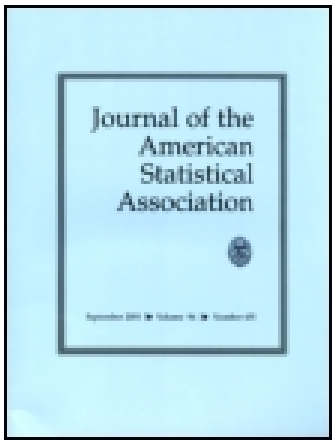

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

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

C. Wang[a], M. J. Daniels[a], D. O. Scharfstein[a] & S. Land[a]

[a] C. Wang was Ph.D. Student, Department of Statistics, University of Florida, Gainesville, FL 32611, and is Mathematical Statistician, Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD 20993. M. J. Daniels is Professor and Chair, Department of Statistics, University of Florida, Gainesville, FL 32611 . D. O. Scharfstein is Professor, Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21205. S. Land is Research Associate Professor, Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15213. This research was supported by NIH grants R01-CA85295, U10-CA37377, U10-CA69974, R01AI078835, P30MH086043, and R01HL79457. The authors are grateful to oncologist Patricia Ganz at UCLA for providing her expertise for the MNAR analysis.
Published online: 01 Jan 2012.

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

C. WANG, M. J. DANIELS, D. O. SCHARFSTEIN, and S. LAND

We consider inference in randomized longitudinal studies with missing data that is generated by skipped clinic visits and loss to followup. In this setting, it is well known that full data estimands are not identified unless unverified assumptions are imposed. We assume a nonfuture dependence model for the drop-out mechanism and partial ignorability for the intermittent missingness. We posit an exponential tilt model that links nonidentifiable distributions and distributions identified under partial ignorability. This exponential tilt model is indexed by nonidentified 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 by, and applied to, data from the Breast Cancer Prevention Trial.

KEY WORDS: Informative drop-out; Intermittent missingness; Prior elicitation.

1. INTRODUCTION

1.1 Breast Cancer Prevention Trial

The Breast Cancer Prevention Trial (BCPT) was a large multicenter, double-blinded, placebo-controlled, chemoprevention trial of the National Surgical Adjuvant Breast and Bowel Project (NSABP) designed to test the efficacy of 20 mg/day tamoxifen in preventing breast cancer and coronary heart disease (Fisher et al. 1998). The targeted population of the study was women at increased risk of developing breast cancer, including women who were age 60 or older, who were age 35–59 and had 5-year predicted risk for breast cancer at least equivalent to that of women 60 years or older, or who had a history of lobular carcinoma in situ. The study was open to accrual from June 1, 1992 through September 30, 1997 and 13,338 women were enrolled in the study during this interval. The primary objective was to determine whether long-term tamoxifen therapy was 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, 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 10,739 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 followup (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 followup, 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 followup 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 be missing not at random (MNAR); otherwise the data are said to be missing at random (MAR).

C. Wang was Ph.D. Student, Department of Statistics, University of Florida, Gainesville, FL 32611, and is Mathematical Statistician, Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD 20993. M. J. Daniels is Professor and Chair, Department of Statistics, University of Florida, Gainesville, FL 32611 (E-mail: *mdaniels@stat.ufl.edu*). D. O. Scharfstein is Professor, Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21205. S. Land is Research Associate Professor, Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15213. This research was supported by NIH grants R01-CA85295, U10-CA37377, U10-CA69974, R01AI078835, P30MH086043, and R01HL79457. The authors are grateful to oncologist Patricia Ganz at UCLA for providing her expertise for the MNAR analysis.

DOI: 10.1198/jasa.2010.ap09321

1.2 Missing Data in Longitudinal Studies

In this paper, we will concern ourselves with inference in longitudinal studies, where individuals miss visits intermittently and/or do not return for subsequent visits (i.e., drop-out). In such a setting, MNAR is often referred to as informative missingness.

1.2.1 Drop-Out. There are two main inferential paradigms for analyzing longitudinal studies with informative drop-out: likelihood (parametric) and nonlikelihood (semiparametric). 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, e.g., Heckman 1979; Diggle and Kenward 1994; Baker 1995; Fitzmaurice, Molenberghs, and Lipsitz 1995; Molenberghs, Kenward, and Lesaffre 1997; Liu, Waternaux, and Petkova 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, e.g., Little 1993, 1994, 1995; Hogan and Laird 1997b; Daniels and Hogan 2000; Fitzmaurice and Laird 2000; Birmingham and Fitzmaurice 2002; Thijs et al. 2002; Pauler, McCoy, and Moinpour 2003; Roy 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, e.g., Wu and Carroll 1988; DeGruttola and Tu 1994; Pulkstenis, Ten Have, and Landis 1998; Ten Have et al. 1998, 2000; 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, Rosenberger, and DerSimonian 1992; Little 1994; Little and Rubin 1999; Kurland and Heagerty 2004; Daniels and Hogan 2008) have promoted the use of global sensitivity analysis, whereby nonidentified 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.

Nonlikelihood 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 nonparametrically or semiparametrically and the conditional distribution of the drop-out process given the outcome process is modeled semiparametrically or fully parametrically. In the case where the drop-out process is assumed to depend only on observable outcomes (i.e., MAR), Robins, Rotnitzky, and Zhao (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, Robins, and Scharfstein (1998), Scharfstein, Rotnitzky, and Robins (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 (see, e.g., Scharfstein, Manski, and Anthony 2004). An alternative and very natural strategy is to specify an informative prior distribution on the nonidentified 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, Daniels, and Robins (2003) adopted this approach from a semiparametric selection modeling perspective. Kaciroti et al. (2009) proposed a parametric pattern-mixture model for cross-sectional, clustered binary outcomes. Lee, Hogan, and Hitsman (2008) introduced a fully-parametric pattern-mixture approach in the longitudinal setting with binary outcomes. In this paper, we consider a similar setting to Lee, Hogan, and Hitsman (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.

1.2.2 Intermittent Missing Data. In the BCPT, approximately 15% of the responses were intermittently missing, that is, there are missing values prior to drop-out. One approach to handle intermittent missingness is to consider a "monotonized" dataset, whereby all CES-D scores observed on an individual after their first missing score are deleted, as in Land et al. (2002). However, this increases the "drop-out" rate, loses efficiency, and may introduce bias.

Handling informative intermittent missing data is methodologically and computationally challenging and, as a result, the statistics literature is relatively limited. Most methods adopt a likelihood approach and rely on strong parametric assumptions (see, e.g., Troxel, Harrington, and Lipsitz 1998; Albert 2000; Ibrahim, Chen, and Lipsitz 2001; Albert et al. 2002; Lin, McCulloch, and Rosenheck 2004). Semiparametric methods have been proposed by Troxel, Lipsitz, and Harrington (1998) and Vansteelandt, Rotnitzky, and Robins (2007). Troxel, Lipsitz, and Harrington (1998) proposed a marginal model and introduced a pseudo-likelihood estimation procedure. Vansteelandt, Rotnitzky, and Robins (2007) extended the ideas of Rotnitzky, Robins, and Scharfstein (1998), Scharfstein, Rotnitzky, and Robins (1999), and Rotnitzky et al. (2001) to nonmonotone missing data.

Most related to our approach are the (partial ignorability) assumptions proposed in Harel and Schafer (2009) that partition the missing data and allow one (or more) of the partitions to be ignored given the other partition(s) and the observed data. In this paper, we apply a partial ignorability assumption such that the intermittent missing data mechanism can be ignored given drop-out and treatment strata.

1.3 Outline

The paper is organized as follows. In Section 2, we describe the data structure, formalize identification assumptions and prove that the treatment-specific distribution of the full trajectory of longitudinal outcomes is identified under these assumptions. In Section 3, we introduce a saturated model for the distribution of the data that would be observed when there is drop-out, but no intermittent observations. We then illustrate how to apply shrinkage priors to parameters in the saturated model to reduce the dimensionality of the parameter space and how to elicit (conditional) informative priors for nonidentified sensitivity parameters from experts. In Section 4, we assess, by simulation, the behavior of three classes of models: parametric, saturated, and shrinkage. Our analysis of the BCPT trial is presented in Section 5. Section 6 is devoted to a summary and discussion.

2. NOTATION, ASSUMPTIONS, AND IDENTIFIABILITY

We define the following notation for a random individual. When necessary, we use the subscript i to denote data for the ith individual.

Let Z denote the treatment assignment indicator, where $Z = 1$ denotes tamoxifen and $Z = 0$ denotes placebo. Let $\mathbf{Y}$ be the complete response data vector with elements Y_j denoting the binary outcome (i.e., depression) scheduled to be measured at the jth visit [$j = 0$ (baseline), $\ldots, J$] and let $\overline{\mathbf{Y}}_j = (Y_0, \ldots, Y_j)$ denote the history of the outcome process through visit j. Let $\mathbf{R}$ be the vector of missing data indicators with the same dimension as $\mathbf{Y}$, such that $R_j = 1$ indicates Y_j is observed and $R_j = 0$ indicates Y_j is missing. Let $S = \max\{t : R_t = 1\}$ be the last visit at which an individual's depression status is recorded. If $S < J$, then we say that the individual has dropped out and S is referred to as the drop-out time. Let $\mathbf{R}_S = \{R_j : j < S\}$ be the collection of intermittent missing data indicators recorded prior to S.

We will find it useful to distinguish three sets of data for an individual: the complete data $\mathcal{C} = (Z, S, \mathbf{R}_S, \mathbf{Y})$, the full data $\mathcal{F} = (Z, S, \mathbf{R}_S, \overline{\mathbf{Y}}_S)$, and the observed data $\mathcal{O} = (Z, S, \mathbf{R}_S, \mathbf{Y}_{\text{obs}})$, where $\mathbf{Y}_{\text{obs}}$ is the subset of $\mathbf{Y}$ for which $R_j = 1$. It is useful to also define $\mathbf{Y}_{\text{mis}} = (\mathbf{Y}^{\text{I}}_{\text{mis}}, \mathbf{Y}^{\text{C}}_{\text{mis}}, \mathbf{Y}^{\text{F}}_{\text{mis}})$, where $\mathbf{Y}^{\text{I}}_{\text{mis}} = \{Y_j : R_j = 0, j < S\}$ denotes the "intermittent" missing responses, $\mathbf{Y}^{\text{C}}_{\text{mis}} = \{Y_j : j = S + 1, j \le J\}$ denotes the missing response at the time right after drop-out, and $\mathbf{Y}^{\text{F}}_{\text{mis}} = \{Y_j : S + 1 < j \le J\}$ denotes the "future" missing responses. Note that $\overline{\mathbf{Y}}_S = (\mathbf{Y}^{\text{I}}_{\text{mis}}, \mathbf{Y}_{\text{obs}})$.

We assume that individuals are drawn as a simple random sample from a super-population so that we have an iid data structure for $\mathcal{C}$, $\mathcal{F}$, and $\mathcal{O}$. We let the parameters $\boldsymbol{\theta}_z$ index a model for the joint conditional distribution of S and $\overline{\mathbf{Y}}_S$ given $Z = z$ and the parameters $\boldsymbol{\phi}_{s,z}$ index a model for the conditional distribution of $\mathbf{R}_S$ given $S = s$, $\overline{\mathbf{Y}}_s$, and $Z = z$. We assume that the parameters $\boldsymbol{\theta}_z$ and $\boldsymbol{\phi}_z = (\boldsymbol{\phi}_{1,z}, \ldots, \boldsymbol{\phi}_{J,z})$ are distinct.

Our goal is to draw inference about $\mu^*_{z,j} = P[Y_j = 1 | Z = z]$ for $j = 1, \ldots, J$ and $z = 0, 1$. To identify $\mu^*_{z,j}$ from the distribution of the observed data, we make the following three (untestable) assumptions:

Assumption 1. Given Z and S, the intermittent missing data are missing at random, that is,

$$\mathbf{R}_S \perp \mathbf{Y}^{\text{I}}_{\text{mis}} | Z, S, \mathbf{Y}_{\text{obs}}.$$

This assumption plus the assumption that $\boldsymbol{\theta}_z$ is a priori independent of $\boldsymbol{\phi}_z$ implies that the intermittent missingness mechanism is ancillary or ignorable. Specifically, this means that when considering inferences about $\boldsymbol{\theta}_z$ from a likelihood perspective, as we are in this paper, the conditional distribution of $\mathbf{R}_S$ given Z, S, and $\mathbf{Y}_{\text{obs}}$ does not contribute to the likelihood and can be ignored (Harel and Schafer 2009).

The next two assumptions show how we can identify $\mu^*_{z,j}$ from the conditional distribution of S and $\overline{\mathbf{Y}}_S$ given $Z = z$.

Assumption 2 (Non-future dependence). For $j = 1, \ldots, J$,

$$P[S = j - 1 | S \ge j - 1, \mathbf{Y}] = P[S = j - 1 | S \ge j - 1, \overline{\mathbf{Y}}_j].$$

This assumption asserts that for individuals at risk for drop-out at visit j and who share the same history (possibly missing) 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, Molenberghs, and Thijs 2003).

Assumption 3 (Pattern-mixture representation). For $j = 1, \ldots, J$ and $y_j = 0, 1$,

$$P[Y_j = y_j | S = j - 1, \overline{\mathbf{Y}}_{j-1}, Z = z] = \frac{P[Y_j = y_j | S \ge j, \overline{\mathbf{Y}}_{j-1}, Z = z] \exp\{q_{z,j}(\overline{\mathbf{Y}}_{j-1}, y_j)\}}{E\big[\exp\{q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)\} | S \ge j, \overline{\mathbf{Y}}_{j-1}, Z = z\big]},$$

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

Assumption 3 links the nonidentified conditional distribution of Y_j given $S = j - 1$, $\overline{\mathbf{Y}}_{j-1}$, and $Z = z$ to the conditional distribution of Y_j given $S \ge j$, $\overline{\mathbf{Y}}_{j-1}$, and $Z = z$, which is estimable by Assumption 1, using exponential tilting via the specified function $q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)$. Assumption 3 has a selection model representation that is obtained using Bayes' rule.

Assumption 3 (Selection model representation). For $j = 1, \ldots, J$,

$$\text{logit}\{P[S = j - 1 | S \ge j - 1, \overline{\mathbf{Y}}_j, Z = z]\} = h_{z,j}(\overline{\mathbf{Y}}_{j-1}) + q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j),$$

where

$$h_{z,j}(\overline{\mathbf{Y}}_{j-1}) = \text{logit}\, P[S = j - 1 | S \ge j - 1, \overline{\mathbf{Y}}_{j-1}, Z = z] - \log\big\{E\big[\exp\{q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)\} | S \ge j - 1, \overline{\mathbf{Y}}_{j-1}, Z = z\big]\big\}.$$

With this characterization, we see that the function $q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)$ 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. Furthermore, the functions $q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)$ are not identifiable from the distribution of the observed data and their specification places no restrictions on the distribution of the observed data.

The above three assumptions nonparametrically, just-identify $\mu^*_{z,j}$ for all $j = 1, \ldots, J$, and $z = 0, 1$. To see this, consider the

following representation for $\mu^*_{z,j}$ derived using the laws of total and conditional probability:

$$\begin{aligned}\mu^*_{z,j} &= \sum_{\overline{\mathbf{y}}_{j-1}} P[Y_j = 1 | S \geq j, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] \\ &\quad \times \left\{ \prod_{l=1}^{j} P[S \geq l | S \geq l-1, \overline{\mathbf{Y}}_{l-1} = \overline{\mathbf{y}}_{l-1}, Z = z] \right. \\ &\quad \left. \times \prod_{l=0}^{j-1} P[Y_l = y_l | S \geq l, \overline{\mathbf{Y}}_{l-1} = \overline{\mathbf{y}}_{l-1}, Z = z] \right\} \\ &\quad + \sum_{k=1}^{j} \sum_{\overline{\mathbf{y}}_{k-1}} P[Y_j = 1 | S = k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\ &\quad \times P[S = k-1 | S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\ &\quad \times \left\{ \prod_{l=1}^{k-1} P[S \geq l | S \geq l-1, \overline{\mathbf{Y}}_{l-1} = \overline{\mathbf{y}}_{l-1}, Z = z] \right. \\ &\quad \left. \times \prod_{l=0}^{k-1} P[Y_l = y_l | S \geq l, \overline{\mathbf{Y}}_{l-1} = \overline{\mathbf{y}}_{l-1}, Z = z] \right\}.\end{aligned}$$

All quantities on the right-hand side of this equation are estimable under Assumption 1, except $P[Y_j = 1 | S = k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z]$ for $k = 1, \ldots, j$. Under Assumptions 1, 2, and 3, these latter probabilities can be shown to be identified, implying that $\mu^*_{z,j}$ is identified for all j and z.

Theorem 1. $P[Y_j = 1 | S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z]$ and $P[Y_j = 1 | S = k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z]$ are identified for $k = 1, \ldots, j$.

Proof. See Appendix.

The identifiability result shows that, given the functions $q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)$, $\mu^*_{z,j}$ can be expressed as a functional of the distribution of the observed data.

3. MODELING, PRIOR SPECIFICATION, AND POSTERIOR COMPUTATION

3.1 Modeling

We specify saturated models for the conditional distribution of S and $\overline{\mathbf{Y}}_S$ given $Z = z$ via models for $P[Y_j = 1 | S \geq j, \overline{\mathbf{Y}}_{j-1}, Z = z]$ $(j = 0, \ldots, J)$ and for $P[S = j-1 | S \geq j-1, \overline{\mathbf{Y}}_{j-1}, Z]$ $(j = 1, \ldots, J)$. We parameterize these models as follows:

$$\begin{aligned}&P[Y_0 = 1] = \alpha_{z,0}, \\ &P[Y_1 = 1 | S \geq 1, Y_0 = y, Z = z] = \alpha_{z,1,y}, \\ &P[Y_j = 1 | S \geq j, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}, Z = z] = \alpha_{z,j,\overline{\mathbf{y}}_{j-2},y}, \\ &P[S = 0 | Y_0 = y, Z = z] = \gamma_{z,0,y}, \\ &P[S = j-1 | S \geq j-1, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}, Z = z] \\ &\quad = \gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y}\end{aligned} \tag{1}$$

for $j = 2, \ldots, J$ and $y = 0, 1$. Let $\boldsymbol{\alpha}_z$ denote the parameters indexing the first set of models for response and $\boldsymbol{\gamma}_z$ denote the parameters indexing the second set of models for drop-out. Recall that we defined $\boldsymbol{\theta}_z$ to denote the parameters of the conditional distribution of S and $\overline{\mathbf{Y}}_S$ given $Z = z$; thus, $\boldsymbol{\theta}_z = (\boldsymbol{\alpha}_z, \boldsymbol{\gamma}_z)$.

Furthermore, we parameterize the functions $q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)$ with parameters $\tau_{z,j,\overline{\mathbf{y}}_{j-1}} = q_{z,j}(\overline{\mathbf{y}}_{j-1}, 1) - q_{z,j}(\overline{\mathbf{y}}_{j-1}, 0)$. Here, $\exp(\tau_{z,j,\overline{\mathbf{y}}_{j-1}})$ 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 versus not depressed at visit j, but share the mental history $\overline{\mathbf{y}}_{j-1}$ through visit $j-1$. We let $\boldsymbol{\tau}_z$ denote the collection of $\tau_{z,j,\overline{\mathbf{y}}_{j-1}}$'s.

3.2 Shrinkage Prior

The saturated model proposed in Section 3.1 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 $\overline{\mathbf{y}}_{j-1}$ (i.e., "cells") which will be sparsely represented in the dataset. In the BCPT trial, about 50% of the possible realizations of $\overline{\mathbf{Y}}_7$ have less than two observations and about 15% have no observations. From a frequentist perspective, this implies that components of $\boldsymbol{\theta}_z$ 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). 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

$$\begin{aligned}\alpha_{z,j,\overline{\mathbf{y}}_{j-2},y} &\sim \text{Beta}\big(m^{(\alpha)}_{z,j,y}/\eta^{(\alpha)}_{z,j,y}, \big(1 - m^{(\alpha)}_{z,j,y}\big)/\eta^{(\alpha)}_{z,j,y}\big), \\ \gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y} &\sim \text{Beta}\big(m^{(\gamma)}_{z,j-1,y}/\eta^{(\gamma)}_{z,j-1,y}, \big(1 - m^{(\gamma)}_{z,j-1,y}\big)/\eta^{(\gamma)}_{z,j-1,y}\big)\end{aligned} \tag{2}$$

for $j = 2, \ldots, J$ and $y = 0, 1$. For $\alpha_{z,0}$, $\alpha_{z,1,y}$ and $\gamma_{z,0,y}$ for $y = 0, 1$, we assign Unif$(0, 1)$ priors. Let $\mathbf{m}^{(\alpha)}_z$ $(\mathbf{m}^{(\gamma)}_z)$ and $\boldsymbol{\eta}^{(\alpha)}_z$ $(\boldsymbol{\eta}^{(\gamma)}_z)$ denote the parameters $m^{(\alpha)}_{z,j,y}$ $(m^{(\gamma)}_{z,j-1,y})$ and $\eta^{(\alpha)}_{z,j,y}$ $(\eta^{(\gamma)}_{z,j-1,y})$, respectively.

Note that for a random variable X that follows a Beta$(m/\eta, (1-m)/\eta)$ distribution, we have

$$E[X] = m \qquad \text{and} \qquad \text{Var}[X] = m(1-m) \times \frac{\eta}{\eta + 1}.$$

For fixed m, $\text{Var}[X] \to 0$ as $\eta \to 0$, indicating shrinkage of the distribution of X toward the mean. Thus, $\eta^{(\alpha)}_{z,j,y}$ and $\eta^{(\gamma)}_{z,j-1,y}$ serve as shrinkage parameters for $\alpha_{z,j,\overline{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y}$, respectively. As the shrinkage parameters go to zero, the distribution of the probabilities $\alpha_{z,j,\overline{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y}$ are shrunk toward the mean of the probabilities that do not depend on $\overline{\mathbf{y}}_{j-2}$, namely $m^{(\alpha)}_{z,j,y}$ and $m^{(\gamma)}_{z,j-1,y}$, respectively. In essence, the model is being shrunk toward a first-order Markov model. The shrinkage priors allow "neighboring cells" to borrow information from each other and provide more precise inferences.

We specify independent Unif$(0, 1)$ priors for $m^{(\alpha)}_{z,j,y}$ and $m^{(\gamma)}_{z,j-1,y}$. For the shrinkage parameters $\eta^{(\alpha)}_{z,j,y}$ and $\eta^{(\gamma)}_{z,j-1,y}$, we

specify independent, uniform shrinkage priors (Daniels 1999) as follows:

$$\eta^{(\alpha)}_{z,j,y} \sim \frac{g(\mathbf{E}^{(\alpha)}_{z,j,y})}{(g(\mathbf{E}^{(\alpha)}_{z,j,y})\eta^{(\alpha)}_{z,j,y}+1)^2} \quad \text{and}$$
$$\eta^{(\gamma)}_{z,j-1,y} \sim \frac{g(\mathbf{E}^{(\gamma)}_{z,j-1,y})}{(g(\mathbf{E}^{(\gamma)}_{z,j-1,y})\eta^{(\gamma)}_{z,j-1,y}+1)^2}, \tag{3}$$

where:

- $g(\cdot)$ is a summary function (e.g., minimum, median, or maximum, as suggested in Christiansen and Morris 1997).
- $\mathbf{E}^{(\alpha)}_{z,j,y} = \{e^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y}$: the expected number of subjects with $S \geq j, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}, Z = z\}$.
- $\mathbf{E}^{(\gamma)}_{z,j-1,y} = \{e^{(\gamma)}_{z,j-1,y,\overline{\mathbf{y}}_{j-2}}$: the expected number of subjects with $S \geq j-1, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}, Z = z\}$.

The expected number of subjects with $S \geq j, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}, Z = z$ and with $S \geq j-1, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}, Z = z$ can be computed as

$$\begin{aligned} e^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y} &= n_z \sum_{s=j}^{J} \sum_{y_j, y_{j+1}, \ldots, y_s} P\big[S = s, Y_s = y_s, \ldots, \\ &\qquad Y_j = y_j, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2} | Z = z\big], \\ e^{(\gamma)}_{z,j-1,\overline{\mathbf{y}}_{j-2},y} &= n_z \sum_{s=j-1}^{J} \sum_{y_j, y_{j+1}, \ldots, y_s} P\big[S = s, Y_s = y_s, \ldots, \\ &\qquad Y_j = y_j, Y_{j-1} = y, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2} | Z = z\big], \end{aligned} \tag{4}$$

where the probabilities on the right-hand side of the above equations are estimable under Assumption 1.

The expected sample sizes above are used in the prior instead of the observed binomial sample sizes which are not completely determined due to the intermittent missingness. Thus, our formulation of these priors induces a small additional amount of data dependence beyond its standard dependence on the binomial sample sizes. This additional dependence impacts the median of the prior but not its diffuseness.

3.3 τ_z Given θ_z

The sensitivity parameters in Assumption 3, defined formally in Section 3.1, 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, Gilbert, and Mehrotra 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 τ_z.

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 (depressed or not depressed). Suppose that the probability that a randomly selected woman in this group drops out before visit j is p. 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 visit 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 of dropping out only depends on past history and the visit number 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 $r_{z,\min}(p)$, $r_{z,\text{med}}(p)$, and $r_{z,\max}(p)$ denote the elicited minimum, median, and maximum relative risks. Let $p^{(y)}_{z,j}(\overline{\mathbf{y}}_{j-1}) = P[S = j-1 | S \geq j-1, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Y_j = y, Z = z]$ for $y = 0, 1$. By definition,

$$r_z\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}\big) = p^{(1)}_{z,j}(\overline{\mathbf{y}}_{j-1}) / p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1}),$$
$$\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}} = \sum_{y=0}^{1} p^{(y)}_{z,j}(\overline{\mathbf{y}}_{j-1}) \pi^{(y)}_{z,j}(\overline{\mathbf{y}}_{j-1}),$$

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

$$p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1}) = \frac{\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}}{\pi^{(1)}_{z,j}(\overline{\mathbf{y}}_{j-1})(r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}) - 1) + 1}.$$

Since $\pi^{(1)}_{z,j}(\overline{\mathbf{y}}_{j-1}) \in [0,1]$, given $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$ and $r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}})$, $p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1})$ is bounded as follows:

- for $r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}) \geq 1$,
$$\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}} / r_z\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}\big) \leq p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1}) \leq \min\big\{\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}, 1\big\},$$
- for $r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}) \leq 1$,
$$\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}} \leq p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1}) \leq \min\big\{\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}} / r_z\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}\big), 1\big\}.$$

We will use these bounds to construct our prior.

We construct the conditional prior of $\tau_{z,j,\overline{\mathbf{y}}_{j-1}}$ given $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$ 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 $m \in \{\min, \text{med}, \max\}$, interpolate the elicited $r_{z,m}(p)$ at different drop-out probabilities (see Figure 1) to find $r_{z,m}(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}})$ for any $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$.

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

$$\text{Unif}\big(r_{z,\min}\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}\big), r_{z,\text{med}}\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}\big)\big)$$

and

$$\text{Unif}\big(r_{z,\text{med}}\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}\big), r_{z,\max}\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}\big)\big)$$

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

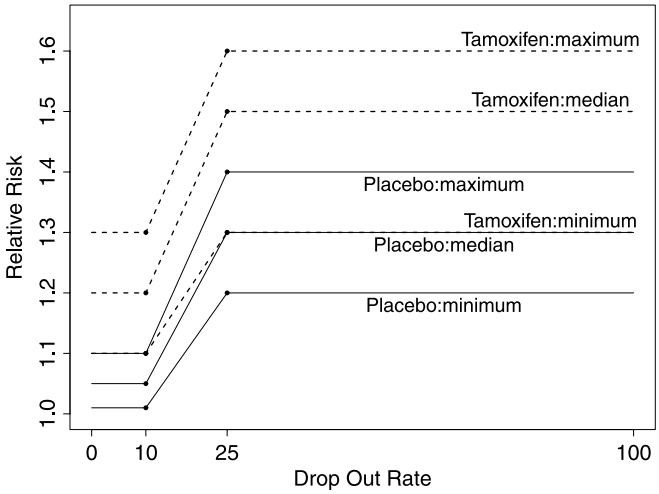

Figure 1. Extrapolation of the elicited relative risks.

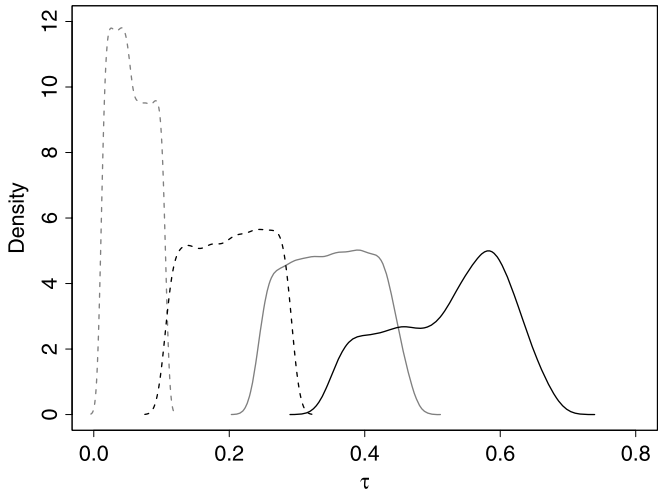

Figure 2. Prior conditional density $\tau_{z,j,\overline{\mathbf{y}}_{j-1}}$ given $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$. Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}} = 0.25$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

Step 3. Construct a conditional prior of $p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1})$ given $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$ and $r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}})$ as a uniform distribution with lower bound

$$\frac{\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}}{\max\{r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}), 1\}}$$

and upper bound

$$\min\left\{\frac{\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}}{\min\{r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}), 1\}}, \frac{1}{\max\{r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}), 1\}}\right\}.$$

The bounds were derived above.

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

$$\tau_{z,j,\overline{\mathbf{y}}_{j-1}} = \log\left(\frac{r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}})(1-p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1}))}{1-r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}})p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1})}\right),$$

that is, $\tau_{z,j,\overline{\mathbf{y}}_{j-1}}$ is a deterministic function of $r_z(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}})$ and $p^{(0)}_{z,j}(\overline{\mathbf{y}}_{j-1})$.

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

Figure 2 shows the density of $\tau_{z,j,\overline{\mathbf{y}}_{j-1}}$ given $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$ equal .10 and .25 for the tamoxifen and placebo arms. For two patients with the same depression 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 $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$, the log odds ratio is higher for patients in the tamoxifen versus placebo arms.

3.4 Posterior Computation

The following steps are used to simulate draws from the posterior of $\mu^*_{z,j}$:

1. Sample $P(\boldsymbol{\theta}_z, \mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}|\mathbf{Y}_{\mathrm{obs}}, S, \mathbf{R}_S, Z=z)$ using Gibbs sampling with data augmentation (see details in Appendix). Continue sampling until convergence.
2. For each draw of $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y_{j-1}}$, draw $\tau_{z,j,\overline{\mathbf{y}}_{j-1}}$ based on the conditional priors described in Section 3.3.
3. Compute $\mu^*_{z,j}$ by plugging the draws of $\alpha_{z,j,\overline{\mathbf{y}}_{j-2},y}$, $\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y}$, and $\tau_{z,j,\overline{\mathbf{y}}_{j-1}}$ into the identification algorithm discussed in Section 2.

4. ASSESSMENT OF MODEL PERFORMANCE VIA SIMULATION

We simulated observed data (no intermittent missingness) from a parametric model of the following form:

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

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

Table 1. Elicited percentiles of relative risks of dropping out

		Drop out rate	
Treatment	Percentile	10%	25%
Tamoxifen	Minimum (100% confident the number is above)	1.10	1.30
	Median (best guess)	1.20	1.50
	Maximum (100% confident the number is below)	1.30	1.60
Placebo	Minimum (100% confident the number is above)	1.01	1.20
	Median (best guess)	1.05	1.30
	Maximum (100% confident the number is below)	1.10	1.40

$$\begin{aligned}
&\operatorname{logit} P[Y_j = 1 | S \geq j, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] \\
&\quad = \alpha_{z,j,0} + \alpha_{z,j,1} y_{j-1} + \alpha_{z,j,2} y_{j-2}, \\
&\operatorname{logit} P[S = 0 | Y_0 = y_0, Z = z] = \gamma_{z,0,0} + \gamma_{z,0,1} y_0, \\
&\operatorname{logit} P[S = j-1 | S \geq j-1, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] \\
&\quad = \gamma_{z,j-1,0} + \gamma_{z,j-1,1} y_{j-1} + \gamma_{z,j-1,2} y_{j-2}
\end{aligned}$$

for $j = 2$ to 7. Our "true" model is a second-order Markov model.

We compared the performance of our shrinkage model with (1) a correct parametric model, (2) an incorrect parametric model (first-order Markov model), and (3) a saturated model (with diffuse priors). Our shrinkage model uses the shrinkage priors proposed in Section 3.2. Note that the shrinkage priors shrink the saturated model to an incorrect parametric model (i.e., a first-order Markov model). For the saturated model, instead of the shrinkage priors, we specify independent Unif(0, 1) on $\boldsymbol{\alpha}_z$'s and $\boldsymbol{\gamma}_z$'s.

To determine the parameters of the data generating model, we fit this model to the "monotonized" BCPT data in `WinBUGS` with noninformative priors. We used the posterior mean of the of parameters $\boldsymbol{\alpha}_z$ and $\boldsymbol{\gamma}_z$ as the true parameters. We compute the "true" values of $\mu^*_{z,j}$ by (1) drawing 10,000 values from the elicited prior of $\boldsymbol{\tau}_z$ given $\boldsymbol{\gamma}_z$ given in Table 1, (2) computing $\mu^*_{z,j}$ using the identification algorithm in Section 2 for each draw, and (3) average the resulting $\mu^*_{z,j}$'s. The model parameters and the "true" depression rates $\mu^*_{z,j}$, are given in Table 2.

We considered small (500), moderate (2000), large (5000), and very large (1,000,000) sample sizes for each treatment arm; for each sample size, we simulated 500 datasets. We assessed model performance using mean squared error (MSE).

In Table 3 (sample size 1,000,000 not shown), we report the MSE's of $P[Y_j = 1 | S \geq j, \overline{Y}_{j-1}, Z = z]$ and $P[S = j-1 | S \geq j-1, \overline{Y}_{j-1}, Z = z]$ averaged over all j and all $\overline{Y}_{j-1}$ (see columns 3 and 4, respectively). We also report the MSE's for $\mu^*_{z,j}$ (see columns 6–12). For reference, the MSE's associated with the true data generating model are bolded. At all sample sizes, the shrinkage model has lower MSE's for the rates of depression at times 3–7 than the incorrectly specified parametric model and the saturated model. Our simulation results show that as sample size goes to infinity (e.g., very large, 1,000,000), both the shrinkage model and the saturated model converge to the true values of $\mu^*_{z,j}$, whereas the incorrectly specified parametric model yields biased estimates.

In addition, the MSE's for the parameters $\mu^*_{z,j}$ in 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.

5. APPLICATION: BREAST CANCER PREVENTION TRIAL (BCPT)

Table 4 displays the treatment-specific drop-out and intermittent missing rates in the BCPT. By the 7th study visit (36 months), more than 30% of patients had dropped out in each treatment arm, with a slightly higher percentage in the tamoxifen arm.

5.1 Model Fit

We fit the shrinkage model to the observed data using `R`, with multiple chains of 5000 iterations and 1000 burn-in. Convergence was checked by examining trace plots of the multiple chains. We defined $g(\cdot)$ in the priors for the hyperparameters [Equation (3)] to be the maximum function. To compute the expected number of subjects $e^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y}$ and $e^{(\gamma)}_{z,j-1,\overline{\mathbf{y}}_{j-2},y}$ in Equation (4), we assigned a point mass prior at 0.5 to all $\mathbf{m}^{(\alpha)}_z$, $\mathbf{m}^{(\gamma)}_z$, $\boldsymbol{\eta}^{(\alpha)}_z$, and $\boldsymbol{\eta}^{(\gamma)}_z$ [which corresponds to Unif(0, 1) priors on $\alpha_{z,\overline{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,\overline{\mathbf{y}}_{j-2},y}$] and sampled $\alpha_{z,\overline{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,\overline{\mathbf{y}}_{j-2},y}$ using Step 1 in the algorithm described in Section 3.4. To avoid data sparsity, we calculated $P[S = s, \overline{\mathbf{Y}}_s = \overline{\mathbf{y}}_s]$ using the posterior mean of $\alpha_{z,\overline{\mathbf{y}}_{j-2},y}$ and $\gamma_{z,\overline{\mathbf{y}}_{j-2},y}$ rather than the empirical probabilities.

To assess model fit, we compared the empirical rates and posterior means (with 95% credible intervals) of $P[Y_j = 1, S \geq$

Table 2. Simulation scenario

	Time point j (month)							
Parameter	0 (0)	1 (3)	2 (6)	3 (12)	4 (18)	5 (24)	6 (30)	7 (36)
Tamoxifen								
$\alpha_{1,j,0}$	−2.578	−2.500	−2.613	−2.752	−2.626	−2.789	−2.811	−2.895
$\alpha_{1,j,1}$		2.460	1.978	1.940	2.023	2.072	1.885	2.007
$\alpha_{1,j,2}$			1.500	1.599	1.389	1.612	1.639	1.830
$\gamma_{1,j-1,0}$		−2.352	−2.871	−2.625	−2.513	−2.281	−2.217	−2.536
$\gamma_{1,j-1,1}$		0.611	0.397	0.460	0.247	0.320	0.127	0.228
$\gamma_{1,j-1,2}$			0.121	0.422	0.261	0.035	0.293	0.204
Depression rate	0.066	0.097	0.119	0.124	0.139	0.126	0.126	0.123
Placebo								
$\alpha_{0,j,0}$	−2.653	−2.632	−2.59	−2.663	−2.598	−2.884	−2.853	−3.035
$\alpha_{0,j,1}$		2.708	2.304	1.874	2.104	2.068	2.123	2.243
$\alpha_{0,j,2}$			1.241	1.608	1.471	1.693	1.540	1.989
$\gamma_{0,j-1,0}$		−2.308	−2.970	−2.729	−2.474	−2.410	−2.460	−2.673
$\gamma_{0,j-1,1}$		0.466	0.468	0.469	0.272	0.376	0.088	0.001
$\gamma_{0,j-1,2}$			−0.293	0.323	0.278	0.288	0.241	0.428
Depression rate	0.071	0.107	0.118	0.120	0.132	0.130	0.126	0.125

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

		Observed		$\mu^*_{j,z}$ (month)						
Model	Treat	Y	R	1 (3)	2 (6)	3 (12)	4 (18)	5 (24)	6 (30)	7 (36)
Sample size 500										
True	P	**6.209**	**2.474**	**0.199**	**0.225**	**0.258**	**0.313**	**0.319**	**0.352**	**0.390**
	T	**6.790**	**2.789**	**0.205**	**0.228**	**0.297**	**0.344**	**0.331**	**0.405**	**0.428**
Parametric	P	33.351	1.511	0.199	0.227	0.260	0.317	0.323	0.349	0.388
	T	32.323	1.602	0.205	0.226	0.292	0.345	0.333	0.403	0.425
Shrinkage	P	29.478	2.310	0.202	0.226	0.252	0.303	0.312	0.337	0.372
	T	28.410	2.365	0.212	0.232	0.294	0.336	0.330	0.390	0.419
Saturated	P	57.107	111.263	0.202	0.228	0.302	0.490	1.083	2.401	4.427
	T	55.582	104.882	0.211	0.245	0.383	0.657	1.352	3.167	5.782
Sample size 2000										
True	P	**1.474**	**0.586**	**0.052**	**0.058**	**0.063**	**0.078**	**0.081**	**0.086**	**0.097**
	T	**1.610**	**0.634**	**0.050**	**0.063**	**0.062**	**0.080**	**0.093**	**0.095**	**0.101**
Parametric	P	30.507	0.543	0.051	0.055	0.064	0.081	0.090	0.091	0.108
	T	29.168	0.495	0.050	0.064	0.071	0.086	0.101	0.110	0.121
Shrinkage	P	23.545	0.647	0.053	0.056	0.063	0.078	0.082	0.084	0.095
	T	22.598	0.615	0.050	0.063	0.063	0.080	0.093	0.095	0.102
Saturated	P	40.322	77.627	0.053	0.057	0.069	0.100	0.188	0.457	0.946
	T	38.943	72.731	0.050	0.064	0.067	0.110	0.218	0.560	1.223
Sample size 5000										
True	P	**0.594**	**0.234**	**0.020**	**0.024**	**0.026**	**0.033**	**0.031**	**0.040**	**0.036**
	T	**0.623**	**0.265**	**0.024**	**0.024**	**0.028**	**0.035**	**0.033**	**0.039**	**0.040**
Parametric	P	29.983	0.379	0.020	0.025	0.029	0.037	0.043	0.049	0.055
	T	28.616	0.298	0.024	0.025	0.035	0.048	0.045	0.060	0.059
Shrinkage	P	18.830	0.394	0.020	0.024	0.026	0.033	0.031	0.039	0.036
	T	18.055	0.322	0.024	0.024	0.028	0.036	0.034	0.040	0.041
Saturated	P	30.071	54.454	0.020	0.024	0.027	0.038	0.052	0.130	0.270
	T	29.156	50.590	0.024	0.024	0.029	0.039	0.059	0.148	0.373

$j|Z = z]$ and $P[S < j|Z = z]$. As shown in Figure 3, the shrinkage model fits the observed data well.

Figure 4 illustrates the effect of shrinkage on the model fit by comparing the difference between the empirical rates and posterior means of $P[Y_j = 1|S \geq j, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z]$ for the tamoxifen arm ($Z = 1$) and $j = 6, 7$. We use the later time points to illustrate this since the observed data were more sparse and the shrinkage effect was more apparent. The empirical depression rates often reside on the boundary (0 or 1). In some cases, there are no observations within "cells," thus the empirical rates were undefined. From the simulation results in Section 4, we know that the empirical estimates are less reliable for later time points. Via the shrinkage priors, the probabilities $P[Y_j = 1|S \geq j, Y_{j-1} = y_{j-1}, \overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}, Z = z]$ with the same $\overline{\mathbf{y}}_{j-2}$ are shrunk together and away from the boundaries. By borrowing information across neighboring cells, we are able to estimate $P[Y_j = 1|S \geq j, \overline{\mathbf{Y}}_{j-1}, Z = z]$ for all j, z, and $\overline{\mathbf{Y}}_{j-1}$ with better precision. The differences between the empirical rates and the posterior means illustrate the magnitude of the shrinkage effect. In the BCPT, the depression rate was (relatively) low

Table 4. Missingness by scheduled measurement time

	Time point j (month)						
	1 (3)	2 (6)	3 (12)	4 (18)	5 (24)	6 (30)	7 (36)
Tamoxifen (total $N = 5364$, overall missing 34.94%)							
Intermittent missing	330	224	190	200	203	195	
Drop-out at j	160	122	259	280	332	352	369
Cumulative drop-out	160	282	541	821	1153	1505	1874
Placebo (total $N = 5375$, overall missing 31.83%)							
Intermittent missing	347	215	153	181	199	197	
Drop-out at j	157	106	247	287	309	272	333
Cumulative drop-out	157	263	510	797	1106	1378	1711

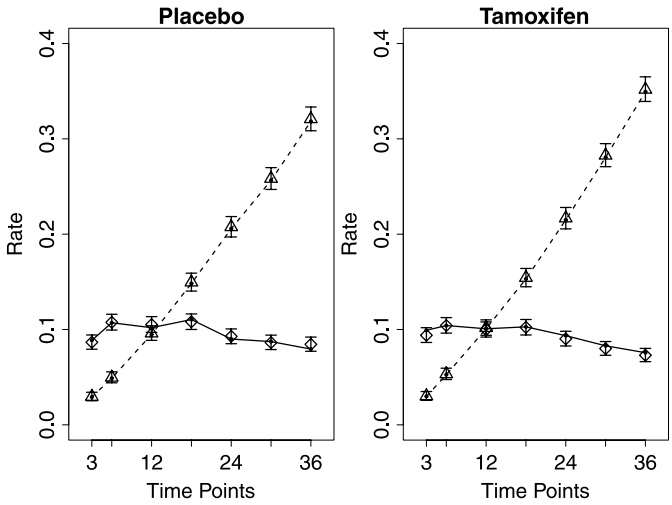

Figure 3. Solid and dashed lines represent the empirical rate of $P[Y_j = 1, S \geq j | Z = z]$ and $P[S < j | Z = z]$, respectively. The posterior means of $P[Y_j = 1, S \geq j | Z = z]$ (diamond) and $P[S < j | Z = z]$ (triangle) and their 95% credible intervals are displayed at each time point.

and there were few subjects at the later times that were observed with a history of mostly depression at the earlier visits; as a result, the differences were larger when $\overline{\mathbf{Y}}_{j-1}$ had a lot of 1's (depression).

5.2 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 τ at zero) is also presented (dashed lines). The observed depression rates (i.e., complete case analysis) were 0.124 and 0.112 for the placebo and tamoxifen arms, respectively. Under the MNAR analysis (using the elicited priors), the posterior mean of the depression rates at month 36 were 0.133 (95% CI: 0.122, 0.144) and 0.125 (95% CI: 0.114, 0.136) for the placebo and tamoxifen arms; the difference was -0.007 (95% CI: $-0.023, 0.008$). Under MAR, the rates were 0.132 (95% CI: 0.121, 0.143) and 0.122 (95% CI: 0.111, 0.133) for the placebo and tamoxifen arms; the difference was -0.01 (95% CI: $-0.025, 0.005$). The

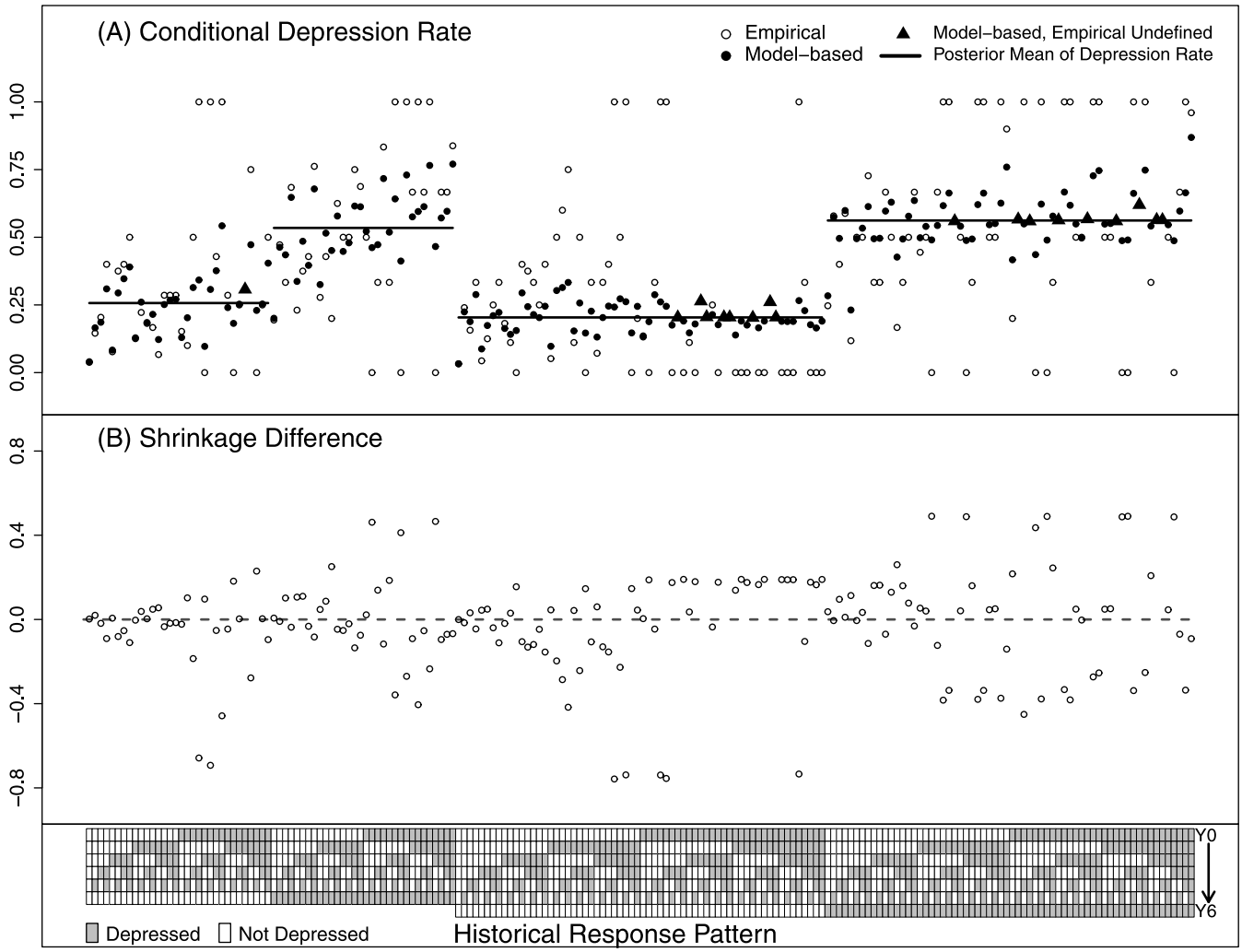

Figure 4. (A) The empirical rate and model-based posterior mean of $P[Y_j = 1 | S \geq j, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z]$ for $Z = 2$ and $j = 6, 7$. (B) The difference between the empirical and model-based posterior mean of the depression rate. The x-axis is the pattern of historical response data $\overline{\mathbf{Y}}_{j-1}$.

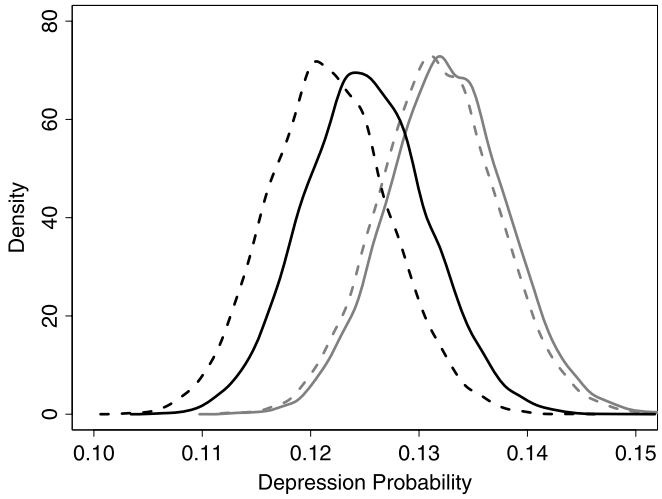

Figure 5. Posterior distribution of $P[Y_7 = 1|Z = z]$. Black and gray lines represent tamoxifen and placebo arms, respectively. Solid and dashed lines are for MNAR and MAR, respectively.

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 1), 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 any measurement time (95% credible intervals all cover zero) under both MNAR and MAR. Similar (nonsignificant) treatment differences were seen when examining treatment comparisons conditional on depression status at baseline.

5.3 Sensitivity of Inference to the Priors

To assess the sensitivity of inference on the 36-month depression rates to the elicited (informative) priors $\{r_{\min}, r_{\text{med}}, r_{\max}\}$, we considered several alternative scenarios based on Table 1. In the first scenario, we made the priors more or less informative by scaling the range, but leaving the median unchanged. That is, we considered increasing (or decreasing) the range by a scale factor v to $\{r_{\text{med}} - v(r_{\text{med}} - r_{\min}), r_{\text{med}}, r_{\text{med}} + v(r_{\max} - r_{\text{med}})\}$. In the second scenario, we shifted the prior by a factor u, $\{u + r_{\min}, u + r_{\text{med}}, u + r_{\max}\}$.

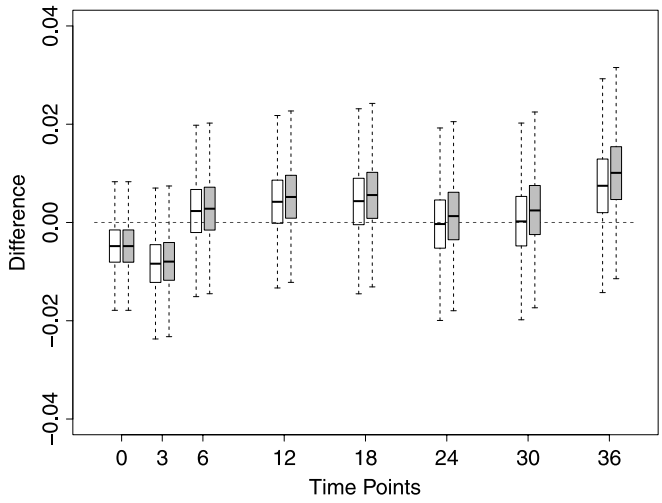

Figure 6. Posterior mean and 95% credible interval of difference of $P[Y_j = 1|Z = z]$ between placebo and tamoxifen arms. The gray and white boxes are for MAR and MNAR, respectively.

The posterior mean and between-treatment difference of the depression rate at month 36 with 95% CI are given in Tables 5 and 6. None of the scenarios considered resulted in the 95% CI for the difference in rates of depression at 36 months that excluded zero except for the (extreme) scenario where the elicited tamoxifen intervals were shifted by 0.5 and the elicited placebo intervals were shifted by -0.5.

We also assessed the impact of switching the priors for the placebo and tamoxifen arms; in this case, the posterior means were 0.135 (95% CI: 0.124, 0.146) and 0.123 (95% CI: 0.112, 0.134) for the placebo and tamoxifen arms respectively, while the difference was -0.012 (95% CI: -0.027, 0.004).

6. SUMMARY AND DISCUSSION

In this paper, we have presented a Bayesian shrinkage approach for longitudinal binary data with informative missingness (both intermittent and drop-out). Our model provides a framework that incorporates expert opinion about nonidentifiable 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 nonidentifiable parameters can be implemented by separate independent data analysts. This feature can be used to increase the objectivity of necessarily subjective inferences.

Penalized likelihood (Wahba 1990; Green and Silverman 1994; Fan and Li 2001) is another approach for high-dimensional statistical modeling. There are similarities between the penalized likelihood approach and our shrinkage model. In fact, the shrinkage priors on the saturated model parameters proposed in our approach can be viewed as a specific form for the penalty.

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 drop-out; 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, withdrawing consent and institution-specific causes such as understaffing and staff turnover. Therefore, some missingness is less likely to be informative. In addition, institutional differences might be addressed by allowing institution-specific parameters with priors that shrink them toward a common set of parameters. Finally, we might consider alternatives to the partial ignorability assumption (Assumption 1) which has been widely used, but questioned by some (Robins 1997).

APPENDIX

Proof of Theorem 1

Under Assumption 1, we know that the parameters of the conditional joint distribution of S and $\overline{\mathbf{Y}}_S$ given $Z = z$ are estimable

Table 5. Sensitivity to the elicited prior

		Scenario (T: tamoxifen, P: placebo)							
		$v^T = 5, v^P = 5$		$v^T = 0.2, v^P = 0.2$		$u^T = 0.5, u^P = 0.5$		$u^T = -0.5, u^P = -0.5$	
Treatment	Percentile	10%	25%	10%	25%	10%	25%	10%	25%
Tamoxifen	Minimum	0.79	0.50	1.18	1.46	1.60	1.80	0.60	0.80
	Median	1.20	1.50	1.20	1.50	1.70	2.00	0.70	1.00
	Maximum	1.70	2.00	1.22	1.52	1.80	2.10	0.80	1.10
	$P[Y_7 = 1]$ (95% CI)	0.125 (0.114, 0.136)		0.125 (0.114, 0.136)		0.132 (0.120, 0.143)		0.117 (0.107, 0.128)	
Placebo	Minimum	0.85	0.80	1.04	1.28	1.51	1.70	0.51	0.70
	Median	1.05	1.30	1.05	1.30	1.55	1.80	0.55	0.80
	Maximum	1.30	1.80	1.06	1.32	1.60	1.90	0.60	0.90
	$P[Y_7 = 1]$ (95% CI)	0.133 (0.122, 0.144)		0.133 (0.122, 0.144)		0.139 (0.128, 0.150)		0.125 (0.114, 0.135)	
Difference of $P[Y_7 = 1]$ (95% CI)		−0.008 (−0.024, 0.008)		−0.007 (−0.023, 0.008)		−0.007 (−0.023, 0.009)		−0.008 (−0.023, 0.007)	

from the distribution of the observed data. Next, we show, via backward induction, that $P[Y_j = 1|S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z]$ and $P[Y_j = 1|S = k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z]$ $(k = 1, \ldots, j)$ can be written in terms of the the conditional joint distribution of S and $\overline{\mathbf{Y}}_S$ given $Z = z$.

Consider $k = j$. By Assumption 2,

$$P[Y_j = 1|S = j-1, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] = \frac{P[Y_j = 1|S \geq j, \overline{\mathbf{Y}}_{j-1}, Z = z] \exp\{q_{z,j}(\overline{\mathbf{Y}}_{j-1}, 1)\}}{E[\exp\{q_{z,j}(\overline{\mathbf{Y}}_{j-1}, Y_j)\}|S \geq j, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z]}.$$

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

$$\begin{aligned} &P[Y_j = 1|S \geq j-1, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] \\ &= P[Y_j = 1|S \geq j, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] \\ &\quad \times P[S \geq j|S \geq j-1, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] \\ &\quad + P[Y_j = 1|S = j-1, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z] \\ &\quad \times P[S = j-1|S \geq j-1, \overline{\mathbf{Y}}_{j-1} = \overline{\mathbf{y}}_{j-1}, Z = z]. \end{aligned}$$

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

Suppose that $P[Y_j = 1|S = k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z]$ and $P[Y_j = 1|S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z]$ 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

$$\begin{aligned} &P[Y_j = 1|S = k'-1, \overline{\mathbf{Y}}_{k'-1} = \overline{\mathbf{y}}_{k'-1}, Z = z] \\ &= P[Y_j = 1|S = k-2, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\ &= \sum_{y_{k-1}=0}^{1} P[Y_j = 1|S = k-2, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\ &\quad \times P[Y_{k-1} = y_{k-1}|S = k-2, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\ &= \sum_{y_{k-1}=0}^{1} P[Y_j = 1|S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\ &\quad \times \big(P[Y_{k-1} = y_{k-1}|S \geq k-1, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\ &\quad \times \exp\{q_{z,k-1}(\overline{\mathbf{Y}}_{k-2}, y_{k-1})\} \\ &\quad /E\big[\exp\{q_{z,k-1}(\overline{\mathbf{Y}}_{k-2}, Y_{k-1})\}|S \geq k-1, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z\big]\big). \end{aligned}$$

The third equality follows by Assumptions 2 and 3. Since all the quantities on the right-hand side of the last equality are identified, $P[Y_j = 1|S = k'-1, \overline{\mathbf{Y}}_{k'-1} = \overline{\mathbf{y}}_{k'-1}, Z = z]$ is identified. Further,

$$\begin{aligned} &P[Y_j = 1|S \geq k'-1, \overline{\mathbf{Y}}_{k'-1} = \overline{\mathbf{y}}_{k'-1}, Z = z] \\ &= P[Y_j = 1|S \geq k-2, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \end{aligned}$$

Table 6. Sensitivity to the elicited prior

		Scenario (T: tamoxifen, P: placebo)							
		$v^T = 5, v^P = 0.2$		$v^T = 0.2, v^P = 5$		$u^T = 0.5, u^P = -0.5$		$u^T = -0.5, u^P = -0.5$	
Treatment	Percentile	10%	25%	10%	25%	10%	25%	10%	25%
Tamoxifen	Minimum	0.79	0.50	1.18	1.46	1.60	1.80	0.60	0.80
	Median	1.20	1.50	1.20	1.50	1.70	2.00	0.70	1.00
	Maximum	1.70	2.00	1.22	1.52	1.80	2.10	0.80	1.10
	$P[Y_7 = 1]$ (95% CI)	0.125 (0.114, 0.136)		0.125 (0.114, 0.136)		0.132 (0.121, 0.143)		0.117 (0.107, 0.128)	
Placebo	Minimum	1.04	1.28	0.85	0.80	0.51	0.70	1.51	1.70
	Median	1.05	1.30	1.05	1.30	0.55	0.80	1.55	1.80
	Maximum	1.06	1.32	1.30	1.80	0.60	0.90	1.60	1.90
	$P[Y_7 = 1]$ (95% CI)	0.133 (0.122, 0.144)		0.133 (0.122, 0.144)		0.125 (0.114, 0.135)		0.139 (0.128, 0.150)	
Difference of $P[Y_7 = 1]$ (95% CI)		−0.008 (−0.024, 0.008)		−0.008 (−0.023, 0.008)		0.007 (−0.008, 0.023)		−0.022 (−0.037, −0.006)	

$$
\begin{aligned}
&= \sum_{y_{k-1}=0}^{1} P[Y_j = 1 | S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\
&\quad \times P[Y_{k-1} = y_{k-1} | S \geq k-1, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\
&\quad \times P[S \geq k-1 | S \geq k-2, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\
&\quad + \sum_{y_{k-1}=0}^{1} P[Y_j = 1 | S = k-2, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\
&\quad \times P[Y_{k-1} = y_{k-1} | S = k-1, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\
&\quad \times P[S = k-2 | S \geq k-2, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\
&= \sum_{y_{k-1}=0}^{1} P[Y_j = 1 | S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\
&\quad \times P[Y_{k-1} = y_{k-1} | S \geq k-1, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\
&\quad \times P[S \geq k-1 | S \geq k-2, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\
&\quad + \sum_{y_{k-1}=0}^{1} P[Y_j = 1 | S \geq k-1, \overline{\mathbf{Y}}_{k-1} = \overline{\mathbf{y}}_{k-1}, Z = z] \\
&\quad \times \big(P[Y_{k-1} = y_{k-1} | S \geq k-1, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z] \\
&\quad \times \exp\{q_{z,k-1}(\overline{\mathbf{Y}}_{k-2}, y_{k-1})\} \\
&\quad / E\big[\exp\{q_{z,k-1}(\overline{\mathbf{Y}}_{k-2}, Y_{k-1})\} | S \geq k-1, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z\big]\big) \\
&\quad \times P[S = k-2 | S \geq k-2, \overline{\mathbf{Y}}_{k-2} = \overline{\mathbf{y}}_{k-2}, Z = z].
\end{aligned}
$$

The third equality follows by Assumptions 2 and 3. Since all the quantities on the right-hand side of the last equality are identified, $P(Y_j = 1 | S \geq k' - 1, \overline{\mathbf{Y}}_{k'-1} = \overline{\mathbf{y}}_{k'-1}, Z = z)$ is identified.

Gibbs Sampler for Posterior Computation

In the first step of the Gibbs sampler, we draw, for each subject with intermittent missing data, from the full conditional of $\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}$ given $\boldsymbol{\alpha}_z$, $\boldsymbol{\gamma}_z$, $\mathbf{m}_z^{(\alpha)}$, $\boldsymbol{\eta}_z^{(\alpha)}$, $\mathbf{m}_z^{(\gamma)}$, $\boldsymbol{\eta}_z^{(\gamma)}$, $\mathbf{Y}_{\mathrm{obs}}$, S, $\mathbf{R}_S$, and $Z = z$. The full conditional distribution can be expressed as

$$
\begin{aligned}
&P\big[\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}} = \mathbf{y}^{\mathrm{I}}_{\mathrm{mis}} | \boldsymbol{\alpha}_z, \boldsymbol{\gamma}_z, \mathbf{m}_z^{(\alpha)}, \boldsymbol{\eta}_z^{(\alpha)}, \mathbf{m}_z^{(\gamma)}, \boldsymbol{\eta}_z^{(\gamma)}, \\
&\qquad \mathbf{Y}_{\mathrm{obs}} = \mathbf{y}_{\mathrm{obs}}, S = s, \mathbf{R}_s = \mathbf{r}_s, Z = z\big] \\
&= P\big[\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}} = \mathbf{y}^{\mathrm{I}}_{\mathrm{mis}}, \mathbf{Y}_{\mathrm{obs}} = \mathbf{y}_{\mathrm{obs}}, S = s | \\
&\qquad \boldsymbol{\alpha}_z, \boldsymbol{\gamma}_z, \mathbf{m}_z^{(\alpha)}, \boldsymbol{\eta}_z^{(\alpha)}, \mathbf{m}_z^{(\gamma)}, \boldsymbol{\eta}_z^{(\gamma)}, Z = z\big] \\
&\qquad \Big/ \sum_{\text{all } \mathbf{y}^{\mathrm{I}}_{\mathrm{mis}}} P\big[\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}} = \mathbf{y}^{\mathrm{I}}_{\mathrm{mis}}, \mathbf{Y}_{\mathrm{obs}} = \mathbf{y}_{\mathrm{obs}}, S = s | \\
&\qquad \boldsymbol{\alpha}_z, \boldsymbol{\gamma}_z, \mathbf{r}_z^{(\alpha)}, \boldsymbol{\eta}_z^{(\alpha)}, \mathbf{r}_z^{(\gamma)}, \boldsymbol{\eta}_z^{(\gamma)}, Z = z\big]
\end{aligned}
$$

where the right-hand side can be expressed as a function of $\mathbf{y}^{\mathrm{I}}_{\mathrm{mis}}$, $\mathbf{y}_{\mathrm{obs}}$, s, $\boldsymbol{\alpha}_z$, and $\boldsymbol{\gamma}_z$.

In the second step, we draw from the full conditional of $\mathbf{m}_z^{(\alpha)}$ given $\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}$, $\boldsymbol{\alpha}_z$, $\boldsymbol{\gamma}_z$, $\boldsymbol{\eta}_z^{(\alpha)}$, $\mathbf{m}_z^{(\gamma)}$, $\boldsymbol{\eta}_z^{(\gamma)}$, $\{\mathbf{Y}_{\mathrm{obs}}\}$, $\{S\}$, $\{\mathbf{R}_S\}$, and $\{Z\} = z$, where the notation $\{\mathscr{D}\}$ denotes data $\mathscr{D}$ for all the individuals on the study. The full conditional can be expressed as

$$
\prod_{j=2}^{J} \prod_{y=0}^{1} f\big(m_{z,j,y}^{(\alpha)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\alpha}_z, \eta_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big),
$$

where

$$
\begin{aligned}
&f\big(m_{z,j,y}^{(\alpha)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\alpha}_z, \eta_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big) \\
&\qquad \propto \prod_{\substack{S_i \geq j, Y_{i,j-1} = y \\ i : Z_i = z}} B\big(\alpha_{z,j,\overline{\mathbf{Y}}_{i,j-2},y}; m_{z,j,y}^{(\alpha)} / \eta_{z,j,y}^{(\alpha)}, \big(1 - m_{z,j,y}^{(\alpha)}\big) / \eta_{z,j,y}^{(\alpha)}\big)
\end{aligned}
$$

and $B(\alpha; c, d)$ is a Beta density with parameters c and d.

In the third step, we draw from the full conditional of $\mathbf{m}_z^{(\gamma)}$ given $\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}$, $\boldsymbol{\alpha}_z$, $\boldsymbol{\gamma}_z$, $\boldsymbol{\eta}_z^{(\alpha)}$, $\mathbf{m}_z^{(\alpha)}$, $\boldsymbol{\eta}_z^{(\gamma)}$, $\{\mathbf{Y}_{\mathrm{obs}}\}$, $\{S\}$, $\{\mathbf{R}_S\}$, and $\{Z\} = z$. The full conditional can be expressed as

$$
\prod_{j=2}^{J} \prod_{y=0}^{1} f\big(m_{z,j-1,y}^{(\gamma)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\gamma}_z, \eta_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big),
$$

where

$$
\begin{aligned}
&f\big(m_{z,j-1,y}^{(\gamma)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\gamma}_z, \eta_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big) \\
&\qquad \propto \prod_{\substack{S_i \geq j-1, Y_{i,j-1} = y \\ i : Z_i = z}} B\big(\gamma_{z,j-1,\overline{\mathbf{Y}}_{i,j-2},y}; m_{z,j-1,y}^{(\gamma)} / \eta_{z,j-1,y}^{(\gamma)}, \\
&\qquad\qquad \big(1 - m_{z,j-1,y}^{(\gamma)}\big) / \eta_{z,j-1,y}^{(\gamma)}\big).
\end{aligned}
$$

In the fourth step, we draw from the full conditional of $\boldsymbol{\eta}_z^{(\alpha)}$ given $\mathbf{m}_z^{(\alpha)}$, $\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}$, $\boldsymbol{\alpha}_z$, $\boldsymbol{\gamma}_z$, $\mathbf{m}_z^{(\gamma)}$, $\boldsymbol{\eta}_z^{(\gamma)}$, $\{\mathbf{Y}_{\mathrm{obs}}\}$, $\{S\}$, $\{\mathbf{R}_S\}$, and $\{Z\} = z$. The full conditional can be expressed as

$$
\prod_{j=2}^{J} \prod_{y=0}^{1} f\big(\eta_{z,j,y}^{(\alpha)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\alpha}_z, m_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big),
$$

where

$$
\begin{aligned}
&f\big(\eta_{z,j,y}^{(\alpha)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\alpha}_z, m_{z,j,y}^{(\alpha)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big) \\
&\qquad \propto \frac{g(\mathbf{E}_{z,j,y}^{(\alpha)})}{(g(\mathbf{E}_{z,j,y}^{(\alpha)}) \eta_{z,j,y}^{(\alpha)} + 1)^2} \\
&\qquad\quad \times \prod_{\substack{S_i \geq j, Y_{i,j-1} = y \\ i : Z_i = z}} B\big(\alpha_{z,j,\overline{\mathbf{Y}}_{i,j-2},y}; m_{z,j,y}^{(\alpha)} / \eta_{z,j,y}^{(\alpha)}, \big(1 - m_{z,j,y}^{(\alpha)}\big) / \eta_{z,j,y}^{(\alpha)}\big).
\end{aligned}
$$

In the fifth step, we draw from the full conditional of $\boldsymbol{\eta}_z^{(\gamma)}$ given $\mathbf{m}_z^{(\alpha)}$, $\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}$, $\boldsymbol{\alpha}_z$, $\boldsymbol{\gamma}_z$, $\mathbf{m}_z^{(\gamma)}$, $\boldsymbol{\eta}_z^{(\alpha)}$, $\{\mathbf{Y}_{\mathrm{obs}}\}$, $\{S\}$, $\{\mathbf{R}_S\}$, and $\{Z\} = z$. The full conditional can be expressed as

$$
\prod_{j=2}^{J} \prod_{y=0}^{1} f\big(\eta_{z,j-1,y}^{(\gamma)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\gamma}_z, m_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big),
$$

where

$$
\begin{aligned}
&f\big(\eta_{z,-1j,y}^{(\gamma)} | \{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, \boldsymbol{\gamma}_z, m_{z,j-1,y}^{(\gamma)}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big) \\
&\qquad \propto \frac{g(\mathbf{E}_{z,j-1,y}^{(\gamma)})}{(g(\mathbf{E}_{z,j-1,y}^{(\gamma)}) \eta_{z,j-1,y}^{(\gamma)} + 1)^2} \\
&\qquad\quad \times \prod_{\substack{S_i \geq j-1, Y_{i,j-1} = y \\ i : Z_i = z}} B\big(\gamma_{z,j-1,\overline{\mathbf{Y}}_{i,j-2},y}; m_{z,j-1,y}^{(\gamma)} / \eta_{z,j-1,y}^{(\gamma)}, \\
&\qquad\qquad \big(1 - m_{z,j-1,y}^{(\gamma)}\big) / \eta_{z,j-1,y}^{(\gamma)}\big).
\end{aligned}
$$

To draw from the full conditionals for steps two to five, we use slice sampling (Neal 2003).

In the sixth step, we draw from the full conditional of $\boldsymbol{\alpha}_z$ given $\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}$, $\boldsymbol{\gamma}_z$, $\mathbf{m}_z^{(\alpha)}$, $\boldsymbol{\eta}_z^{(\alpha)}$, $\mathbf{m}_z^{(\gamma)}$, $\boldsymbol{\eta}_z^{(\gamma)}$, $\{\mathbf{Y}_{\mathrm{obs}}\}$, $\{S\}$, $\{\mathbf{R}_s\}$, and $\{Z\} = z$. The full conditional can be expressed as

$$\prod_{j=2}^{J}\prod_{y=0}^{1}\prod_{\text{all }\overline{\mathbf{y}}_{j-2}} f\big(\alpha_{z,j,\overline{\mathbf{y}}_{j-2},y}|\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, m^{(\alpha)}_{z,j,y}, \eta^{(\alpha)}_{z,j,y}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big),$$

where

$$\begin{aligned} &f\big(\alpha_{z,j,\overline{\mathbf{y}}_{j-2},y}|\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, m^{(\alpha)}_{z,j,y}, \eta^{(\alpha)}_{z,j,y}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big) \\ &\quad = B\big(\alpha_{z,j,\overline{\mathbf{y}}_{j-2},y}; m^{(\alpha)}_{z,j,y}/\eta^{(\alpha)}_{z,j,y} + o^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y}, \\ &\qquad \big(1 - m^{(\alpha)}_{z,j,y}\big)/\eta^{(\alpha)}_{z,j,y} + n^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y} - o^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y}\big), \end{aligned}$$

$n^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y}$ is the number of subjects with $S \geq j$, $Y_{j-1} = y$, $\overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}$, and $Z = z$, and $o^{(\alpha)}_{z,j,\overline{\mathbf{y}}_{j-2},y}$ is the number of subjects with $S \geq j$, $Y_{j-1} = y$, $\overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}$, $Z = z$, and $Y_j = 1$.

Finally, we draw from the full conditional of $\boldsymbol{\gamma}_z$ given $\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}$, $\boldsymbol{\alpha}_z$, $\mathbf{r}_z^{(\alpha)}$, $\boldsymbol{\eta}_z^{(\alpha)}$, $\mathbf{r}_z^{(\gamma)}$, $\boldsymbol{\eta}_z^{(\gamma)}$, $\{\mathbf{Y}_{\mathrm{obs}}\}$, $\{S\}$, $\{\mathbf{R}_s\}$, and $\{Z\} = z$. The full conditional can be expressed as

$$\prod_{j=2}^{J}\prod_{y=0}^{1}\prod_{\text{all }\overline{\mathbf{y}}_{j-2}} f\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y}|\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, m^{(\gamma)}_{z,j-1,y}, \eta^{(\gamma)}_{z,j-1,y}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big),$$

where

$$\begin{aligned} &f\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y}|\{\mathbf{Y}^{\mathrm{I}}_{\mathrm{mis}}\}, m^{(\gamma)}_{z,j-1,y}, \eta^{(\gamma)}_{z,j-1,y}, \{\mathbf{Y}_{\mathrm{obs}}\}, \{S\}, \{Z\} = z\big) \\ &\quad = B\big(\gamma_{z,j-1,\overline{\mathbf{y}}_{j-2},y}; m^{(\gamma)}_{z,j-1,y}/\eta^{(\gamma)}_{z,j-1,y} + o^{(\gamma)}_{z,j-1,\overline{\mathbf{y}}_{j-2},y}, \\ &\qquad \big(1 - m^{(\gamma)}_{z,j-1,y}\big)/\eta^{(\gamma)}_{z,j-1,y} + n^{(\gamma)}_{z,j-1,\overline{\mathbf{y}}_{j-2},y} - o^{(\gamma)}_{z,j-1,\overline{\mathbf{y}}_{j-2},y}\big), \end{aligned}$$

$n^{(\gamma)}_{z,j-1,\overline{\mathbf{y}}_{j-2},y}$ is the number of subjects with $S \geq j-1$, $Y_{j-1} = y$, $\overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}$, and $Z = z$, and $o^{(\gamma)}_{z,j-1,\overline{\mathbf{y}}_{j-2},y}$ is the number of subjects with $S = j-1$, $Y_{j-1} = y$, $\overline{\mathbf{Y}}_{j-2} = \overline{\mathbf{y}}_{j-2}$, and $Z = z$.

[Received May 2009. Revised February 2010.]

REFERENCES

Albert, P. S. (2000), "A Transitional Model for Longitudinal Binary Data Subject to Nonignorable Missing Data," *Biometrics*, 56 (2), 602–608. [1334]

Albert, P. S., Follmann, D. A., Wang, S. A., and Suh, E. B. (2002), "A Latent Autoregressive Model for Longitudinal Binary Data Subject to Informative Missingness," *Biometrics*, 58 (3), 631–642. [1334]

Baker, S. G. (1995), "Marginal Regression for Repeated Binary Data With Outcome Subject to Non-Ignorable Non-Response," *Biometrics*, 51 (3), 1042–1052. [1334]

Baker, S. G., Rosenberger, W. F., and DerSimonian, R. (1992), "Closed-Form Estimates for Missing Counts in Two-Way Contingency Tables," *Statistics in Medicine*, 11, 643–657. [1334]

Birmingham, J., and Fitzmaurice, G. M. (2002), "A Pattern-Mixture Model for Longitudinal Binary Responses With Nonignorable Nonresponse," *Biometrics*, 58 (4), 989–996. [1334]

Christiansen, C. L., and Morris, C. N. (1997), "Hierarchical Poisson Regression Modeling," *Journal of the American Statistical Association*, 92, 618–632. [1337]

Daniels, M. J. (1999), "A Prior for the Variance in Hierarchical Models," *The Canadian Journal of Statistics*, 27 (3), 567–578. [1337]

Daniels, M. J., and Hogan, J. W. (2000), "Reparameterizing the Pattern Mixture Model for Sensitivity Analyses Under Informative Dropout," *Biometrics*, 56 (4), 1241–1248. [1334]

——— (2008), *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, New York: Chapman & Hall/CRC. [1334]

DeGruttola, V., and Tu, X. M. (1994), "Modelling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time," *Biometrics*, 50 (4), 1003–1014. [1334]

Diggle, P., and Kenward, M. G. (1994), "Informative Drop-Out Longitudinal Data Analysis," *Applied Statistics*, 43 (1), 49–93. [1334]

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588. [1342]

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96 (456), 1348–1361. [1342]

Fisher, B., Costantino, J. P., Wickerham, D. L., Redmond, C. K., Kavanah, M., Cronin, W. M., Vogel, V., Robidoux, A., Dimitrov, N., Atkins, J., Daly, M., Wieand, S., Tan-Chiu, E., Ford, L., Wolmark, N., and other National Surgical Adjuvant Breast and Bowel Project Investigators (1998), "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. [1333]

Fitzmaurice, G. M., and Laird, N. M. (2000), "Generalized Linear Mixture Models for Handling Non-Ignorable Dropouts in Longitudinal Studies," *Biostatistics*, 1 (2), 141–156. [1334]

Fitzmaurice, G. M., Molenberghs, G., and Lipsitz, S. R. (1995), "Regression Models for Longitudinal Binary Responses With Informative Drop-Outs," *Journal of the Royal Statistical Society, Ser. B*, 57 (4), 691–704. [1334]

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, New York: Chapman & Hall/CRC. [1342]

Harel, O., and Schafer, J. L. (2009), "Partial and Latent Ignorability in Missing-Data Problems," *Biometrika*, 96 (1), 37. [1334,1335]

Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1), 153–161. [1334]

Hogan, J. W., and Laird, N. M. (1997a), "Model-Based Approaches to Analysing Incomplete Longitudinal and Failure Time Data," *Statistics in Medicine*, 16 (3), 259–272. [1334]

——— (1997b), "Mixture Models for the Joint Distribution of Repeated Measures and Event Times," *Statistics in Medicine*, 16 (3), 239–257. [1334]

Ibrahim, J. G., Chen, M. H., and Lipsitz, S. R. (2001), "Missing Responses in Generalised Linear Mixed Models When the Missing Data Mechanism is Nonignorable," *Biometrika*, 88 (2), 551. [1334]

Kaciroti, N. A., Schork, M. A., Raghunathan, T., and Julius, S. (2009), "A Bayesian Sensitivity Model for Intention-to-Treat Analysis on Binary Outcomes With Dropouts," *Statistics in Medicine*, 28, 572–585. [1334]

Kenward, M. G., and Molenberghs, G. (1999), "Parametric Models for Incomplete Continuous and Categorical Longitudinal Data," *Statistical Methods in Medical Research*, 8 (1), 51. [1334]

Kenward, M. G., Molenberghs, G., and Thijs, H. (2003), "Pattern-Mixture Models With Proper Time Dependence," *Biometrika*, 90 (1), 53–71. [1335]

Kurland, B. F., and Heagerty, P. J. (2004), "Marginalized Transition Models for Longitudinal Binary Data With Ignorable and Non-Ignorable Drop-Out," *Statistics in Medicine*, 23 (17), 2673–2695. [1334]

Land, S., Wieand, S., Day, R., Ten Have, T., Costantino, J. P., Lang, W., and Ganz, P. A. (2002), "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," in *Statistical Methods for Quality of Life Studies*: *Design, Measurements and Analysis*, eds. M. Mesbah, B. F. Cole, and M.-L. Ting Lee, New York: Springer, pp. 71–85. [1333,1334]

Lee, J. Y., Hogan, J. W., and Hitsman, B. (2008), "Sensitivity Analysis and Informative Priors for Longitudinal Binary Data With Outcome-Related Drop-Out," technical report, Brown University. [1334]

Lin, H., McCulloch, C. E., and Rosenheck, R. A. (2004), "Latent Pattern Mixture Models for Informative Intermittent Missing Data in Longitudinal Studies," *Biometrics*, 60 (2), 295–305. [1334]

Little, R. J. A. (1993), "Pattern-Mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association*, 88 (421), 125–134. [1334]

——— (1994), "A Class of Pattern-Mixture Models for Normal Incomplete Data," *Biometrika*, 81 (3), 471–483. [1334]

——— (1995), "Modeling the Drop-Out Mechanism in Repeated-Measures Studies," *Journal of the American Statistical Association*, 90 (431), 1112–1121. [1334]

Little, R. J. A., and Rubin, D. B. (1999), Comment on "Adjusting for Non-Ignorable Drop-Out Using Semiparametric Models," by D. O. Scharfstein, A. Rotnitsky, and J. M. Robins, *Journal of the American Statistical Association*, 94, 1130–1132. [1334]

Liu, X., Waternaux, C., and Petkova, E. (1999), "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, Ser. C*, 48 (1), 103–115. [1334]

Molenberghs, G., and Kenward, M. G. (2007), *Missing Data in Clinical Studies*, New York: Wiley. [1334]

Molenberghs, G., Kenward, M. G., and Lesaffre, E. (1997), "The Analysis of Longitudinal Ordinal Data With Nonrandom Drop-Out," *Biometrika*, 84 (1), 33–44. [1334]

Neal, R. M. (2003), "Slice Sampling," *The Annals of Statistics*, 31 (3), 705–741. [1344]

Nordheim, E. V. (1984), "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. [1334]

Pauler, D. K., McCoy, S., and Moinpour, C. (2003), "Pattern Mixture Models for Longitudinal Quality of Life Studies in Advanced Stage Disease," *Statistics in Medicine*, 22 (5), 795–809. [1334]

Pulkstenis, E. P., Ten Have, T. R., and Landis, J. R. (1998), "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. [1334]

Radloff, L. S. (1977), "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population," *Applied Psychological Measurement*, 1 (3), 385. [1333]

Robins, J. M. (1997), "Non-Response Models for the Analysis of Non-Monotone Non-Ignorable Missing Data," *Statistics in Medicine*, 16 (1), 21–37. [1342]

Robins, J. M., and Ritov, Y. A. A. (1997), "Toward a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-Parametric Models," *Statistics in Medicine*, 16 (3), 285–319. [1336]

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89 (427), 846–866. [1334]

——— (1995), "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, 90 (429), 106–121. [1334]

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998), "Semiparametric Regression for Repeated Outcomes With Nonignorable Nonresponse," *Journal of the American Statistical Association*, 93 (444), 1321–1322. [1334]

Rotnitzky, A., Scharfstein, D. O., Su, T.-L., and Robins, M. J. (2001), "Methods for Conducting Sensitivity Analysis of Trials With Potentially Nonignorable Competing Causes of Censoring," *Biometrics*, 57, 103–113. [1334]

Roy, J. (2003), "Modeling Longitudinal Data With Nonignorable Dropouts Using a Latent Dropout Class Model," *Biometrics*, 59 (4), 829–836. [1334]

Roy, J., and Daniels, M. J. (2008), "A General Class of Pattern Mixture Models for Nonignorable Dropout With Many Possible Dropout Times," *Biometrics*, 64, 538–545. [1334]

Scharfstein, D. O., Daniels, M. J., and Robins, J. M. (2003), "Incorporating Prior Beliefs About Selection Bias Into the Analysis of Randomized Trials With Missing Outcomes," *Biostatistics*, 4 (4), 495. [1334]

Scharfstein, D. O., Halloran, M. E., Chu, H., and Daniels, M. J. (2006), "On Estimation of Vaccine Efficacy Using Validation Samples With Selection Bias," *Biostatistics*, 7 (4), 615. [1337]

Scharfstein, D. O., Manski, C. F., and Anthony, J. C. (2004), "On the Construction of Bounds in Prospective Studies With Missing Ordinal Outcomes: Application to the Good Behavior Game Trial," *Biometrics*, 60 (1), 154–164. [1334]

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999), "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models," *Journal of the American Statistical Association*, 94 (448), 1096–1146. [1334]

Shepherd, B. E., Gilbert, P. B., and Mehrotra, D. V. (2007), "Eliciting a Counterfactual Sensitivity Parameter," *The American Statistician*, 61 (1), 56. [1337]

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., and Landis, J. R. (1998), "Mixed Effects Logistic Regression Models for Longitudinal Binary Response Data With Informative Drop-Out," *Biometrics*, 54 (1), 367–383. [1334]

Ten Have, T. R., Miller, M. E., Reboussin, B. A., and James, M. K. (2000), "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. [1334]

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., and Curran, D. (2002), "Strategies to Fit Pattern-Mixture Models," *Biostatistics*, 3 (2), 245. [1334]

Troxel, A. B., Harrington, D. P., and Lipsitz, S. R. (1998), "Analysis of Longitudinal Data With Non-Ignorable Non-Monotone Missing Values," *Journal of the Royal Statistical Society, Ser. C*, 47 (3), 425–438. [1334]

Troxel, A. B., Lipsitz, S. R., and Harrington, D. P. (1998), "Marginal Models for the Analysis of Longitudinal Measurements With Nonignorable Non-Monotone Missing Data," *Biometrika*, 85 (3), 661. [1334]

Tsiatis, A. A. (2006), *Semiparametric Theory and Missing Data*, New York: Springer. [1334]

van der Laan, M. J., and Robins, J. M. (2003), *Unified Methods for Censored Longitudinal Data and Causality*, New York: Springer. [1334]

Vansteelandt, S., Goetghebeur, E., Kenward, M. G., and Molenberghs, G. (2006), "Ignorance and Uncertainty Regions as Inferential Tools in a Sensitivity Analysis," *Statistica Sinica*, 16 (3), 953–979. [1334]

Vansteelandt, S., Rotnitzky, A., and Robins, J. (2007), "Estimation of Regression Models for the Mean of repeated Outcomes Under Nonignorable Nonmonotone Nonresponse," *Biometrika*, 94 (4), 841. [1334]

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia, PA: Society for Industrial and Applied Mathematics. [1342]

Wu, M. C., and Carroll, R. J. (1998), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process," *Biometrics*, 44 (1), 175–188. [1334]

Yuan, Y., and Little, R. J. A. (2009), "Mixed-Effect Hybrid Models for Longitudinal Data With Nonignorable Drop-Out," *Biometrics*, 65 (2), 478–486. [1334]