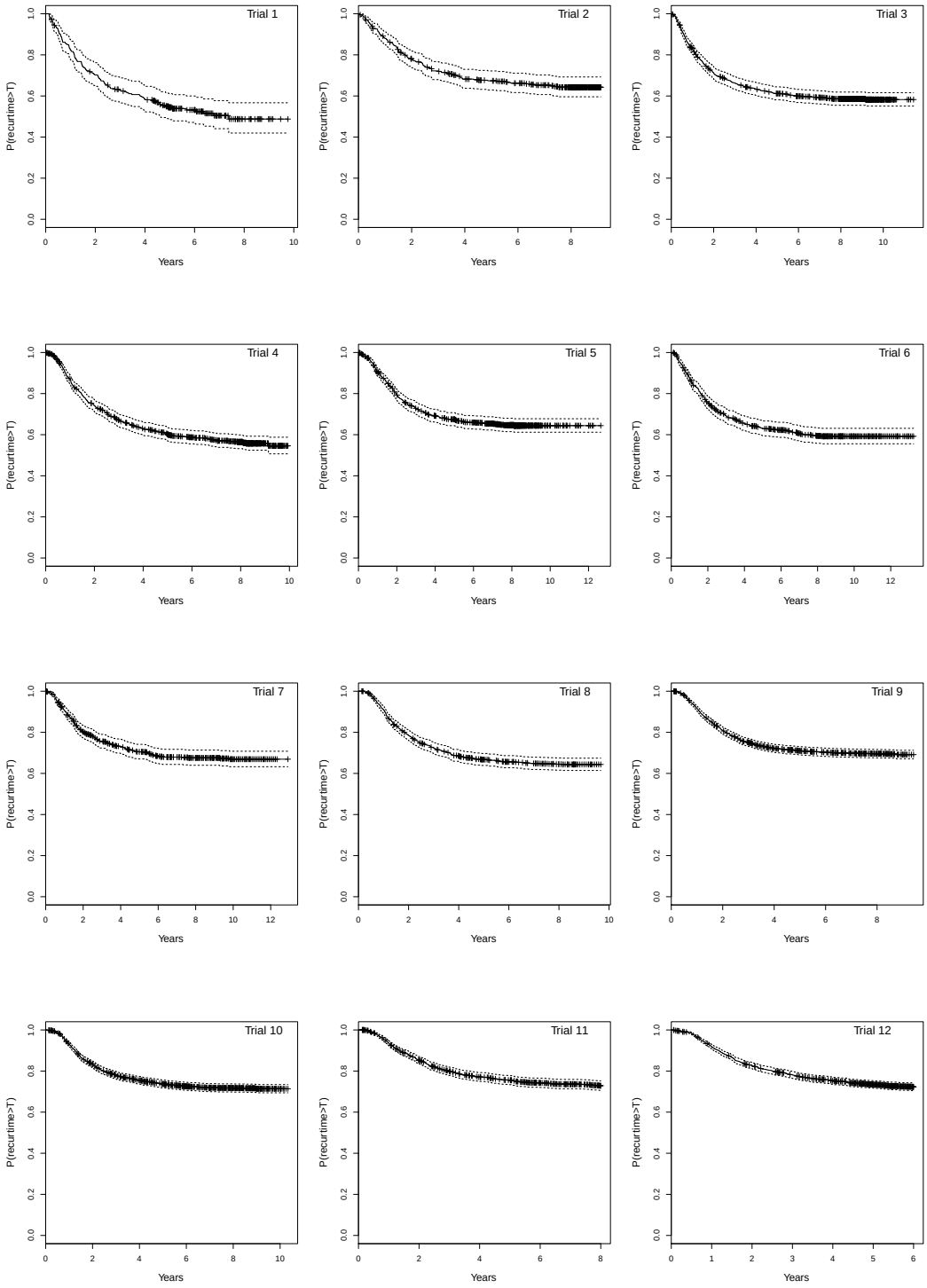
Figure H.2: QQ plots to assess bivariate normality for transformed colorectal cancer data



APPENDIX I

Kaplan-Meier plots of time to recurrence for 12 trials in conlon cancer

Figure I.1: Kaplan-Meier plots of time to recurrence for the 12 trials. Patients who died without recurrence are censored for recurrence at that time.



APPENDIX J

Details of multi-state cure model estimation procedure

Define the following indicator functions:

$RD_i = I(\delta_{ir} = 1, \delta_{id} = 1, Y_{ir} < Y_{id})$

$RDs_i = I(\delta_{ir} = 1, \delta_{id} = 1, Y_{ir} = Y_{id})$

$RA_i = I(\delta_{ir} = 1, \delta_{id} = 0, Y_{ir} < Y_{id})$

$RAs_i = I(\delta_{ir} = 1, \delta_{id} = 0, Y_{ir} = Y_{id})$

$RAs_i = I(\delta_{ir} = 1, \delta_{id} = 0, Y_{ir} = Y_{id})$

$NRD_i = I(\delta_{ir} = 0, \delta_{id} = 1, Y_{ir} < Y_{id})$

$NRA_i = I(\delta_{ir} = 0, \delta_{id} = 0, Y_{ir} < Y_{id})$

$NRDs_i = I(\delta_{ir} = 0, \delta_{id} = 1, Y_{ir} = Y_{id})$

$NRAs_i = I(\delta_{ir} = 0, \delta_{id} = 0, Y_{ir} = Y_{id})$

$Z_i = I(c_i = 1)$

The observed data likelihood is given by:

$\prod_{i=1}^{n} \{ \ [(1 - p_i)S_2(Y_{ir})\lambda_{23}(Y_{ir})S_3(Y_{id})\lambda_{34}(Y_{id} - Y_{ir})]^{RD_i(1-Z_i)}$

$[(1 - p_i) \int_0^{Y_{ir}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id} - u)^{\delta_{id}} S_3(Y_{id} \mid u)du]^{RDs_i(1-Z_i)}$

$[(1-p_i)S_2(Y_{ir})\lambda_{23}(Y_{ir})S_3(Y_{id})]^{RA_i(1-Z_i)}$

$[(1-p_i)S_2(Y_{ir})\lambda_{23}(Y_{ir})]^{RAs_i(1-Z_i)}$

$[p_i\lambda_{14}(Y_{id})S_1(Y_{id})]^{NRD_iZ_i}$

$[(1-p_i)\lambda_{24}(Y_{id})S_2(Y_{id}) + (1-p_i)\int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id}-u)S_3(Y_{id} \mid u)du]^{NRD_i(1-Z_i)}$

$[p_iS_1(Y_{id})\lambda_{14}(Y_{id})]^{NRDs_iZ_i}$

$[(1-p_i)S_2(Y_{id})\lambda_{24}(Y_{id})]^{NRDs_i(1-Z_i)}$

$[p_iS_1(Y_{id})]^{NRA_iZ_i}$

$[(1-p_i)S_2(Y_{id}) + (1-p_i)\int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u)S_2(u)S_3(Y_{id} \mid u)du]^{NRA_i(1-Z_i)}$

$[p_iS_1(Y_{id})]^{NRAs_iZ_i}$

$[(1-p_i)S_2(Y_{id})]^{NRAs_i(1-Z_i)}\}$

Where:

$p_i = \frac{exp(\gamma_0+\gamma_{trt}T_i+\gamma_{st}S_i+\gamma_{age}A_i)}{1+exp(\gamma_0+\gamma_{trt}T_i+\gamma_{st}S_i+\gamma_{age}A_i)}$

$S_1(t) = exp\left(-\left(\frac{t}{\lambda_{14}}\right)^{\rho_{14}} exp(\beta_{trt_{14}}T_i + \beta_{st_{14}}S_i + \beta_{age_{14}}A_i)\right)$

$S_2(t) = exp\left(-\left(\frac{t}{\lambda_{23}}\right)^{\rho_{23}} exp(\beta_{trt_{23}}T_i + \beta_{st_{23}}S_i + \beta_{age_{23}}A_i) - \left(\frac{t}{\lambda_{24}}\right)^{\rho_{24}} exp(\beta_{trt_{24}}T_i + \beta_{st_{24}}S_i + \beta_{age_{24}}A_i\right.$

$S_3(t) = exp\left(-\left(\frac{t-Y_{ir}}{\lambda_{34}}\right)^{\rho_{34}} exp(\beta_{trt_{34}}T_i + \beta_{st_{34}}S_i + \beta_{age_{34}}A_i + \beta_{Tr_{34}}Y_{ir})\right)$

$\lambda_{23}(t) = \left(\frac{\rho_{23}}{\lambda_{23}}\right)\left(\frac{t}{\lambda_{23}}\right)^{\rho_{23}-1} exp(\beta_{trt_{23}}T_i + \beta_{st_{23}}S_i + \beta_{age_{23}}A_i)$

$\lambda_{24}(t) = \left(\frac{\rho_{24}}{\lambda_{24}}\right)\left(\frac{t}{\lambda_{24}}\right)^{\rho_{24}-1} exp(\beta_{trt_{24}}T_i + \beta_{st_{24}}S_i + \beta_{age_{24}}A_i)$

$\lambda_{34}(t-Y_{ir}) = \left(\frac{\rho_{34}}{\lambda_{34}}\right)\left(\frac{(t-Y_{ir})}{\lambda}\right)^{\rho_{34}-1} exp(\beta_{trt_{34}}T_i + \beta_{st_{34}}S_i + \beta_{age_{34}}A_i + \beta_{Tr_{34}}Y_{ir})$

The integrals:

$\int_0^{Y_{ir}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id}-u)^{\delta_{id}}S_3(Y_{id} \mid u)du$

$\int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u)S_2(u)\lambda_{34}(Y_{id}-u)^{\delta_{id}}S_3(Y_{id} \mid u)du$

were computed by adaptive quadrature using the 'integrate' function in R.

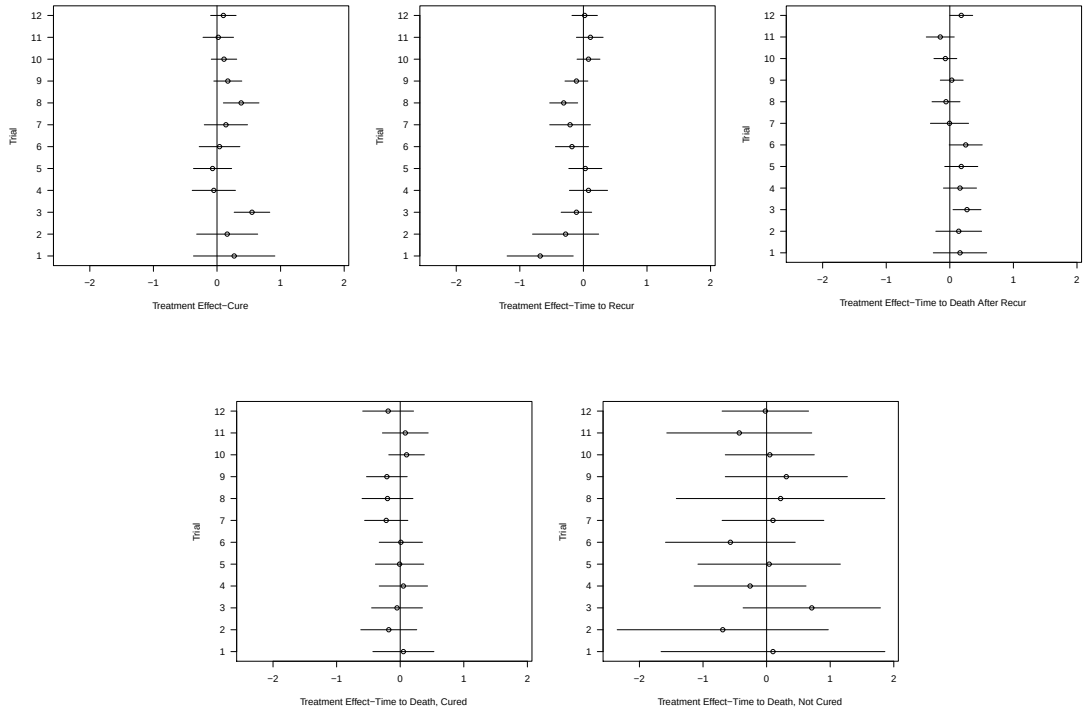
The Metropolis-Hastings algorithm is used to draw parameters. The chain is run for 50,000 iterations, after a 10,000 iteration burn-in period with 5,000 draws from the

posterior distribution saved for each parameter by taking every 100th draw from the post burn-in iterations. All of the proposal distributions are normal and centered at the most recent parameter draw. For the shape parameters, the proposal distribution is truncated at 0. For each study, the variance of the proposal distribution for each parameter is adjusted so that each of the resulting acceptance rates are close to 40%.

APPENDIX K

Estimated treatment effects from multi-state cure model

Figure K.1: Treatment effect estimates for each of the five model components for 12 trials. Each line represents the 95% credible interval for the coefficient



APPENDIX L

Multi-state cure model Cox-Snell residual plots

Figure L.1: Cox-Snell residual plots for time to recur. Results from 12 trials.

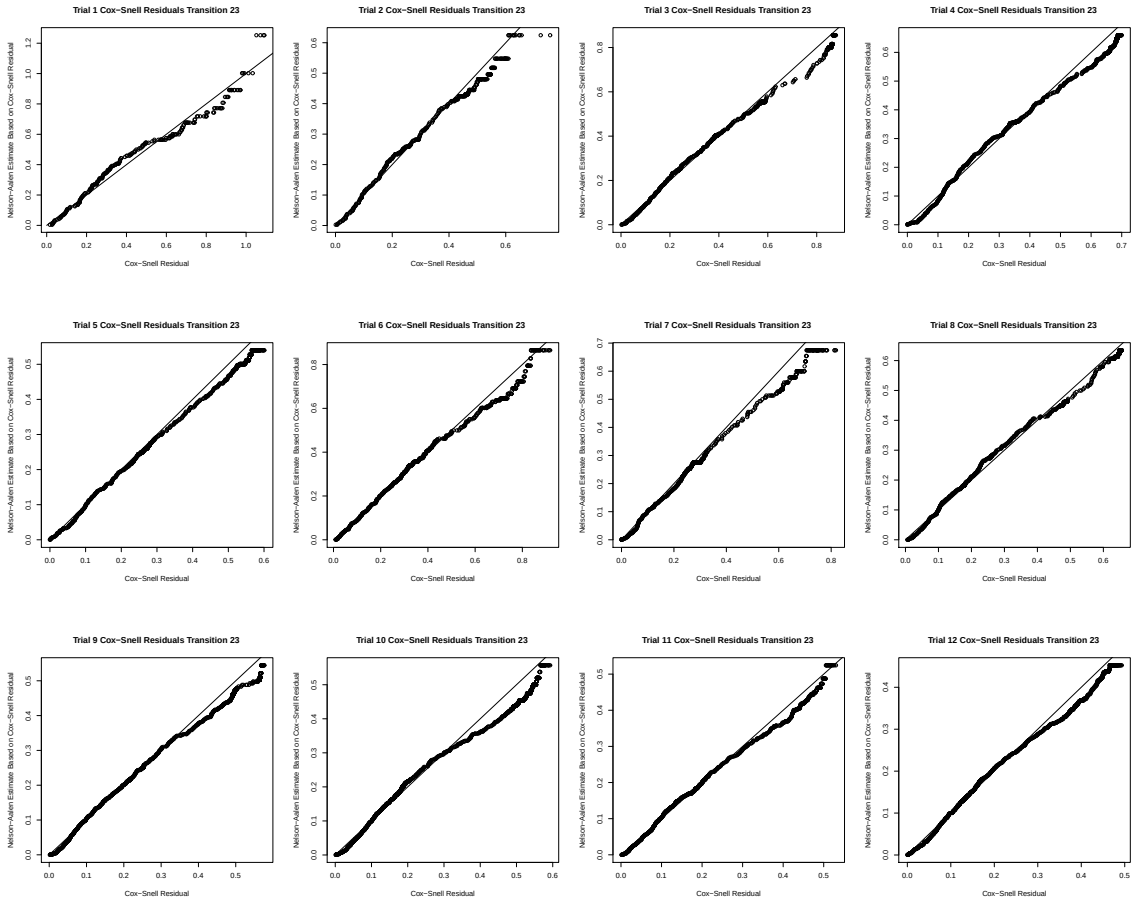
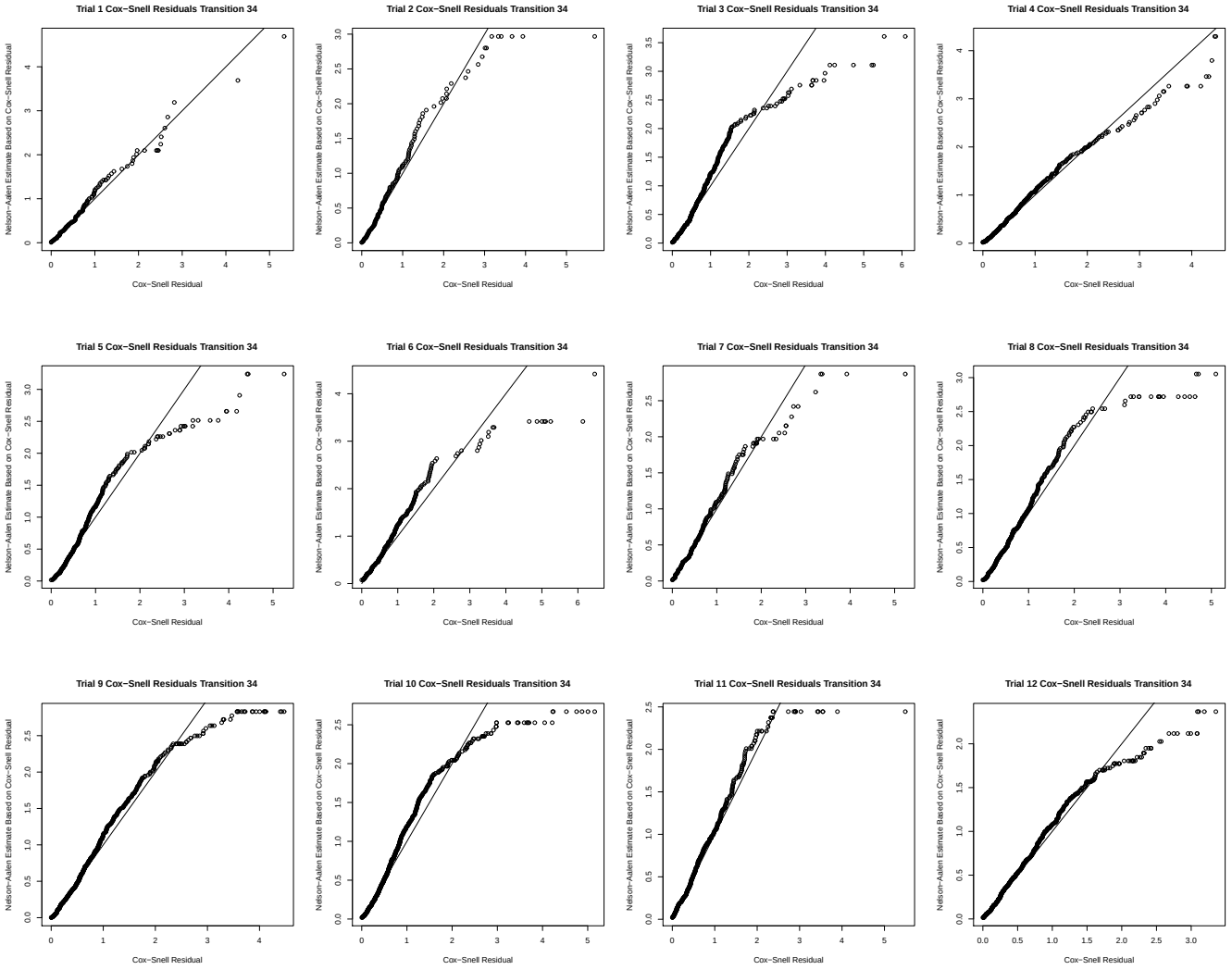


Figure L.2: Cox-Snell residual plots for time to death after recurrence. Results for 12 trials.



APPENDIX M

Multi-state cure model deviance residual plots

Figure M.1: Deviance residual plots for time to recurrence plotted against age.

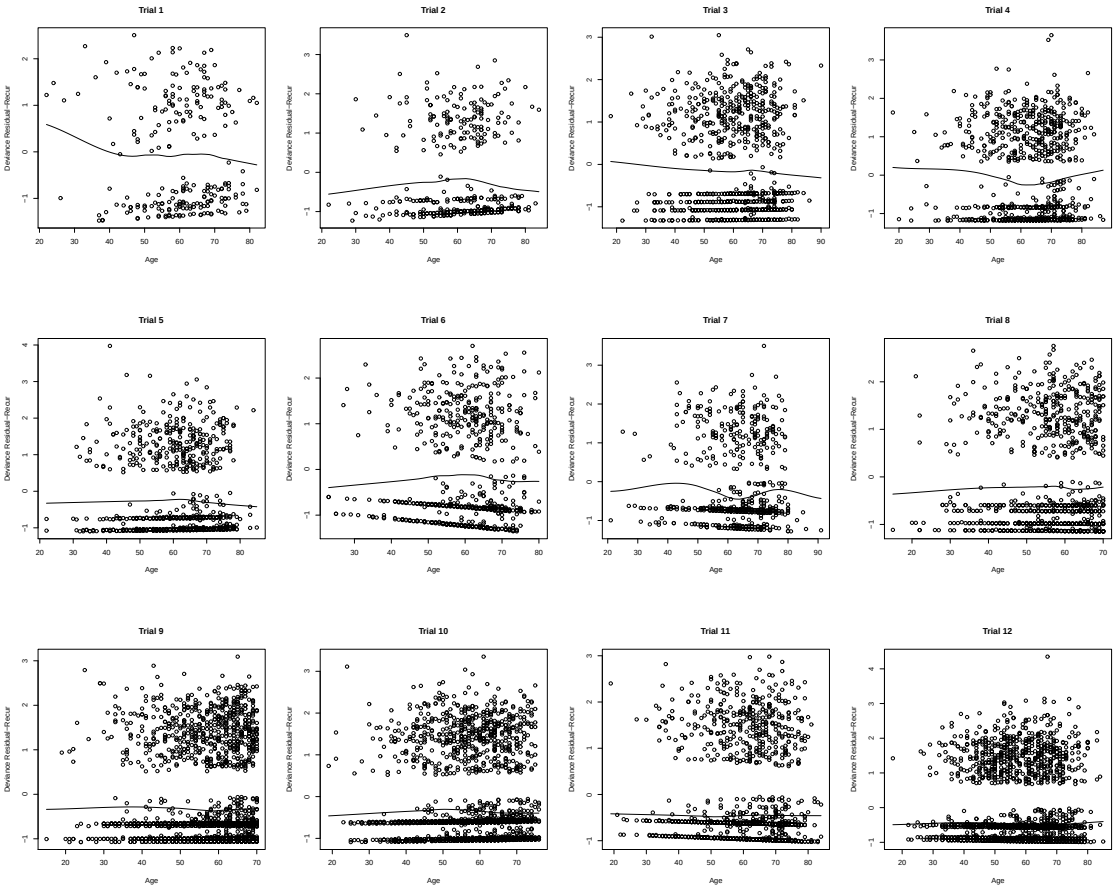


Figure M.2: Deviance residual plots for time to death after recurrence plotted against age.

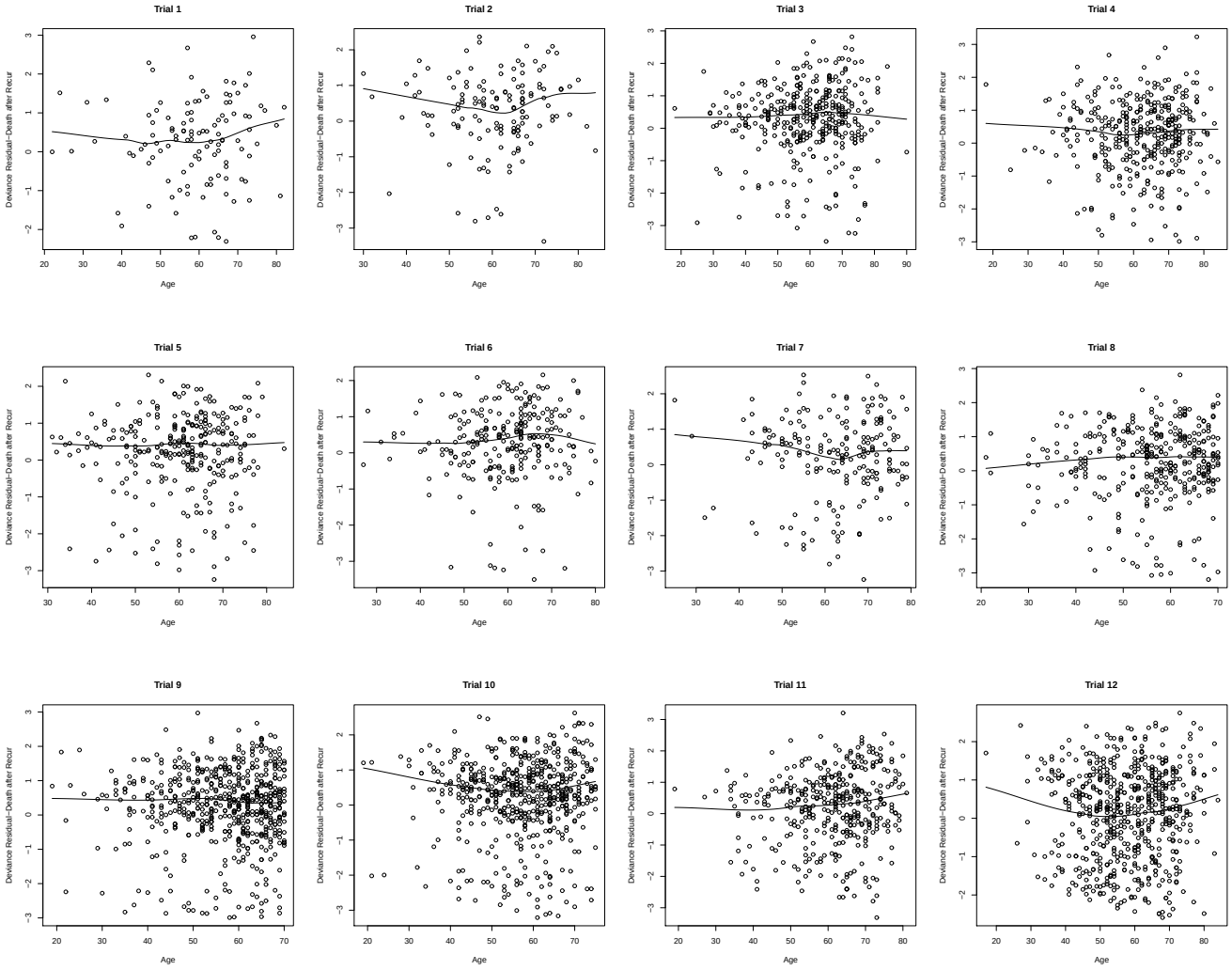


Figure M.3: Deviance residual plots for time to death after recurrence plotted against time to recurrence.

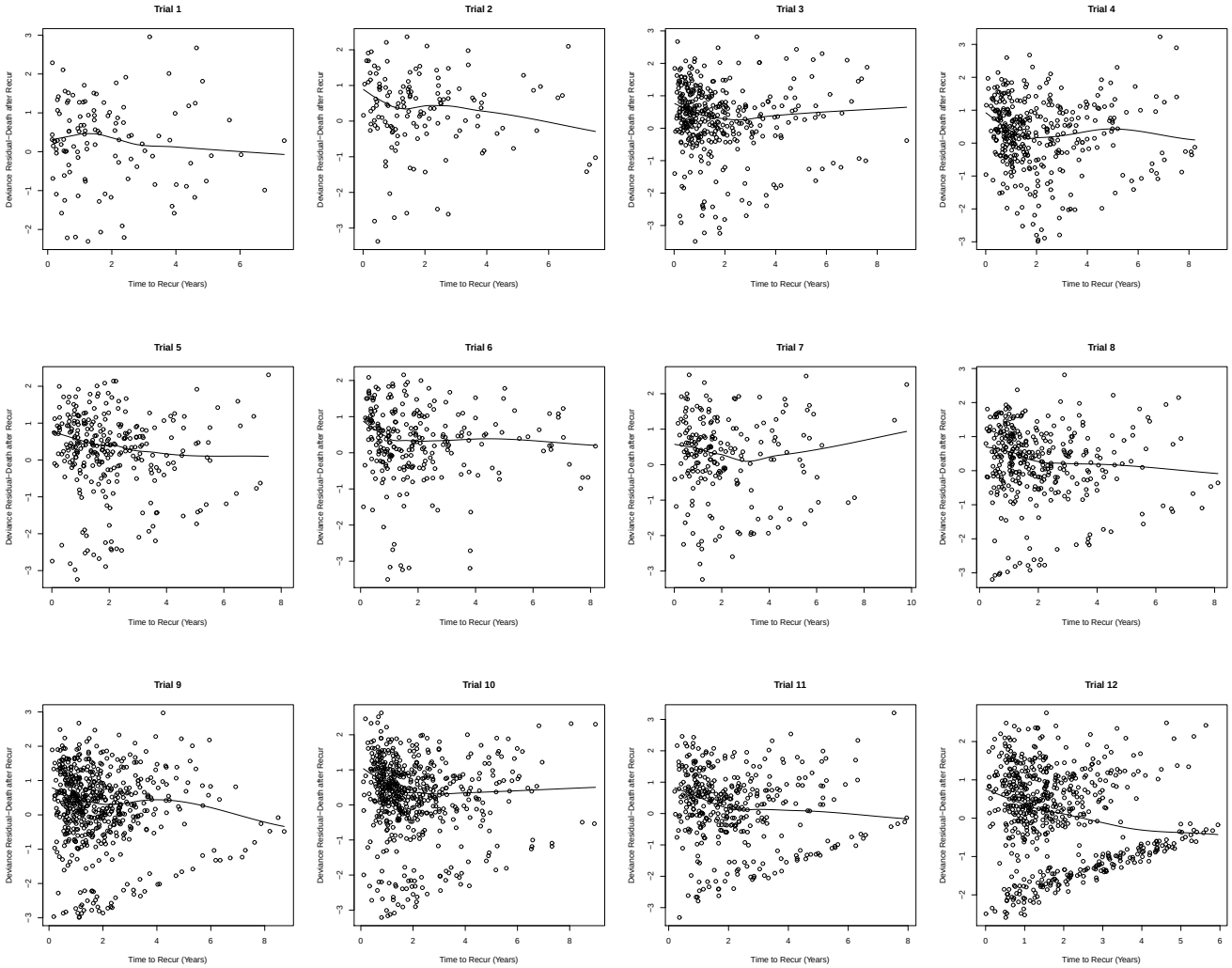
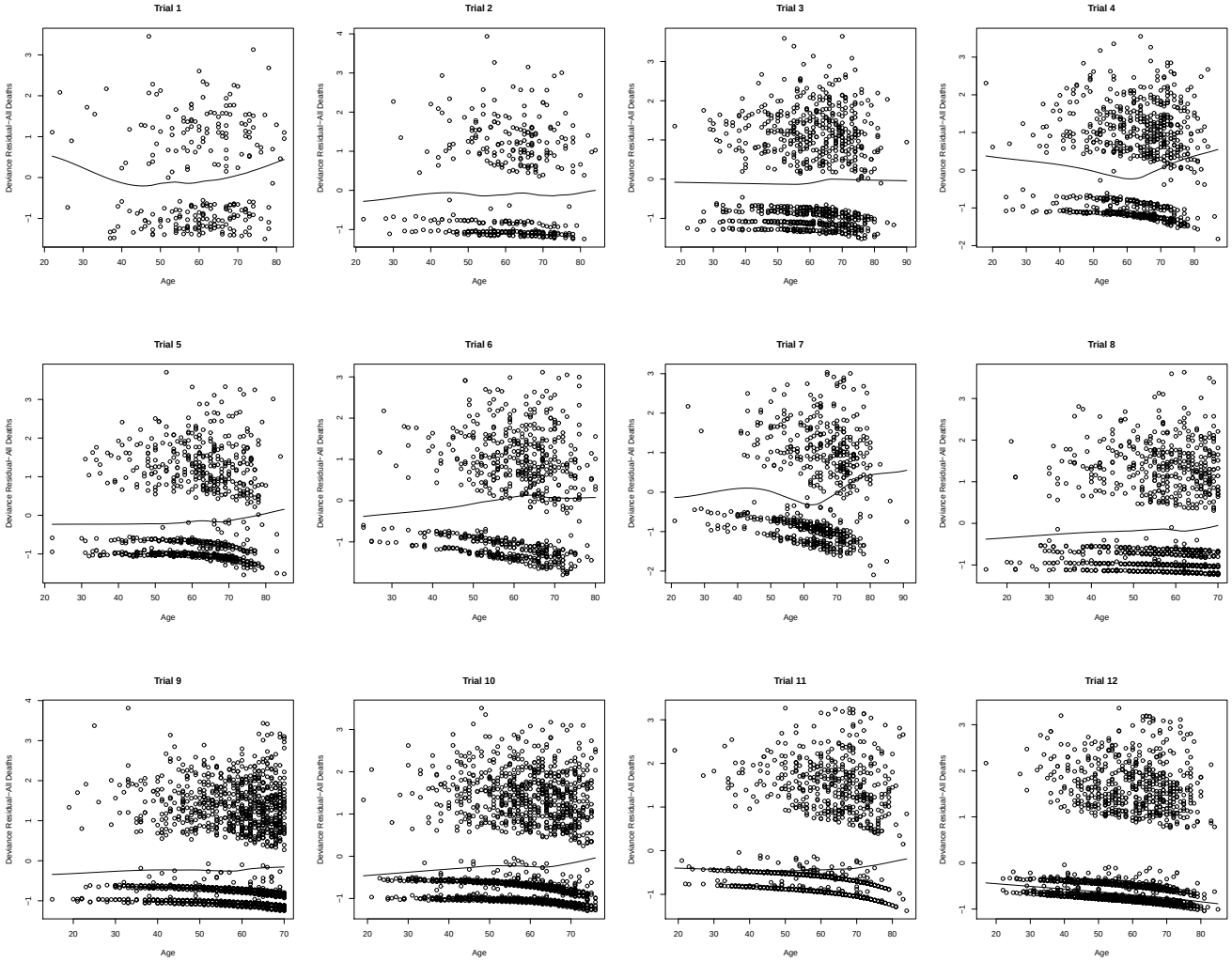


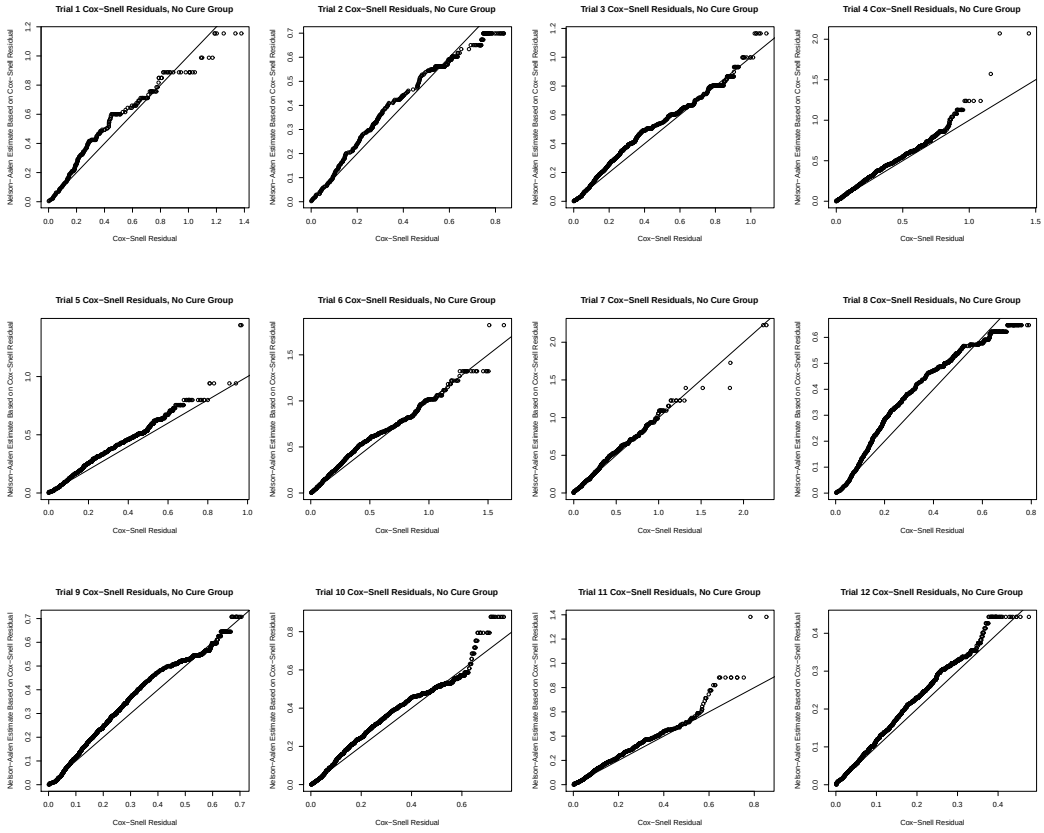
Figure M.4: Deviance residual plots for time to death plotted against age.



APPENDIX N

Cox-Snell residual plots for multi-state model with no cured fraction

Figure N.1: Cox-Snell residual plots for time to death for model without a cured fraction. Results from 12 trials.



APPENDIX O

Follow-up details for colon cancer trials

Table O.1: Accrual and follow-up details for 12 trials in colon cancer

Study	Accrual Length (Years)	Longest Follow-Up: Original (Years)	Longest Follow-Up: Artificially Censored (Years)
1	5.7	9.9	7.6
2	1.5	9.1	6
3	3.6	11.4	5.6
4	2	9.9	5.5
5	4.8	12.6	6.8
6	5.2	13.2	7.2
7	4.9	12.9	6.9
8	1.7	9.7	5.7
9	1.4	9.4	5.9
10	2.3	10.3	5.8
11	2	8	5.5
12	2.8	6	5.8

BIBLIOGRAPHY

Alonso, A., and Molenberghs, G. (2008). Evaluating time to event cancer recurrence as a surrogate marker for survival from an information theory perspective. *Statistics in Medicine* **17**, 497–504.

Anderson, P.K., Esbjerg, S., and Sorenson, T.I.A. (2000). Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine* **19**, 587–599.

Anderson, P.K. and Keiding, N. (2002). Milti-state models for event history analysis. *Statistical Methods in Medical Research* **11,** 91–115.

Barnard, J., McCulloch, R. and Meng, X-L (2000). Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Statistica Sinica* **10,** 1281–1311.

Bartolucci, F., and Grilli, L. (2011). Modeling partial compliance through copulas in a principal stratification framework. *Journal of the American Statistical Association* **106,** 469–479.

Berkson, J. and Gage, R.P. (1952). Survival Curve for Cancer Patients Following Treatment. *Journal of the American Statistical Association***47**, 501–515.

Broglio, K.R., and Berry, D.A. (2009). Detecting an overall survival benefit that is derived from progression-free survival. *Journal of the National Cancer Institute* **101**(23), 1642–1649.

Burzykowski, T., Molenberghs, G., and Buyse, M. (2004). The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. *Journal of the Royal Statistical Association, Series A* **167**, 103–124.

Buyse, M., and Molenberghs, G. (1998). Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* **54**, 1014–1029.

Buyse, M., Molenberghs, G., Burzykowski, D., et al. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* **1,** 49–67.

Buyse, M., and Piedbois, P. (1996). On the relationship between response to treatment and survival. *Statistics in Medicine* **15**, 2797–2812.

Chen, H., Geng, Z., and Jia, J. (2007). Criteria for surrogate end points. *Journal of the Royal Statistical Society, Series B* **69**, 919–932.

Chen, T.T., Simon, R.M., Kom, E.L., Anderson, S.J., Lindblad, A.S., Wieand, H.S., Douglass, H.O., Fisher, B., Hamilton, J.M., and Friedman, M.A. (1998). Investigation of disease-free survival as a surrogate endpoint for survival in cancer clinical trials. *Communications in Statistics-Theory and Methods* **27(6)**, 1363–1378.

Conlon, A.S.C., Taylor, J.M.G., and Elliott, M.R. (2013). Surrogacy assessment using principal stratification when surrogate and outcome measures are multivariate normal. University of Michigan Department of Biostatistics Working Paper #99. Available at http://www.bepress.com/umichbiostat/

Conlon, A.S.C., Taylor, J.M.G., Sargent, D.J., Yothers, G. (2011). Using cure models and multiple imputation to utilize recurrence as an auxiliary variable for overall survival. *Clinical Trials* **8**, 581-590.

Cook, R.J., and Lawless, J.F. (2001). Some comments on efficiency gains from auxiliary information for right-censored data. *Journal of Statistical Planning and Inference* **96**, 191–202.

Dabrowska, D.M., Sun, G.W., and Horowitz, M.M. (1994). Cox regression in a Markov renewal model: An application to the analysis of bone-marrow transplant data. *Journal of the American Statistical Association* **89**, 867–877

de Leon, A.R., Wu, B. (2011). Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine* **30**, 175–185.

Ellenberg, S.S., and Hamilton, J.M. (1989). Surrogate endpoints in clinical trials: cancer. *Statistics in Medicine* **8**, 405–413.

Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38,** 1041–1046.

Farewell, V.T. (1986). Mixture models in survival analysis: are the worth the risk?. *Canadian Journal of Statistics* **14,** 257–262.

Faucett C.L., Schenker N., and Taylor J.M.G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* **58**, 37–47.

Finkelstein D.M., and Schoenfeld D.A. (1994). Analysing survival in the presence of an auxiliary variable. *Statistics in Medicine* **13**, 1747–1754.

Fleming, T.R. and Prentice, R.L. and Pepe, M.S. and Glidden, D. (1994). Surrogate and auxiliary endpoints in clinical trials with potential applications in cancer and AIDS research. *Statistics in Medicine* **13,** 955–968.

Foucher, Y. and Mathieu, E. and Saint-Pierre, P. and Durand, J.F. and Daurés, J.P. (2005). A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal* **47(6),** 1–9.

Frangakis, C.E., and Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics* **58,** 21–29.

Freedman, L., Graubard, B., and Schatzkin, A. (1992). Statistical evaluation of intermediate endpoints for chronic diseases. *Statistics in Medicine* **11,** 167–178.

Gilbert, P.B., and Hudgens, M.G. (2008). Evaluating candidate principal surrogate endpoints. *Biometrics* **64,** 1146–1154.

Gilbert, P., Qin, L., and Self, S. (2009). Response to Andrew Dunning's comment on "Evaluating a surrogate endpoint at three levels, with application to vaccine development". *Statistics in Medicine* **28,** 716–719.

Gustafson, P. (2010). Bayesian inference for partially identified models. *The International Journal of Biostatistics* **Vol. 6: Iss. 2,** Article 17.

Holgersson, H.E.T. (2006). A graphical method for assessing multivariate normality. *Computational Statistics* **21,** 141-149.

Hsu C.-H., Taylor J.M.G., Murray S., and Commenges D. (2006). Survival analysis using auxiliary variables via nonparametric multiple imputation. *Statistics in Medicine* **25**, 3503–3517.

Joffe, M.M., and Greene, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics* **65,** 530–538.

Klein, J.P., Keiding, N., and Copelan, E.A. (1994). Plotting summary predictions in multi-state survival models- probabilities of relapse and death in remission for bone-marrow transplantation patients. *Statistics in Medicine 12*, 2315–2332.

Kneib, T., and Hennerfeind, A. (2008). Bayesian semi parametric multi-state models. *Statistical Modelling* **8**, 169–198.

Kosorok, M. and Fleming, T. (1993). Using surrogate failure time data to increase cost effectiveness in clinical trials. *Biometrika* **80,** 823–833.

Lagakos, S.W. (1977). Using auxiliary variables for improved estimates of survival time. *Biometrics* **33**, 399–404.

Lagakos, S.W., Sommer, C.J., and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrics* **65**, 311–317.

Li K.-H., Meng, X.-L., Raghunathen, T.E., and Rubin, D.B. (1991). Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica* **1**, 65–92.

Li, Y., Taylor, J.M.G., and Elliott, M.R. (2010). A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics* **66,** 523–531.

Li, Y., Taylor, J.M.G., and Little, R.J.A. (2011). A shrinkage approach for estimating a treatment effect using intermediate biomarker data in clinical trials. *Biometrics* **67**, 1434–1441.

Meira-Machado, L. and de Uña-*Á*lvarez, J. and Cadarso-Su*á*rez, C. and Anderson, P.K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research* **18,** 195–222.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* **9,** 538–557.

Meta-Analysis Group in Cancer (1998). Efficacy of intravenous continuous infusion of 5-fluorouracil compared with bolus administration in patients with advanced colorectal cancer. *Journal of Clinical Oncology* **16**, 301–308.

Molenberghs, G., Geys, H., and Buyse, M. (2001). Evaluation of surrogate end-points in randomized experiments with mixed discrete and continuous outcomes. *Statistics in Medicine* **20**, 3023–3038.

Nelsen, R.B. (2006). An Introduction to Copulas, 2^{nd} Edition. *New York: Springer-Verlag* .

Olsson, U., Drasgow, F., Dorans, N.J. (1982). The polyserial correlation coefficient. *Psychometrika* **47**, 337–347.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* **82(4),** 669–710.

Peng, Y. and Dear, K.B.G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56**, 237–243.

Prentice, R.L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* **8,** 431–440.

Putter, H. and Fiocco, M. and Geskus, R.B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* **26,** 2389–2430.

Putter, H., van der Hage, J., de Bock, G.H., Elgalta, R., and van de Velde, C.J.H. (2006). Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal* **48**, 366-380.

Qin, L., Gilbert, P.B., Follmann, D., and Li, D. (2008). Assessing surrogate endpoints in vaccine trials with case-cohort sampling and the Cox model. *The Annals of Applied Statistics* **Vol. 2, No. 1,** 386–407.

Ritter, C., and Tanner, M.A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association* **87**, 861–868.

Robins, J.M., and Wang, N. (2000). Inference for Imputation Estimators. *Biometrika* **87**, 113–124.

Rosenbaum, P. (1984). The consequences of adjustment for a concomitant variable that has been affected by treatment. *Journal of the Royal Statistical Society, Ser. B* **147,** 656–666.

Rosenbaum, P. (1984). From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment. *Journal of the American Statistical Association* **79,** 41–48.

Rubin, D.B. (1978). Bayesian-inference for causal effects-role of randomization. *The Annals of Statistics* **6,** 34–58.

Rubin, D.B. (1978). Multiple imputation in sample surveys-a phenomenological Bayesian approach to nonresponse. *Journal of the American Statstical Association* 20–34.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. *New York: Wiley*

Rubin, D.B. (1996). Multiple Imputation after 18+ years. *Journal of the American Statistical Association* **91,** 473–489.

Sargent, D.J., Wieand, H.S., Haller, D.G., et al. (2005). Disease-free versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *Journal of Clinical Oncology* **23(34),** 8664–8670.

Sargent, D.J., Patiyil, S., Yothers, G., et al. (2007). End points for colon cancer adjuvant trials: Observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT group. *Journal of Clinical Oncology* **25,** 4569–4574.

Satten, G., and Sternberg, M. (1999). Fitting semi-markov models to interval-censored data with unknown initiation times. *Biometrics* **55**, 507–513.

Schwartz, S.L., Li, F., and Mealli, F. (2011). A Bayesian semiparametric approach to intermediate variables in causal inference. *Journal of the American Statistical Association* **106,** 1331–1344.

Sklar, A. (1959). Fontions de répartition *à n* dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* **8,** 229–231.

Song, P.X. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics* **27,** 305–320.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B* **64**(4), 583–639.

Sy, J.P., and Taylor, J.M.G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56**, 227–236.

Taylor, J.M.G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics* **51**, 899–907.

Torri, V., Simon, R., Russek-Cohen, E., Midthune, D., and Friedman, M. (1992). Statistical model to determine the relationship of response and survival in patients with advanced ovarian cancer treated with chemotherapy. *Journal of the National Cancer Institute* **84**, 407–414.

VanderWeele, T.J. (2011). Principal Stratification - Uses and Limitations. *The International Journal of Biostatistics* **Volume 7, Issue 1,** Article 28.

VanderWeele, T.J. (2013). Surrogate measures and consistent surrogates. *Biometrics* **69**, 561–581.

Wang, Y., Mogg, R., and Lunceford, J. (2012). Evaluating correlation-based metric for surrogate marker qualification within a causal correlation framework. *Biometrics* **68,** 617–627.

Wang, Y., and Taylor, J.M.G. (2003). A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics* **58**, 803–812.

Weir, C.J., and Walley, R.J. (2006). Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Statistics in Medicine* **25,** 183–203.

Wu, Z.G., He, P., and Geng, Z. (2011). Sufficient conditions for concluding surrogacy based on observed data. *Statistics in Medicine* **30**, 2422–2434.

Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association* **87**, 284–292.

Zigler, C.M., and Belin, T.R. (2012). A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint. *Biometrics* **68**, 922–932.