



---

Johns Hopkins University, Dept. of Biostatistics Working Papers

---

5-22-2009

# A BAYESIAN SHRINKAGE MODEL FOR INCOMPLETE LONGITUDINAL BINARY DATA WITH APPLICATION TO THE BREAST CANCER PREVENTION TRIAL

C. Wang

*Department of Statistics, University of Florida*

M.J. Daniels

*Department of Statistics, University of Florida*

Daniel O. Scharfstein

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, dscharf@jhsp.edu*

S. Land

*Department of Biostatistics, University of Pittsburgh*

---

## Suggested Citation

Wang, C.; Daniels, M.J.; Scharfstein, Daniel O.; and Land, S., "A BAYESIAN SHRINKAGE MODEL FOR INCOMPLETE LONGITUDINAL BINARY DATA WITH APPLICATION TO THE BREAST CANCER PREVENTION TRIAL" (May 2009). *Johns Hopkins University, Dept. of Biostatistics Working Papers*. Working Paper 188. <http://biostats.bepress.com/jhubiostat/paper188>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

# A Bayesian Shrinkage Model for Incomplete Longitudinal Binary Data with Application to the Breast Cancer Prevention Trial

C. Wang <sup>\*</sup>; M.J. Daniels <sup>†</sup>; D. O. Scharfstein <sup>‡</sup>; S. Land <sup>§</sup>

*Abstract:* We consider inference in randomized studies, in which repeatedly measured outcomes may be informatively missing due to drop out. In this setting, it is well known that full data estimands are not identified unless unverified assumptions are imposed. We assume a non-future dependence model for the drop-out mechanism and posit an exponential tilt model that links non-identifiable and identifiable distributions. This model is indexed by non-identified parameters, which are assumed to have an informative prior distribution, elicited from subject-matter experts. Under this model, full data estimands are shown to be expressed as functionals of the distribution of the observed data. To avoid the curse of dimensionality, we model the distribution of the observed data using a Bayesian shrinkage model. In a simulation study, we compare our approach to a fully parametric and a fully saturated model for the distribution of the observed data. Our methodology is motivated and applied to data from the Breast Cancer Prevention Trial.

*KEY WORDS:* Informative drop-out; Prior elicitation.

---

<sup>\*</sup>Department of Statistics, University of Florida, Gainesville, FL 32611

<sup>†</sup>Department of Statistics, University of Florida, Gainesville, FL 32611

<sup>‡</sup>Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21205

<sup>§</sup>Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15213

# 1 Introduction

## 1.1 Breast Cancer Prevention Trial

The Breast Cancer Prevention Trial (BCPT) was a large multi-center, double-blinded, placebo-controlled, chemoprevention trial of the National Surgical Adjuvant Breast and Bowel Project (NSABP) designed to test the efficacy of 20mg/day tamoxifen in preventing breast cancer and coronary heart disease in healthy women at risk for breast cancer (Fisher et al., 1998). The study was open to accrual from June 1, 1992 through September 30, 1997 and 13,338 women aged 35 or older were enrolled in the study during this interval. The primary objective was to determine whether long-term tamoxifen therapy is effective in preventing the occurrence of invasive breast cancer. Secondary objectives included quality of life (QOL) assessments to evaluate benefit as well as risk resulting from the use of tamoxifen.

Monitoring QOL was of particular importance for this trial since the participants were healthy women and there had been concerns voiced by researchers about the association between clinical depression and tamoxifen use. Accordingly, data on depression symptoms was scheduled to be collected at baseline prior to randomization, at 3 months, at 6 months and every 6 months thereafter for up to 5 years. The primary instrument used to monitor depressive symptoms over time was the Center for Epidemiologic Studies Depression Scale (CES-D)(Radloff, 1977). This self-test questionnaire is composed of 20 items, each of which is scored on a scale of 0-3. A score of 16 or higher is considered as a likely case of clinical depression.

The trial was unblinded on March 31, 1998, after an interim analysis showed a dramatic reduction in the incidence of breast cancer in the treatment arm. Due to the potential loss of the control arm, we focus on QOL data collected on the 10,982 participants who were enrolled during the first two years of accrual and had their CES-D score recorded at baseline. All women in this cohort had the potential for three years of follow-up (before the unblinding).

In the BCPT, the clinical centers were not required to collect QOL data on women after they stopped their assigned therapy. This design feature aggravated the problem of missing QOL data in the trial. As reported in Land et al. (2002), more than 30% of the CES-D scores were missing at the 36-month follow-up, with a slightly higher percentage in the tamoxifen group. They also showed that women with higher baseline CES-D scores had higher rates of missing data at each follow-up visit and the mean observed CES-D scores preceding a missing measurement were higher than those preceding an observed measurement; there was no evidence that these relationships differed by treatment group.

While these results suggest that the missing data process is associated with observed QOL outcomes, one cannot rule out the possibility that the process is further related to unobserved outcomes and that this relationship is modified by treatment. In particular, investigators were concerned (a priori) that, between assessments, tamoxifen might be causing depression in some individuals, who then do not return for their next assessment. If this occurs, the data are said to be missing not at random (MNAR); otherwise the data are said to be missing at random (MAR).

## 1.2 Informative Drop-out in Longitudinal Studies

In this paper, we will concern ourselves with inference in longitudinal studies, where individuals who miss visits do not return for subsequent visits (i.e., drop-out). In such a setting, MNAR is often referred to as informative drop-out. While there were some intermittent responses in the BCPT, we will, as in Land et al. (2002), consider a “monotonized” dataset, whereby all CES-D scores observed on an individual after their first missing score have been deleted (this increases the “dropout” rate).

There are two main inferential paradigms for analyzing longitudinal studies with informative drop-out: likelihood (parametric) and non-likelihood (semi-parametric). Articles by Little (1995), Hogan and Laird (1997a) and Kenward and Molenberghs (1999) as well as recent books by Molenberghs and Kenward (2007) and Daniels and Hogan (2008) provide

a comprehensive review of likelihood-based approaches, including selection models, pattern-mixture models, and shared-parameter models. These models differ in the way the joint distribution of the outcome and missing data processes are factorized. In selection models, one specifies a model for the marginal distribution of the outcome process and a model for the conditional distribution of the drop-out process given the outcome process (see, for example, Heckman, 1979; Diggle and Kenward, 1994; Baker, 1995; Fitzmaurice et al., 1995; Molenberghs et al., 1997; Liu et al., 1999; Albert, 2000); in pattern-mixture models, one specifies a model for the conditional distribution of the outcome process given the drop-out time and the marginal distribution of the drop-out time (see, for example, Little, 1993, 1994, 1995; Hogan and Laird, 1997b; Fitzmaurice and Laird, 2000; Daniels and Hogan, 2000; Roy, 2003; Birmingham and Fitzmaurice, 2002; Thijs et al., 2002; Pauler et al., 2003; Roy and Daniels, 2008); and in shared-parameter models, the outcome and drop-out processes are assumed to be conditionally independent given shared random effects (see, for example, Wu and Carroll, 1988; DeGruttola and Tu, 1994; Ten Have et al., 1998, 2000; Pulkstenis et al., 1998; Land et al., 2002; Yuan and Little, 2009). Traditionally, these models have relied on very strong distributional assumptions in order to obtain model identifiability.

Without these strong distributional assumptions, specific parameters from these models would not be identified from the distribution of the observed data. To address this issue within a likelihood-based framework, several authors (Nordheim, 1984; Baker et al., 1992; Little, 1994; Little and Rubin, 1999; Kurland and Heagerty, 2004; Daniels and Hogan, 2008) have promoted the use of global sensitivity analysis, whereby non- or weakly- identified, interpretable parameters are fixed and then varied to evaluate the robustness of the inferences. Scientific experts can be employed to constrain the range of these parameters.

Non-likelihood approaches to informative drop-out in longitudinal studies have been primarily developed from a selection modeling perspective. Here, the marginal distribution of the outcome process is modeled non- or semi-parametrically and the conditional distribution of the drop-out process given the outcome process is modeled semi- or fully- parametrically.

In the case where the drop-out process is assumed to depend only on observable outcomes (i.e., MAR), Robins et al. (1994, 1995), van der Laan and Robins (2003) and Tsiatis (2006) developed inverse-weighted and augmented inverse-weighted estimating equations for inference. For informative drop-out, Rotnitzky et al. (1998), Scharfstein et al. (1999) and Rotnitzky et al. (2001) introduced a class of selection models, in which the model for drop-out is indexed by interpretable sensitivity parameters that express departures from MAR. Inference using inverse-weighted estimating equations was proposed.

The problem with the aforementioned sensitivity analysis approaches is that the ultimate inferences can be cumbersome to display. Vansteelandt et al. (2006) developed a method for reporting ignorance and uncertainty intervals (regions) that contain the true parameter(s) of interest with a prescribed level of precision, when the true data generating model is assumed to fall within a plausible class of models (as an example, see Scharfstein et al., 2004). An alternative and very natural strategy is specify an informative prior distribution on the non- or weakly- identified parameters and conduct a fully Bayesian analysis, whereby the ultimate inferences are reported in terms of posterior distributions. In the cross-sectional setting with a continuous outcome, Scharfstein et al. (2003) adopted this approach from a semi-parametric selection modeling perspective. Kaciroti et al. (2009) proposed a parametric pattern-mixture model for cross-sectional, clustered binary outcomes . Lee et al. (2008) introduced a fully-parametric pattern-mixture approach in the longitudinal setting with binary outcomes. In this paper, we consider the same setting as Lee et al. (2008), but offer a more flexible strategy. In the context of BCPT, the longitudinal outcome will be the indicator that the CES-D score is 16 or higher.

### 1.3 Outline

The paper is organized as follows. In Section 2, we describe the data structure. In Section 3 and 4, we formalize identification assumptions and prove that the full-data distribution is identified under these assumptions. We introduce a saturated model for the distribution

of the observed data in Section 5. In Section 6, we illustrate how to apply shrinkage priors to high-order interaction parameters in the saturated model to reduce the dimensionality of the parameter space and how to elicit (conditional) informative priors for non-identified sensitivity parameters from experts. In Section 7, we assess, by simulation, the behavior of three classes of models for the distribution of observed data; parametric, saturated, and shrinkage. Our analysis of the BCPT trial is presented in Section 8. Section 9 is devoted to a summary and discussion.

## 2 Data Structure and Notation

Let  $Z$  denote the treatment assignment indicator, where  $Z = 1$  denotes tamoxifen and  $Z = 0$  denotes placebo. Let  $Y_j$  denote the binary outcome (i.e., depression) scheduled to be measured at the  $j$ th visit ( $j = 0(\text{baseline}), \dots, J$ ) and let  $\bar{\mathbf{Y}}_j = (Y_0, \dots, Y_j)$  denote the history of the outcome process through visit  $j$ . Let  $R_j$  denote the indicator that an individual has her depression status recorded at visit  $j$ . We assume that  $R_0 = 1$  (i.e.,  $Y_0$  is always observed) and  $R_j = 0$  implies that  $R_{j+1} = 0$  (i.e., monotone missing data). Let  $C = \max\{t : R_t = 1\}$  be the last visit at which an individual's depression status is recorded. The full and observed data for an individual are  $F = (Z, C, \bar{\mathbf{Y}}_J)$  and  $O = (Z, C, \bar{\mathbf{Y}}_C)$ , respectively. We assume that we observe  $n$  i.i.d., copies of  $O$ . We will use the subscript  $i$  to denote data for the  $i$ th individual.

Our goal is to draw inference about  $\mu_{z,j}^* = P[Y_j = 1 | Z = z]$  for  $j = 1, \dots, J$  and  $z = 0, 1$ .

## 3 Assumptions

To identify  $\mu_{z,j}^*$  from the distribution of the observed data, we make the following two untestable assumptions:

**Assumption 1 (Non-Future Dependence):**  $R_j$  is independent of  $(Y_{j+1}, \dots, Y_J)$  given  $R_{j-1} = 1$  and  $\bar{\mathbf{Y}}_j$ , for  $j = 1, \dots, J - 1$ .

This assumption asserts that for individuals at risk for drop-out at visit  $j$  and who share the same history of outcomes up to and including visit  $j$ , the distribution of future outcomes is the same for those who are last seen at visit  $j$  and those who remain on study past visit  $j$ . This assumption has been referred to as non-future dependence (Kenward et al., 2003).

**Assumption 2 (Pattern-Mixture Representation):** For  $j = 1, \dots, J$  and  $y_j = 0, 1$ ,

$$P[Y_j = y_j | R_j = 0, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z] = \frac{P[Y_j = y_j | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z = z] \exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, y_j)\}}{E[\exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)\} | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z = z]}$$

where  $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$  is a specified function of its arguments.

Assumption 2 links the non-identified conditional distribution of  $Y_j$  given  $R_j = 0$ ,  $R_{j-1} = 1$ ,  $\bar{\mathbf{Y}}_{j-1}$ , and  $Z = z$  to the identified conditional distribution of  $Y_j$  given  $R_j = 1$ ,  $\bar{\mathbf{Y}}_{j-1}$ , and  $Z = z$  using exponential tilting via the specified function  $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ . Assumption (2) has a selection model representation that is obtained using Bayes' rule.

**Assumption 2 (Selection Model Representation):** For  $j = 1, \dots, J$ ,

$$\text{logit} \{P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z]\} = h_{z,j}(\bar{\mathbf{Y}}_{j-1}) + q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$$

where

$$h_{z,j}(\bar{\mathbf{Y}}_{j-1}) = \text{logit} P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z] - \log\{E[\exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)\} | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z = z]\}$$

With this characterization, we see that the function  $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$  quantifies the influence (on a log odds ratio scale) of the potentially unobservable outcome  $Y_j$  on the conditional odds of dropping at time  $j$ .

## 4 Identifiability

The above two assumptions non-parametrically, just-identify  $\mu_{z,j}^*$  for all  $j = 1, \dots, J$  and  $z = 0, 1$ . To see this, consider the following representation of this conditional distribution,



derived using the laws of total and conditional probability:

$$\begin{aligned} \mu_{z,j}^* &= \sum_{\bar{\mathbf{y}}_{j-1}} P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \times \\ &\quad \left\{ \prod_{l=1}^j P[R_l = 1 | R_{l-1} = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \prod_{l=0}^{j-1} P[Y_l = y_l | R_l = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \right\} + \\ &\quad \sum_{k=1}^j \sum_{\bar{\mathbf{y}}_{k-1}} P[Y_j = 1 | R_k = 0, R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] P[R_k = 0 | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \\ &\quad \times \left\{ \prod_{l=1}^{k-1} P[R_l = 1 | R_{l-1} = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \prod_{l=0}^{k-1} P[Y_l = y_l | R_l = 1, \bar{\mathbf{Y}}_{l-1} = \bar{\mathbf{y}}_{l-1}, Z = z] \right\} \end{aligned}$$

All quantities on the right hand side of this equation are identified, without appealing to any assumptions, except  $P[Y_j = 1 | R_k = 0, R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$  for  $k = 1, \dots, j-1$ . Under Assumptions 1 and 2, these probabilities can be shown to be identified, implying that  $\mu_{z,j}^*$  is identified for all  $j$  and  $z$ .

**Theorem 1:**  $P[Y_j = 1 | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$  and  $P[Y_j = 1 | R_k = 0, R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z]$  are identified for  $k = 1, \dots, j$ .

**Proof:** The proof follows by backward induction. Consider  $k = j$ . By Assumption 2,

$$\begin{aligned} P[Y_j = 1 | R_j = 0, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] &= \\ &= \frac{P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, 1)\}}{E[\exp\{q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)\} | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]} \end{aligned}$$

Since the right hand side is identified, we know that  $P[Y_j = 1 | R_j = 0, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  is identified. Further, we can write

$$\begin{aligned} &P[Y_j = 1 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \\ &= \sum_{r=0}^1 P[Y_j = 1 | R_j = r, R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] P[R_j = r | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] \end{aligned}$$

Since all quantities on the right hand side are identified,  $P[Y_j = 1 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  is identified.

Suppose that  $P[Y_j = 1|R_k = 0, R_{k-1} = 1, \bar{Y}_{k-1} = \bar{y}_{k-1}, Z = z]$  and  $P[Y_j = 1|R_{k-1} = 1, \bar{Y}_{k-1} = \bar{y}_{k-1}, Z = z]$  are identified for some  $k$  where  $1 < k < j$ . Then, we need to show that these probabilities are identified for  $k' = k - 1$ . To see this, note that

$$\begin{aligned}
& P[Y_j = 1|R_{k'} = 0, R_{k'-1} = 1, \bar{Y}_{k'-1} = \bar{y}_{k'-1}, Z = z] \\
&= P[Y_j = 1|R_{k-1} = 0, R_{k-2} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \\
&= \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 0, R_{k-2} = 1, \bar{Y}_{k-1} = \bar{y}_{k-1}, Z = z] \times \\
&\quad P[Y_{k-1} = y_{k-1}|R_{k-1} = 0, R_{k-2} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \\
&= \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 1, \bar{Y}_{k-1} = \bar{y}_{k-1}, Z = z] \times \\
&\quad \frac{P[Y_{k-1} = y_{k-1}|R_{k-1} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \exp\{q_{z,k-1}(\bar{Y}_{k-2}, y_{k-1})\}}{E[\exp\{q_{z,k-1}(\bar{Y}_{k-2}, Y_{k-1})\}|R_{k-1} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z]}
\end{aligned}$$

The third equality follows by Assumptions 1 and 2. Since all the quantities on the right hand side of the last equality are identified,  $P[Y_j = 1|R_{k'} = 0, R_{k'-1} = 1, \bar{Y}_{k'-1} = \bar{y}_{k'-1}, Z = z]$  is identified. Further,

$$\begin{aligned}
& P[Y_j = 1|R_{k'-1} = 1, \bar{Y}_{k'-1} = \bar{y}_{k'-1}, Z = z] \\
&= P[Y_j = 1|R_{k-2} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \\
&= \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 1, \bar{Y}_{k-1} = \bar{y}_{k-1}, Z = z] \times \\
&\quad P[Y_{k-1} = y_{k-1}|R_{k-1} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \times \\
&\quad P[R_{k-1} = 1|R_{k-2} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] + \\
&\quad \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 0, R_{k-2} = 1, \bar{Y}_{k-1} = \bar{y}_{k-1}, Z = z] \times \\
&\quad P[Y_{k-1} = y_{k-1}|R_{k-1} = 0, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \times \\
&\quad P[R_{k-1} = 0|R_{k-2} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \\
&= \sum_{y_{k-1}=0}^1 P[Y_j = 1|R_{k-1} = 1, \bar{Y}_{k-1} = \bar{y}_{k-1}, Z = z] \times \\
&\quad P[Y_{k-1} = y_{k-1}|R_{k-1} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] \times \\
&\quad P[R_{k-1} = 1|R_{k-2} = 1, \bar{Y}_{k-2} = \bar{y}_{k-2}, Z = z] +
\end{aligned}$$

$$\begin{aligned} & \sum_{y_{k-1}=0}^1 P[Y_j = 1 | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-1} = \bar{\mathbf{y}}_{k-1}, Z = z] \times \\ & \frac{P[Y_{k-1} = y_{k-1} | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \exp\{q_{z,k-1}(\bar{\mathbf{Y}}_{k-2}, y_{k-1})\}}{E[\exp\{q_{z,k-1}(\bar{\mathbf{Y}}_{k-2}, Y_{k-1})\} | R_{k-1} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z]} \times \\ & P[R_{k-1} = 0 | R_{k-2} = 1, \bar{\mathbf{Y}}_{k-2} = \bar{\mathbf{y}}_{k-2}, Z = z] \end{aligned}$$

The third equality follows by Assumptions 1 and 2. Since all the quantities on the right hand side of the last equality are identified,  $P[Y_j = 1 | R_{k'-1} = 1, \bar{\mathbf{Y}}_{k'-1} = \bar{\mathbf{y}}_{k'-1}, Z = z]$  is identified.  $\square$

The identifiability result shows that, given the functions  $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$ ,  $\mu_{z,j}^*$  can be expressed as functional of the distribution of the observed data. In particular, the functional depends on the conditional distributions of  $Y_j$  given  $R_j = 1$ ,  $\bar{\mathbf{Y}}_{j-1}$ , and  $Z$  for  $j = 0, \dots, J$  and the conditional distributions of  $R_j$  given  $R_{j-1} = 1$ ,  $\bar{\mathbf{Y}}_{j-1}$  and  $Z$  for  $j = 1, \dots, J$ . Furthermore, the functions  $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$  are not identifiable from the distribution of the observed data and their specification places no restrictions on the distribution of the observed data.

## 5 Modeling

We specify saturated models for the observed data via the sequential conditional distributions of  $[Y_j | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z]$  for  $j = 0, \dots, J$  and the conditional hazards  $[R_j | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z]$  for  $j = 1, \dots, J$ . We parameterize these models as follows:

$$\begin{aligned} \text{logit } P[Y_0 = 1 | R_0 = 1, Z = z] &= \alpha_{z,0,0} \\ \text{logit } P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] &= \alpha_{z,j,0} + \alpha_{z,j,1} y_{j-1} + \sum_{k=0}^{j-2} \alpha_{z,j,k}^{(1)} y_k \\ &+ \sum_{k,l \in A_j^{(2)}} \alpha_{z,j}^{(2)} y_k y_l + \sum_{k,l,m \in A_j^{(3)}} \alpha_{z,j}^{(3)} y_k y_l y_m + \dots + \alpha_{z,j}^{(j-1)} y_0 y_1 \dots y_{j-1} \\ \text{logit } P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] &= \gamma_{z,j,0} + \gamma_{z,j,1} y_{j-1} + \sum_{k=0}^{j-2} \gamma_{z,j,k}^{(1)} y_k \\ &+ \sum_{k,l \in A_j^{(2)}} \gamma_{z,j}^{(2)} y_k y_l + \sum_{k,l,m \in A_j^{(3)}} \gamma_{z,j}^{(3)} y_k y_l y_m + \dots + \gamma_{z,j}^{(j-1)} y_0 y_1 \dots y_{j-1} \end{aligned}$$

for  $j = 1, \dots, J$ , where  $A_j^{(t)}$  is the set of all  $t$ -tuples of the integers  $0, \dots, j-1$ . Let  $\boldsymbol{\alpha}$  denote the parameters indexing the conditional distributions  $[Y_j | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z]$ ,  $\boldsymbol{\gamma}$  denote the parameters indexing the conditional distributions  $[R_j | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1}, Z]$  and  $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\gamma}\}$ .

Furthermore, we propose to parameterize the functions  $q_{z,j}(\bar{\mathbf{Y}}_{j-1}, Y_j)$  with parameters  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}} = q_{z,j}((\bar{\mathbf{y}}_{j-1}, 1)) - q_{z,j}((\bar{\mathbf{y}}_{j-1}, 0))$ . Here,  $\exp(\tau_{z,j,\bar{\mathbf{y}}_{j-1}})$  represents, in the context of the BCPT trial, the conditional odds ratio of dropping out between visits  $j-1$  and  $j$  for individuals who are depressed vs. not depressed at visit  $j$ , but share the mental history  $\bar{\mathbf{y}}_{j-1}$  through visit  $j-1$ . We let  $\boldsymbol{\tau}$  denote the collection of  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$ 's.

## 6 Prior Specification and Posterior Computation

For specified sensitivity analysis parameters  $\boldsymbol{\tau}$ , the saturated model proposed in Section 5 provides a perfect fit to the distribution of the observed data. In this model, however, the number of parameters increases exponentially in  $J$ . In contrast, the number of data points increases linearly in  $J$ . As a consequence, there will be many combinations of  $\bar{\mathbf{y}}_{j-1}$  (i.e., "cells") which will be sparsely represented in the dataset. For example, in the BCPT trial, about 50% of the possible realizations of  $\bar{\mathbf{Y}}_7$  have less than two observations and about 15% have no observations. For a frequentist perspective, this implies that components of  $\boldsymbol{\theta}$  will be imprecisely estimated; in turn, this can adversely affect estimation of  $\mu_{z,j}^*$ . This has been called the curse of dimensionality (Robins and Ritov, 1997).

### 6.1 Shrinkage Priors on $\boldsymbol{\theta}$

To address this problem, we introduce data driven shrinkage priors for higher order interactions to reduce the number of parameters in an automated manner. In particular, we assume

$$\alpha_{z,j,k}^{(t)} \sim N(0, \sigma_\alpha^{(t)}) \quad \text{and} \quad \gamma_{z,j,k}^{(t)} \sim N(0, \sigma_\gamma^{(t)}) \quad k \in A_j^{(t)}, 3 \leq t < j \leq J, z = 0, 1 \quad (1)$$

where  $t$  is the order of interactions and the hyper-parameters (shrinkage variances) follow distributions

$$\sigma_{\alpha}^{(t)} \sim \text{Unif}(0, 10) \quad \text{and} \quad \sigma_{\gamma}^{(t)} \sim \text{Unif}(0, 10).$$

When  $\sigma_{\alpha}^{(t)}$  and  $\sigma_{\gamma}^{(t)}$  equal zero for all interactions, the saturated model is reduced to a first order Markov model,

$$\text{logit } P[Y_0 = 1 | R_0 = 1, Z = z] = \alpha_{z,0,0}$$

$$\text{logit } P[Y_j = 1 | R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \alpha_{z,j,0} + \alpha_{z,j,1}y_{j-1}$$

$$\text{logit } P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \gamma_{z,j,0} + \gamma_{z,j,1}y_{j-1}.$$

The shrinkage priors allow the “neighboring” cells in the observed data model to borrow information from each other and provide more precise estimates.

When the first order Markov model is not true, as  $n$  goes to infinity, the posterior means of observed data probabilities will converge to their true values as long as the shrinkage priors are  $O(1)$  (which is the case here) and all the true values of the observed data probabilities,  $P[Y_j | R_j = 1, \bar{\mathbf{Y}}_{j-1}, Z]$  for  $j = 0, \dots, J$  and are in the open interval,  $(0, 1)$ . This follows, since under this latter condition, all combinations of depression histories have a positive probability of being observed and the prior will become swamped by the observed data. However, when the true value of any of the observed data probabilities is zero or one, there exists at least one combination of depression history that will never be observed and thus the influence of the prior will not dissipate as  $n$  increases.

We specify non-informative priors  $N(0, 1000)$  for the non-interaction parameters in  $\boldsymbol{\theta}$ , namely  $\alpha_{z,j,0}$  for  $j = 0, \dots, J$  and  $z = 0, 1$ ,  $\alpha_{z,j,1}$ ,  $\gamma_{z,j,0}$  and  $\gamma_{z,j,1}$  for  $j = 1, \dots, J$  and  $z = 0, 1$ .

## 6.2 $\tau$ given $\theta$

The sensitivity parameters in Assumption 2, defined formally in Section 5, are (conditional) odds ratios. In our experience, subject matter experts often have difficulty thinking in terms of odds ratios; rather, they are more comfortable expressing beliefs about relative risks (Scharfstein et al., 2006; Shepherd et al., 2007). With this in mind, we asked Dr. Patricia Ganz, a medical oncologist and expert on quality of life outcomes in breast cancer, to express her beliefs about the risk of dropping out and its relationship to treatment assignment and depression. We then translated her beliefs into prior distributional assumptions about the odds ratio sensitivity parameters  $\tau$ .

Specifically, we asked Dr. Ganz to answer the following question for each treatment group:

**Q:** *Consider a group of women assigned to placebo (tamoxifen), who are on study through visit  $j - 1$  and who share the same history of depression. Suppose that the probability that a randomly selected woman in this group drops out before visit  $j$  is  $p$  (denoted by the columns in Table 1). For each  $p$ , what is the minimum, maximum and your best guess (median) representing how much more (e.g. twice) or less (e.g., half) likely you consider the risk of dropping out before visit  $j$  for a woman who would be depressed at year  $j$  RELATIVE to a woman who would not be depressed at visit  $j$ ?*

Implicit in this question is the assumption that, for each treatment group, the relative risk only depends on past history and the visit number only through the risk of dropping out between visits  $j - 1$  and  $j$ .

For notational convenience, let  $r_z(p)$  denote the relative risk of drop-out for treatment group  $z$  and drop-out probability  $p$ . Further, let  $r_{z,min}(p)$ ,  $r_{z,med}(p)$  and  $r_{z,max}(p)$  denote the elicited minimum, median, and maximum relative risks (see Table 1). Let  $p_{z,j}(\bar{\mathbf{y}}_{j-1}) = P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  and let  $p_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1}) = P[R_j = 0 | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Y_j = y, Z = z]$  for  $y = 0, 1$ .

By definition,

$$r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) = p_{z,j}^{(1)}(\bar{\mathbf{y}}_{j-1})/p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$$

$$p_{z,j}(\bar{\mathbf{y}}_{j-1}) = \sum_{y=0}^1 p_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1})\pi_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1})$$

where  $\pi_{z,j}^{(y)}(\bar{\mathbf{y}}_{j-1}) = P[Y_j = y | R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  for  $y = 0, 1$ . This implies that

$$p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}) = \frac{p_{z,j}(\bar{\mathbf{y}}_{j-1})}{\pi_{z,j}^{(1)}(\bar{\mathbf{y}}_{j-1})(r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) - 1) + 1}.$$

Since  $\pi_{z,j}^{(1)}(\bar{\mathbf{y}}_{j-1}) \in [0, 1]$ , given  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$  and  $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$ ,  $p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$  is bounded as follows:

for  $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) \geq 1$ ,

$$p_{z,j}(\bar{\mathbf{y}}_{j-1})/r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) \leq p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}) \leq \min\{p_{z,j}(\bar{\mathbf{y}}_{j-1}), 1\}$$

and, for  $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})) \leq 1$ ,

$$p_{z,j}(\bar{\mathbf{y}}_{j-1}) \leq p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}) \leq \min\{p_{z,j}(\bar{\mathbf{y}}_{j-1})/r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}.$$

We will use these bounds to construct our prior.

We construct the conditional prior of  $\tau_{z,j;\bar{\mathbf{y}}_{j-1}}$  given  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$  using Steps 1-4 given below. The general strategy is to use the elicited information on the relative risk at different drop-out probabilities and the bounds derived above to construct the prior of interest.

Step 1. For  $m \in \{min, med, max\}$ , interpolate the elicited  $r_{z,m}(p)$  at different drop-out probabilities (see Figure 1) to find  $r_{z,m}(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$  for any  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ .

Step 2. Construct the prior of  $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$  given  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$  as a 50-50 mixture of

Uniform( $r_{z,min}(p_{z,j}(\bar{\mathbf{y}}_{j-1})), r_{z,med}(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$ )  
 and  
 Uniform( $r_{z,med}(p_{z,j}(\bar{\mathbf{y}}_{j-1})), r_{z,max}(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$ )

random variables. This preserves the elicited percentiles of the relative risk.

Step 3. Construct a conditional prior of  $p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$  given  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$  and  $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$  as a uniform distribution with lower bound

$$\frac{p_{z,j}(\bar{\mathbf{y}}_{j-1})}{\max \{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}}$$

and upper bound

$$\min \left\{ \frac{p_{z,j}(\bar{\mathbf{y}}_{j-1})}{\min \{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}}, \frac{1}{\max \{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1})), 1\}} \right\}.$$

The bounds were derived above.

Step 4. Steps (2) and (3) induce a prior for  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}|\boldsymbol{\theta}$  by noting

$$\tau_{z,j,\bar{\mathbf{y}}_{j-1}} = \log \left( \frac{r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))(1 - p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1}))}{1 - r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})} \right),$$

i.e.,  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$  is a deterministic function of  $r_z(p_{z,j}(\bar{\mathbf{y}}_{j-1}))$  and  $p_{z,j}^{(0)}(\bar{\mathbf{y}}_{j-1})$ .

The relative risks elicited from Dr. Ganz are given in Table 2. We extrapolated the relative risks outside the ranges given in Table 2 as shown in Figure 1.

Figure 2 shows the density of  $\tau$  given  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$  equal 10% and 25% for the tamoxifen and placebo arms. For two patients with the same response history up to time point  $j - 1$ , the log odds ratio of dropping out at time point  $j$ , for the patient that is depressed at time point  $j$  versus the patient that is not, increases as the overall drop out rate at time point  $j$  increases. In general, for a given  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ , the log odds ratio is higher for patients in the tamoxifen versus placebo arms.

### 6.3 Posterior Computation

With the shrinkage priors on  $\boldsymbol{\theta}$ , the elicited conditional priors  $\boldsymbol{\tau}$  given  $\boldsymbol{\theta}$ , and the observed data, the following steps are used to simulate draws from the posterior of  $\mu_{z,j}^*$ :

1. Using the proposed observed data model with the shrinkage priors on  $\boldsymbol{\theta}$ , we simulate draws from the posterior distributions of  $P[Y_j = 1|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  and  $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  for all  $j, z$  and  $\bar{\mathbf{y}}_{j-1}$  in WinBUGS.



2. For each draw of  $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ , we draw  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$  based on the conditional priors described in Section 6.2.
3. We compute  $\mu_{z,j}^*$  by plugging the draws of  $P[Y_j = 1|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$ ,  $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  and  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$  into the identification algorithm discussed in Section 4.

To sample from the posterior distributions of  $P[Y_j = 1|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  and  $P[R_j = 0|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z]$  in WinBUGS we stratify the individual binary data (by previous response history) and analyze as Binomial data; this serves to drastically improve the computational efficiency. Sampling  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$  and computing  $\mu_{z,j}^*$  is implemented separately from the first step using R.

## 7 Assessment of Model Performance via Simulation

Via simulation, we compared the performance of the shrinkage model with a correct parametric model (given below), an incorrect parametric model (first order Markov model) and the saturated model with diffuse priors (given below).

The shrinkage model uses the shrinkage priors proposed in Section 6.1 (shrink the saturated model toward a first order Markov model). Note that the shrinkage priors shrink the saturated model to an incorrect parametric model.

For the saturated model with diffuse priors, we re-parameterize the model as

$$P[Y_j = 1|R_j = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \mu_{z,\bar{\mathbf{y}}_{j-1}}$$

$$P[R_j = 0|R_{j-1} = 1, \bar{\mathbf{Y}}_{j-1} = \bar{\mathbf{y}}_{j-1}, Z = z] = \rho_{z,\bar{\mathbf{y}}_{j-1}}$$

for  $j = 1, \dots, 7$ , and specify independent  $\text{Unif}(0, 1)$  on  $\mu$ 's and  $\rho$ 's.

We simulated observed data from a "true" parametric model of the following form:

$$\text{logit}P[Y_0 = 1|R_0 = 1, Z = z] = \alpha_{z,0,0}$$

$$\text{logit}P[Y_1 = 1|R_1 = 1, Y_0 = y_0, Z = z] = \alpha_{z,1,0} + \alpha_{z,1,1}y_0$$

$$\text{logit}P[R_1 = 0|R_0 = 1, Y_0 = y_0, Z = z] = \gamma_{z,1,0} + \gamma_{z,1,1}y_0$$

$$\text{logit}P[Y_j = 1|R_j = 1, \bar{Y}_{j-1} = \bar{y}_{j-1}, Z = z] = \alpha_{z,j,0} + \alpha_{z,j,1}y_{j-1} + \alpha_{z,j,2}y_{j-2}$$

$$\text{logit}P[R_j = 0|R_{j-1} = 1, \bar{Y}_{j-1} = \bar{y}_{j-1}, Z = z] = \gamma_{z,j,0} + \gamma_{z,j,1}y_{j-1} + \gamma_{z,j,2}y_{j-2},$$

for  $j = 2$  to  $7$ .

We fit the "true" model to the BCPT data and used the posterior mean of parameters  $\alpha$  and  $\gamma$  as the true parameters. We identify the full data model by using the elicited prior from the BCPT (see Table 2). The parameters and the true depression rates  $\mu_{z,j}^*$ , computed by Monte Carlo integration, are given in Table 3.

We considered (relatively) small (3000), moderate (5000), and large (10000) sample sizes for each treatment arm; for each sample size, we simulated 50 datasets. We assessed model performance using the mean squared error (MSE) criterion.

In Table 4, we report the MSEs of  $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$  and  $P[R_j = 1|R_{j-1} = 1, \bar{Y}_{j-1}, Z = z]$  averaged over all  $j$  and all  $\bar{Y}_{j-1}$  (see columns 3 and 4, respectively). We also report the MSEs for  $\mu_{z,j}^*$  (see columns 6-12). For reference, the MSEs associated with the true data generating model are bolded. This table demonstrate that the shrinkage model generally outperforms both the incorrectly specified parametric model and the saturated model at all sample sizes. This improved performance is especially noticeable when comparing the MSEs for the rates of depression at times 3-7.

In addition, the MSEs for the shrinkage model compare favorably with those of the true parametric model for all sample sizes considered, despite the fact that the shrinkage priors were specified to shrink toward an incorrect model.

## 8 Application: Breast Cancer Prevention Trial (BCPT)

Table 5 displays the treatment-specific monotonized drop-out rates in the BCPT. By the 7th study visit, more than 40% of patients had missed one or more assessments, with a slightly higher percentage in the tamoxifen arm.

We fit the shrinkage model to the observed data using WinBUGS, with four chains of 8000 iterations and 1000 burn-in. Convergence was checked by examining trace plots of the multiple chains.

### 8.1 Model Fit and Shrinkage Results

To assess the model fit, we compared the empirical rates and posterior means (with 95% credible intervals) of  $P[Y_j = 1, R_j = 1|Z = z]$  and  $P[R_j = 0|Z = z]$ . As shown in Figure 3, the shrinkage model fits the observed data well. Figure 4 illustrates the effect of shrinkage on the model fits by comparing the difference between the empirical rate and posterior mean of  $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$  for all  $j, z$  and  $\bar{Y}_{j-1}$ . We can see that for early time points, the difference is close to zero since there is little shrinkage applied to the model parameters. For later time points, more higher order interaction coefficients are shrunk toward zero and the magnitude of difference increases and drifts away from zero line. In general, the empirical estimates are less reliable for the later time points (re: the simulation results in Section 7). In some cases, there are no observations within "cells." By shrinking the high order interactions (i.e., borrowing information across neighboring cells), we are able to estimate  $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$  for all  $j, z$  and  $\bar{Y}_{j-1}$  with reasonable precision.

### 8.2 Inference

Figure 5 shows the posterior of  $P[Y_7 = 1|Z = z]$ , the treatment-specific probability of depression at the end of the 36-month follow up (solid lines). For comparison, the posterior under MAR (corresponding to point mass priors for  $\tau$  at zero) is also presented (dashed

lines). The observed depression rates (i.e., complete case analysis) were 0.115 on both the placebo and tamoxifen arms. Under the MNAR analysis (using the elicited priors), the posterior mean of the depression rates at month 36 were 0.126 (95%*CI* : 0.115, 0.138) and 0.130 (95%*CI* : 0.119, 0.143) for the placebo and tamoxifen arms; the difference was 0.004 (95%*CI* :  $-0.012, 0.021$ ). Under MAR, the rates were 0.125 (95%*CI* : 0.114, 0.136) and 0.126 (95%*CI* : 0.115, 0.138) for the placebo and tamoxifen arms; the difference was 0.001 (95%*CI* :  $-0.015, 0.018$ ). The posterior probability of depression was higher under the MNAR analysis than the MAR analysis since researchers believed depressed patients were more likely to drop out (see Table 2), a belief that was captured by the elicited priors. Figure 6 shows that under the two treatments there were no significant differences in the depression rates at every time point (95% credible intervals all cover zero) under both MNAR and MAR. Similar (non-significant) treatment differences were seen when examining treatment comparisons conditional on depression status at baseline.

## 9 Summary and Discussion

In this paper, we have presented a Bayesian shrinkage approach for longitudinal binary data with informative drop-out. Our model provides a framework that incorporates expert opinion about non-identifiable parameters and avoids the curse of dimensionality by using shrinkage priors. In our analysis of the BCPT data, we concluded that there was little (if any) evidence that women on tamoxifen were more depressed than those on placebo.

An important feature of our approach is that the specification of models for the identifiable distribution of the observed data and the non-identifiable parameters can be implemented by separate independent data analysts. This feature can be used to increase the objectivity of necessarily subjective inferences in the FDA review of randomized trials with informative drop-out.

The ideas in this paper can be extended to continuous outcomes. For example, one could

use the mixtures of Dirichlet processes model (Escobar and West, 1995) for the distribution of observed responses. They can also be extended to multiple cause dropout; in this trial, missed assessments were due to a variety of reasons including patient-specific causes such as experiencing a protocol defined event, stopping therapy, or withdrawing consent and institution-specific causes such as understaffing and staff turnover. Therefore, some missingness is less likely to be informative; extensions will need to account for that. In addition, institutional differences might be addressed by allowing institution-specific parameters with priors that shrink them toward a common set of parameters.

For smaller sample sizes, WinBUGS has difficulty sampling from the posterior distribution of the parameters in the shrinkage model. Tailored sampling algorithms can be written to overcome this difficulty. In addition, we are working on alternative parameterizations of the saturated model for the observed data as well as alternative shrinkage prior specifications to improve computational efficiency.

## Acknowledgments

This research was supported by NIH grants R01-CA85295, U10-CA37377, and U10-CA69974. The authors are grateful to oncologist Patricia Ganz at UCLA for providing her expertise for the MNAR analysis.

## References

- P.S. Albert. A Transitional Model for Longitudinal Binary Data Subject to Nonignorable Missing Data. *Biometrics*, 56(2):602–608, 2000.
- S.G. Baker. Marginal regression for repeated binary data with outcome subject to non-ignorable non-response. *Biometrics*, 51(3):1042–1052, 1995.

- S.G. Baker, W.F. Rosenberger, and R. DerSimonian. Closed-form estimates for missing counts in two-way contingency tables. *Statistics in Medicine*, 11:643–657, 1992.
- J. Birmingham and G.M. Fitzmaurice. A Pattern-Mixture Model for Longitudinal Binary Responses with Nonignorable Nonresponse. *Biometrics*, 58(4):989–996, 2002.
- M.J. Daniels and J.W. Hogan. Reparameterizing the Pattern Mixture Model for Sensitivity Analyses Under Informative Dropout. *Biometrics*, 56(4):1241–1248, 2000.
- M.J. Daniels and J.W. Hogan. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman & Hall/CRC, 2008.
- V. DeGruttola and X.M. Tu. Modelling Progression of CD4-lymphocyte Count and its Relationship to Survival Time. *Biometrics*, 50(4):1003–1014, 1994.
- P. Diggle and M.G. Kenward. Informative Drop-out Longitudinal Data Analysis. *Applied Statistics*, 43(1):49–93, 1994.
- M.D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, pages 577–588, 1995.
- B. Fisher, J.P. Costantino, D.L. Wickerham, C.K. Redmond, M. Kavanah, W.M. Cronin, V. Vogel, A. Robidoux, N. Dimitrov, J. Atkins, M. Daly, S. Wieand, E. Tan-Chiu, L. Ford, N. Wolmark, other National Surgical Adjuvant Breast, and Bowel Project Investigators. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *Journal of the National Cancer Institute*, 90:1371–1388, 1998.
- G.M. Fitzmaurice and N.M. Laird. Generalized Linear Mixture Models for Handling Non-ignorable Dropouts in Longitudinal Studies. *Biostatistics*, 1(2):141–156, 2000.
- G.M. Fitzmaurice, G. Molenberghs, and S.R. Lipsitz. Regression Models for Longitudinal

- Binary Responses with Informative Drop-Outs. *Journal of the Royal Statistical Society. Series B. Methodological*, 57(4):691–704, 1995.
- J. Heckman. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153–161, 1979.
- J.W. Hogan and N.M. Laird. Model-Based Approaches to Analysing Incomplete Longitudinal and Failure Time Data. *Statistics in Medicine*, 16(3):259–272, 1997a.
- J.W. Hogan and N.M. Laird. Mixture Models for the Joint Distribution of Repeated Measures and Event Times. *Statistics in Medicine*, 16(3):239–257, 1997b.
- N.A. Kaciroti, M.A. Schork, T Raghunathan, and S. Julius. A Bayesian Sensitivity Model for Intention-to-treat Analysis on Binary Outcomes with Dropouts. *Statistics in Medicine*, 28:572–585, 2009.
- M.G. Kenward and G. Molenberghs. Parametric Models for Incomplete Continuous and Categorical Longitudinal Data. *Statistical Methods in Medical Research*, 8(1):51, 1999.
- M.G. Kenward, G. Molenberghs, and H. Thijs. Pattern-mixture Models with Proper Time Dependence. *Biometrika*, 90(1):53–71, 2003.
- B.F. Kurland and P.J. Heagerty. Marginalized Transition Models for Longitudinal Binary Data with Ignorable and Non-Ignorable Drop-Out. *Statistics in Medicine*, 23(17):2673–2695, 2004.
- S. Land, S. Wieand, R. Day, T. Ten Have, J.P. Costantino, W. Lang, and P.A. Ganz. Methodological Issues In the Analysis of Quality of Life Data in Clinical Trials: Illustrations from the National Surgical Adjuvant Breast And Bowel Project (NSABP) Breast Cancer Prevention Trial. *Statistical Methods for Quality of Life Studies*, pages 71–85, 2002.

J.Y. Lee, J.W. Hogan, and B. Hitsman. Sensitivity analysis and informative priors for longitudinal binary data with outcome-related drop-out. *Technical Report, Brown University*, 2008.

R.J.A. Little. Pattern-Mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

R.J.A. Little. A Class of Pattern-Mixture Models for Normal Incomplete Data. *Biometrika*, 81(3):471–483, 1994.

R.J.A Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431), 1995.

R.J.A. Little and D.B. Rubin. Comment on Adjusting for Non-Ignorable Drop-out Using Semiparametric Models by D.O. Scharfstein, A. Rotnitzky and J.M. Robins. *Journal of the American Statistical Association*, 94:1130–1132, 1999.

X. Liu, C. Waternaux, and E. Petkova. Influence of Human Immunodeficiency Virus Infection on Neurological Impairment: An Analysis of Longitudinal Binary Data with Informative Drop-Out. *Journal of the Royal Statistical Society (Series C): Applied Statistics*, 48(1): 103–115, 1999.

G. Molenberghs, M.G. Kenward, and E. Lesaffre. The Analysis of Longitudinal Ordinal Data with Nonrandom Drop-Out. *Biometrika*, 84(1):33–44, 1997.

Geert Molenberghs and Michael G. Kenward. *Missing Data in Clinical Studies*. Wiley, 2007.

E.V. Nordheim. Inference from Nonrandomly Missing Categorical Data: an Example From a Genetic Study of Turner’s Syndrome. *Journal of the American Statistical Association*, 79(388):772–780, 1984.

D.K. Pauler, S. McCoy, and C. Moinpour. Pattern Mixture Models for Longitudinal Quality of Life Studies in Advanced Stage Disease. *Statistics in Medicine*, 22(5):795–809, 2003.

Hosted by The Berkeley Electronic Press



- E.P. Pulkstenis, T.R. Ten Have, and J.R. Landis. Model for the Analysis of Binary Longitudinal Pain Data Subject to Informative Dropout Through Remedication. *Journal of the American Statistical Association*, 93(442):438–450, 1998.
- L.S. Radloff. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, 1(3):385, 1977.
- J.M. Robins and Y.A.A. Ritov. Toward a curse of dimensionality appropriate(coda) asymptotic theory for semi-parametric models. *Statistics in Medicine*, 16(3):285–319, 1997.
- J.M. Robins, A. Rotnitzky, and L.P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.
- J.M. Robins, A. Rotnitzky, and L.P. Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429), 1995.
- A. Rotnitzky, J.M. Robins, and D.O. Scharfstein. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93(444):1321–1322, 1998.
- A. Rotnitzky, D. O. Scharfstein, T-L Su, and J. M. Robins. Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics*, 57:103–113, 2001.
- J. Roy. Modeling Longitudinal Data with Nonignorable Dropouts Using a Latent Dropout Class Model. *Biometrics*, 59(4):829–836, 2003.
- J. Roy and M. J. Daniels. A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*, 64:538–545, 2008.

- D.O. Scharfstein, A. Rotnitzky, and J.M. Robins. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *Journal of the American Statistical Association*, 94(448):1096–1146, 1999.
- D.O. Scharfstein, M.J. Daniels, and J.M. Robins. Incorporating Prior Beliefs about Selection Bias into the Analysis of Randomized Trials with Missing Outcomes. *Biostatistics*, 4(4):495, 2003.
- D.O. Scharfstein, C.F. Manski, and J.C. Anthony. On the Construction of Bounds in Prospective Studies with Missing Ordinal Outcomes: Application to the Good Behavior Game Trial. *Biometrics*, 60(1):154–164, 2004.
- D.O. Scharfstein, M.E. Halloran, H. Chu, and M.J. Daniels. On estimation of vaccine efficacy using validation samples with selection bias. *Biostatistics*, 7(4):615, 2006.
- B.E. Shepherd, P.B. Gilbert, and D.V. Mehrotra. Eliciting a Counterfactual Sensitivity Parameter. *American Statistician*, 61(1):56, 2007.
- T.R. Ten Have, A.R. Kunselman, E.P. Pulkstenis, and J.R. Landis. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, 54(1):367–383, 1998.
- T.R. Ten Have, M.E. Miller, B.A. Reboussin, and M.K. James. Mixed Effects Logistic Regression Models for Longitudinal Ordinal Functional Response Data with Multiple-Cause Drop-Out from the Longitudinal Study of Aging. *Biometrics*, 56(1):279–287, 2000.
- H. Thijs, G. Molenberghs, B. Michiels, G. Verbeke, and D. Curran. Strategies to fit pattern-mixture models. *Biostatistics*, 3(2):245, 2002.
- A. A. Tsiatis. *Semiparametric theory and missing data*. Springer, New York, 2006.
- M. J. van der Laan and J.M. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. Springer, 2003.

S. Vansteelandt, E. Goetghebeur, M.G. Kenward, and G. Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16(3):953–979, 2006.

M.C. Wu and R.J. Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1):175–188, 1988.

Y. Yuan and R. J.A. Little. Mixed-effect hybrid models for longitudinal data with nonignorable drop-out. *Biometrics (in press)*, 2009.

## Tables and Figures

Table 1: Relative Risks to be Elicited

Question	Relative Risk	Drop out Rate $p$		
		$p_1$	$p_2$	$\dots$
100% confident the number is above	$r_{z,min}(p)$			
Best Guess	$r_{z,med}(p)$			
100% confident the number is below	$r_{z,max}(p)$			

Table 2: Percentiles of Relative Risks Elicited

Treatment	Percentile	Drop out Rate	
		10%	25%
Tamoxifen	Minimum	1.10	1.30
	Median	1.20	1.50
	Maximum	1.30	1.60
Placebo	Minimum	1.01	1.20
	Median	1.05	1.30
	Maximum	1.10	1.40

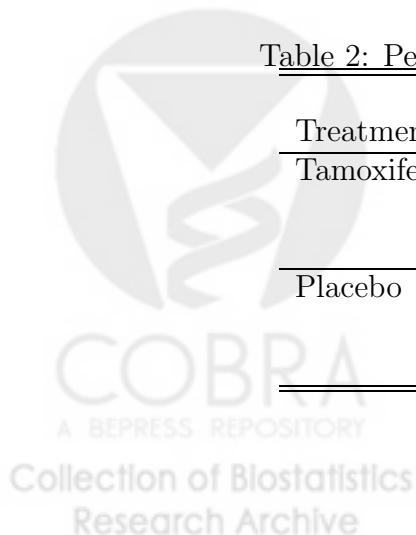


Table 3: Simulation Scenario

Parameter	Time Point							
	0	1	2	3	4	5	6	7
<i>Tamoxifen</i>								
$\alpha_0$	-2.578	-2.500	-2.613	-2.752	-2.626	-2.789	-2.811	-2.895
$\alpha_1$		2.460	1.978	1.940	2.023	2.072	1.885	2.007
$\alpha_2$			1.500	1.599	1.389	1.612	1.639	1.830
$\gamma_0$		-2.352	-2.871	-2.625	-2.513	-2.281	-2.217	-2.536
$\gamma_1$		0.611	0.397	0.460	0.247	0.320	0.127	0.228
$\gamma_2$			0.121	0.422	0.261	0.035	0.293	0.204
<i>Depression Rate</i>	0.066	0.097	0.119	0.124	0.139	0.126	0.126	0.123
<i>Placebo</i>								
$\alpha_0$	-2.653	-2.632	-2.59	-2.663	-2.598	-2.884	-2.853	-3.035
$\alpha_1$		2.708	2.304	1.874	2.104	2.068	2.123	2.243
$\alpha_2$			1.241	1.608	1.471	1.693	1.540	1.989
$\gamma_0$		-2.308	-2.970	-2.729	-2.474	-2.410	-2.460	-2.673
$\gamma_1$		0.466	0.468	0.469	0.272	0.376	0.088	0.001
$\gamma_2$			-0.293	0.323	0.278	0.288	0.241	0.428
<i>Depression Rate</i>	0.071	0.107	0.118	0.120	0.132	0.130	0.126	0.125

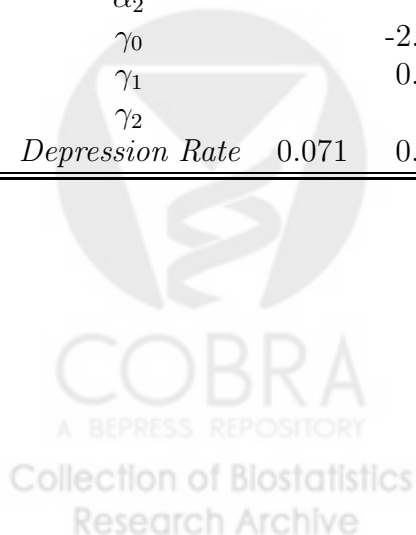


Table 4: Simulation Results: MSE ( $\times 10^3$ ). P and T represent placebo and tamoxifen arms, respectively.

Model	Treat	Observed		$\mu_{j,z}^*$						
		Y	R	1	2	3	4	5	6	7
<i>Sample Size 3000</i>										
True	P	<b>0.946</b>	<b>0.378</b>	<b>0.034</b>	<b>0.048</b>	<b>0.049</b>	<b>0.052</b>	<b>0.050</b>	<b>0.05</b>	<b>0.065</b>
	T	<b>1.075</b>	<b>0.431</b>	<b>0.026</b>	<b>0.035</b>	<b>0.046</b>	<b>0.056</b>	<b>0.052</b>	<b>0.078</b>	<b>0.061</b>
Parametric	P	30.176	0.451	0.034	0.048	0.048	0.062	0.061	0.067	0.086
	T	28.882	0.385	0.026	0.037	0.056	0.078	0.073	0.096	0.077
Shrinkage	P	6.970	1.999	0.033	0.045	0.051	0.059	0.047	0.052	0.066
	T	6.988	2.401	0.024	0.026	0.056	0.053	0.063	0.119	0.073
Saturated	P	35.678	67.171	0.036	0.050	0.054	0.058	0.101	0.231	0.561
	T	34.654	62.606	0.026	0.033	0.045	0.059	0.097	0.329	0.722
<i>Sample Size 5000</i>										
True	P	<b>0.659</b>	<b>0.238</b>	<b>0.023</b>	<b>0.027</b>	<b>0.027</b>	<b>0.028</b>	<b>0.036</b>	<b>0.043</b>	<b>0.058</b>
	T	<b>0.604</b>	<b>0.261</b>	<b>0.020</b>	<b>0.020</b>	<b>0.032</b>	<b>0.027</b>	<b>0.034</b>	<b>0.035</b>	<b>0.044</b>
Parametric	P	30.029	0.381	0.023	0.027	0.028	0.033	0.044	0.052	0.073
	T	28.571	0.277	0.020	0.023	0.042	0.042	0.049	0.056	0.057
Shrinkage	P	4.628	1.188	0.025	0.028	0.032	0.031	0.035	0.048	0.057
	T	4.448	1.414	0.017	0.023	0.028	0.026	0.038	0.033	0.044
Saturated	P	30.274	54.647	0.023	0.028	0.028	0.033	0.061	0.138	0.290
	T	29.599	51.219	0.020	0.020	0.032	0.028	0.051	0.140	0.392
<i>Sample Size 10000</i>										
True	P	<b>0.314</b>	<b>0.121</b>	<b>0.009</b>	<b>0.009</b>	<b>0.015</b>	<b>0.017</b>	<b>0.013</b>	<b>0.014</b>	<b>0.018</b>
	T	<b>0.278</b>	<b>0.126</b>	<b>0.011</b>	<b>0.013</b>	<b>0.013</b>	<b>0.013</b>	<b>0.015</b>	<b>0.019</b>	<b>0.020</b>
Parametric	P	29.849	0.315	0.009	0.010	0.016	0.023	0.028	0.033	0.043
	T	28.418	0.223	0.011	0.014	0.02	0.026	0.028	0.039	0.043
Shrinkage	P	2.392	0.707	0.008	0.010	0.015	0.017	0.014	0.013	0.014
	T	2.474	0.712	0.011	0.015	0.011	0.015	0.016	0.019	0.023
Saturated	P	22.989	37.716	0.009	0.009	0.016	0.018	0.018	0.038	0.094
	T	22.245	34.791	0.011	0.013	0.014	0.014	0.021	0.048	0.128

Table 5: Patients Cumulative Drop Out Rate

	Month	3	6	12	18	24	30	36
Tamoxifen	Available	5364	4874	4597	4249	3910	3529	3163
	Drop out	490	767	1115	1454	1835	2201	2447
	Drop Rate(%)	9.13	14.30	20.79	27.11	34.21	41.03	45.62
Placebo	Available	5375	4871	4624	4310	3951	3593	3297
	Drop out	504	751	1065	1424	1782	2078	2304
	Drop Rate(%)	9.38	13.97	19.81	26.49	33.15	38.66	42.87

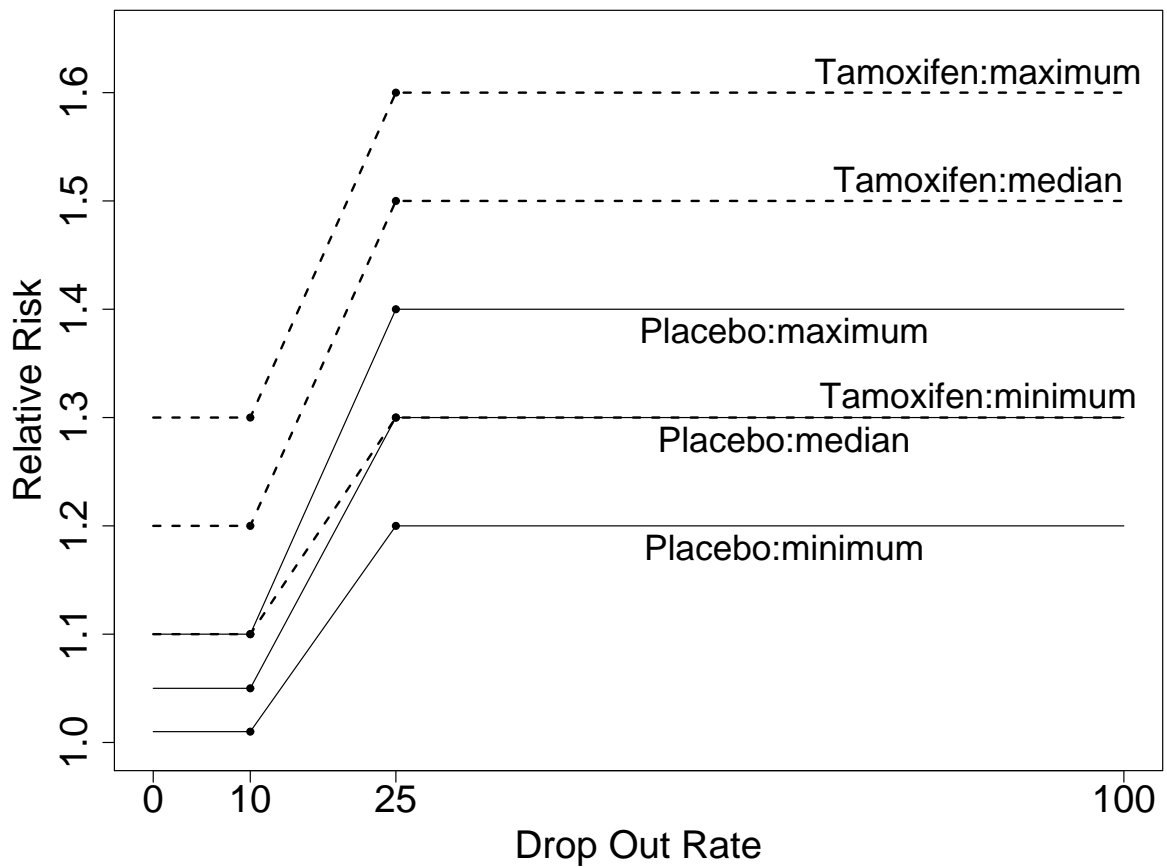


Figure 1: Extrapolation of the elicited relative risks.



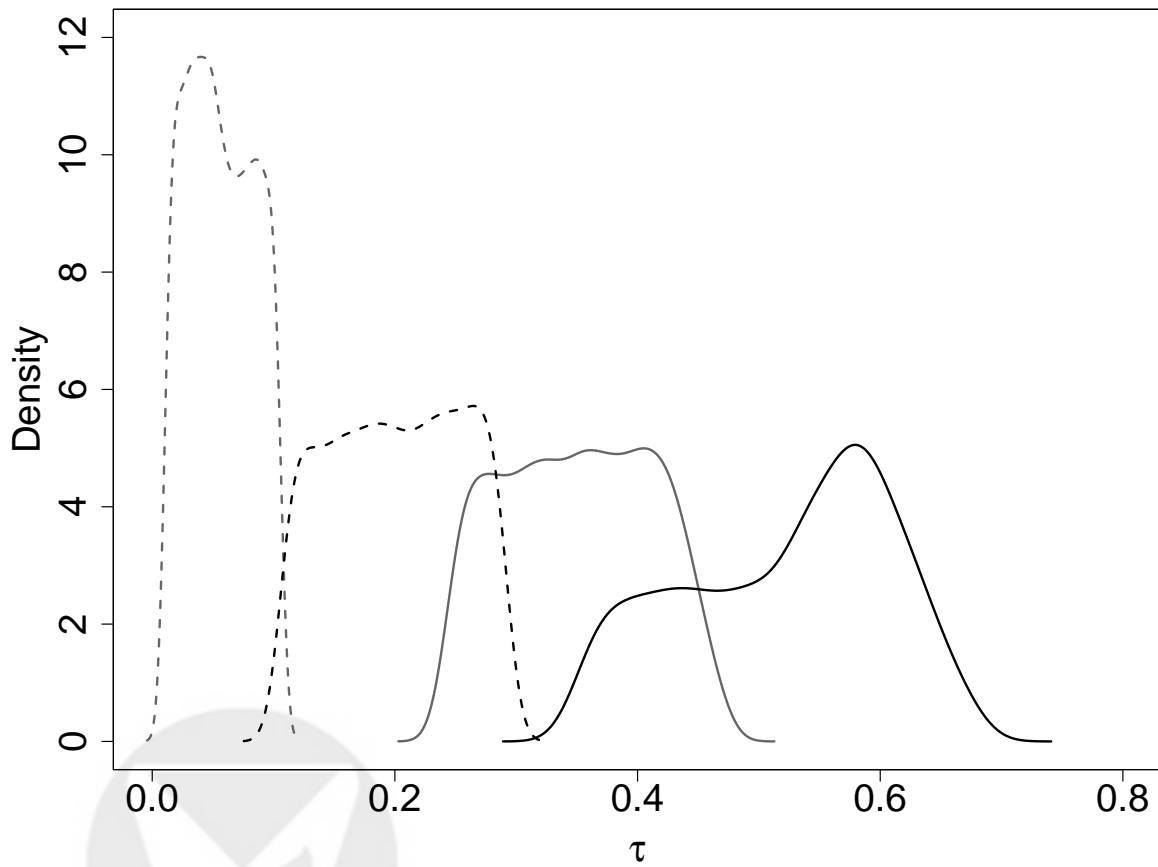


Figure 2: Prior conditional density  $\tau_{z,j,\bar{\mathbf{y}}_{j-1}}$  given  $p_{z,j}(\bar{\mathbf{y}}_{j-1})$ . Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for  $p_{z,j}(\bar{\mathbf{y}}_{j-1}) = 0.25$  and  $p_{z,j}(\bar{\mathbf{y}}_{j-1}) = 0.10$ , respectively.

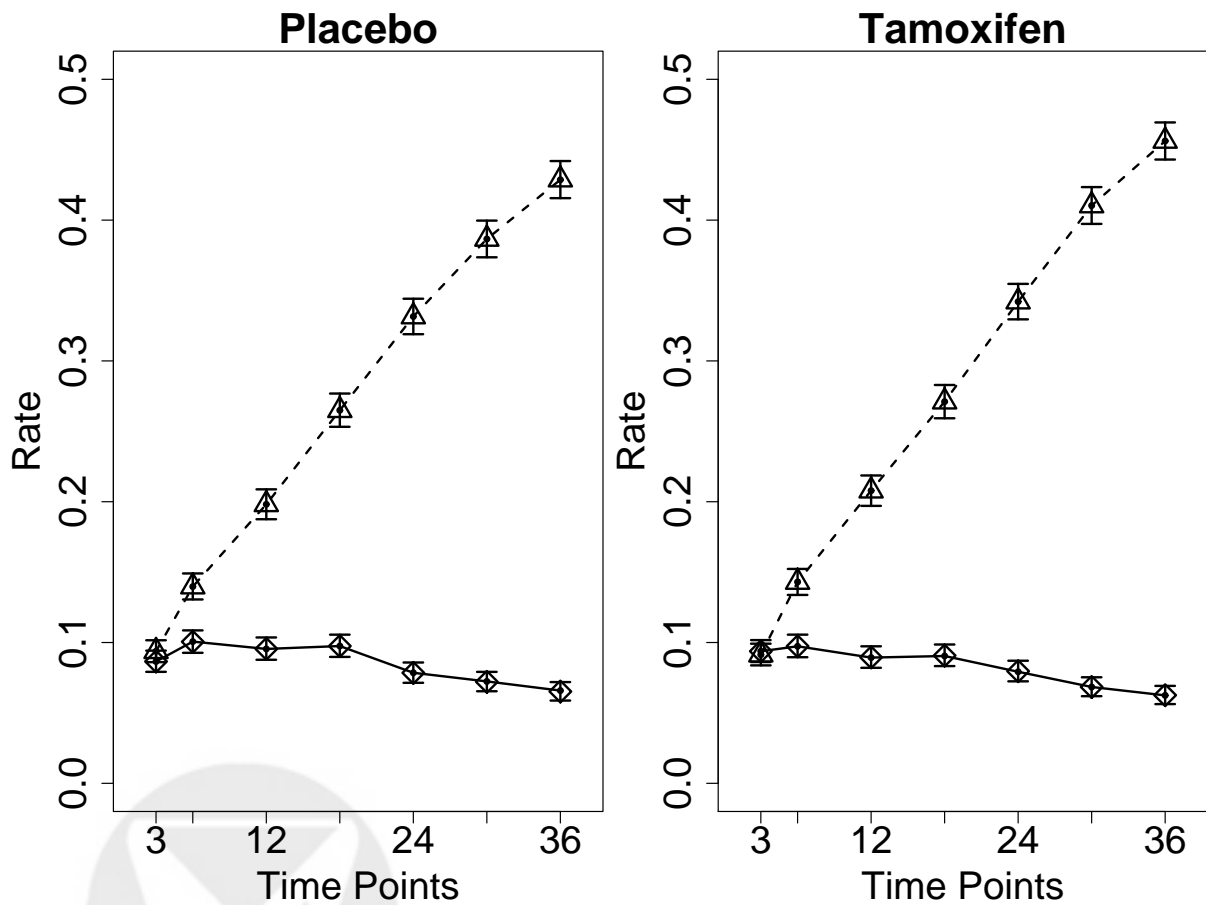
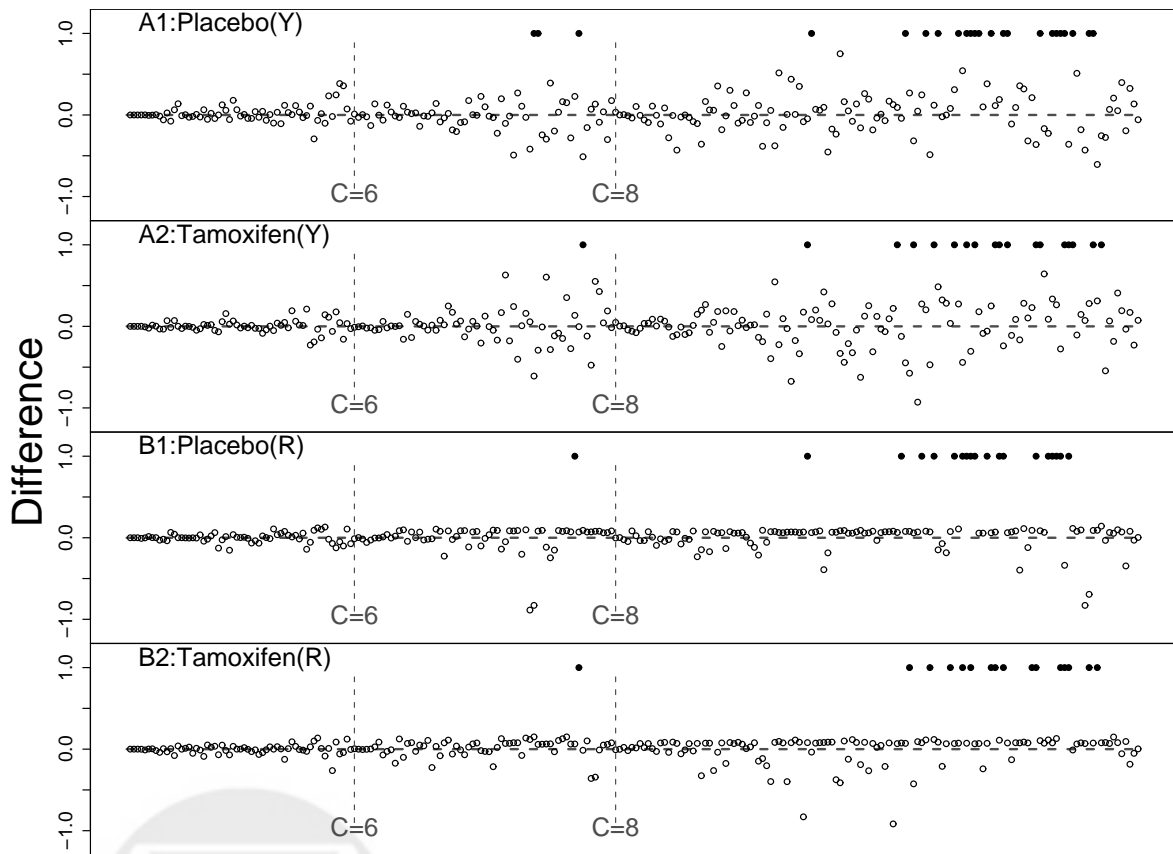


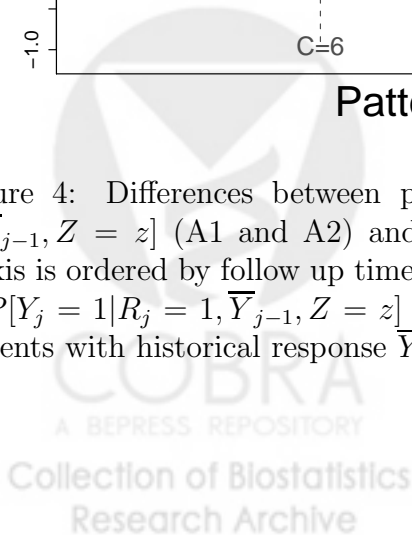
Figure 3: Solid and dashed lines represent the empirical rate of  $P[Y_j = 1, R_j = 1|Z = z]$  and  $P[R_j = 0|Z = z]$ , respectively. The posterior means of  $P[Y_j = 1, R_j = 1|Z = z]$  (diamond) and  $P[R_j = 0|Z = z]$  (triangle) and their 95% credible intervals are displayed at each time point.





Pattern of Historical Response

Figure 4: Differences between posterior mean and empirical rate of  $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$  (A1 and A2) and  $P[R_j = 0|R_{j-1} = 1, \bar{Y}_{j-1}, Z = z]$  (B1 and B2). The x-axis is ordered by follow up time  $C$  ( $\max\{t : R_t = 1\}$ ). The bullets are the posterior mean of  $P[Y_j = 1|R_j = 1, \bar{Y}_{j-1}, Z = z]$  and  $P[R_j = 0|R_{j-1} = 1, \bar{Y}_{j-1}, Z = z]$  when there are no patients with historical response  $\bar{Y}_{j-1}$ .



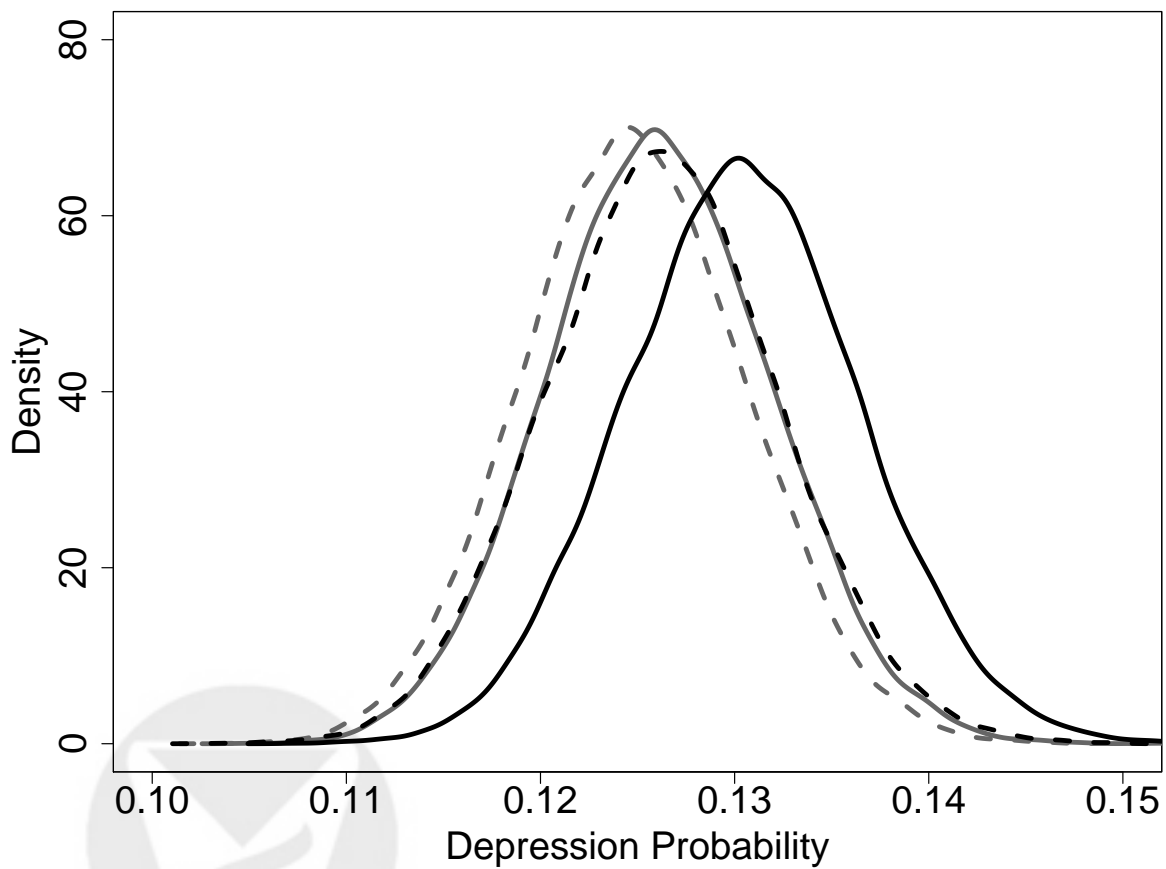


Figure 5: Posterior distribution of  $P[Y_7 = 1|Z = z]$ . Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for MNAR and MAR, respectively.

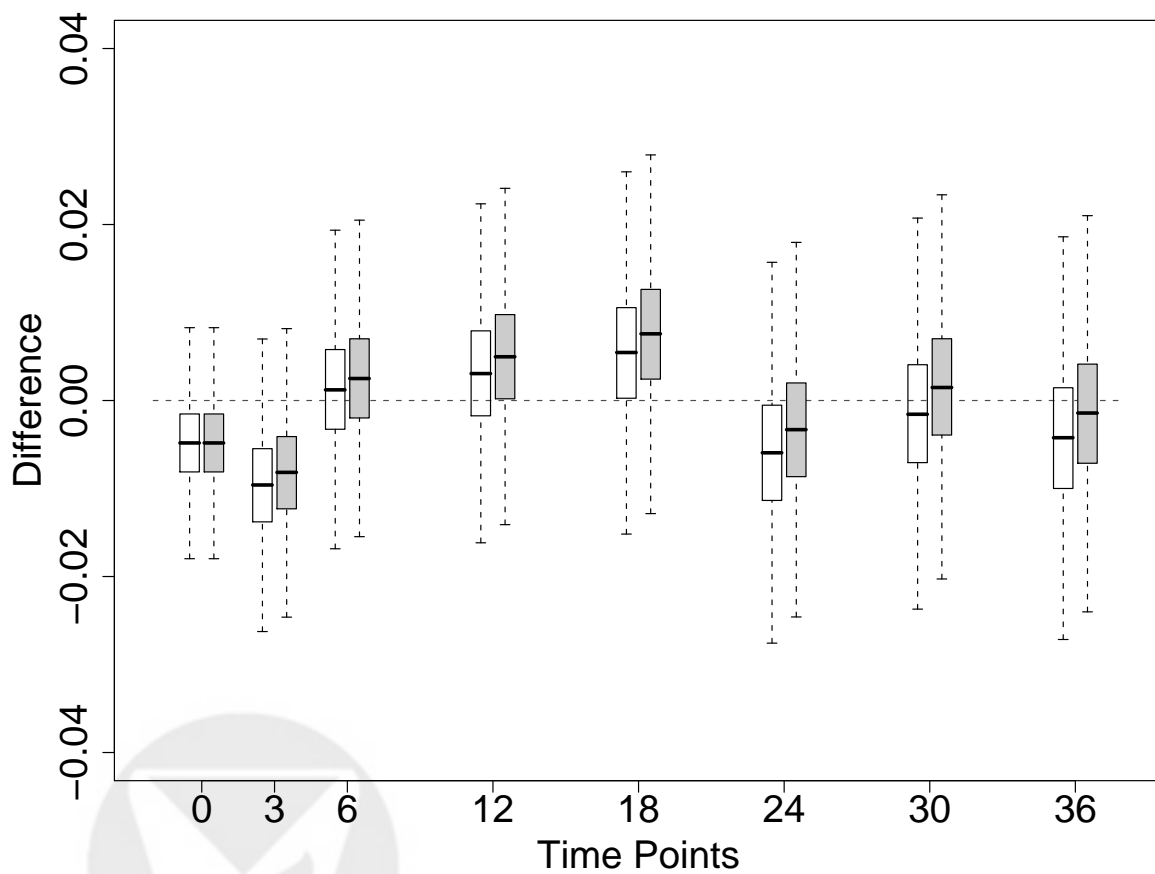


Figure 6: Posterior mean and 95% credible interval of difference of  $P[Y_j = 1|Z = z]$  between placebo and tamoxifen arms. The gray and white boxes are for MAR and MNAR, respectively.