

Bivariate joint modelling for mixed continuous and binary responses in the presence of non-monotone, non-ignorable missingness: The case of prostate cancer

Madiha Liaqat, Shahid Kamal, Florian Fischer

Madiha Liaqat

College of Statistical and Actuarial Sciences (CSAS), University of the Punjab, Lahore, Pakistan
madiha.phd.stat@pu.edu.pk

Prof. Dr. Shahid Kamal

College of Statistical and Actuarial Sciences (CSAS), University of the Punjab, Lahore, Pakistan
kamal0shahid@gmail.com

Dr. Florian Fischer [Corresponding author]

Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany
florian.fischer1@charite.de

Corresponding author:

Dr. Florian Fischer
Charité – Universitätsmedizin Berlin
Institute of Public Health
Charitéplatz 1
10117 Berlin
Germany
florian.fischer1@charite.de

1 **Abstract**

2 Joint modelling for mixed longitudinal responses has played a prominent part in disease decision-
3 making. It is based on a joint strategy of estimating joint likelihood with shared random effects.
4 Non-ignorable missingness in outcomes increases complexity in joint model; a shared parameter
5 model is proposed to incorporate non-ignorable missing data for joint modelling of longitudinal
6 responses and missing data mechanism. Parameters are estimated under the Bayesian paradigm
7 and implemented via Markov chain Monte Carlo (MCMC) methods with Gibbs sampler. To
8 demonstrate the effectiveness of the proposed method, the joint model is applied to analyze a
9 prostate cancer dataset. The objective is to assess whether there is an association between two
10 mixed longitudinal biomarkers, which could have important implications for understanding
11 disease progression and guiding treatment decisions. The dataset contains non-monotone
12 missingness pattern. To evaluate the performance and robustness of the proposed joint model,
13 simulation studies are conducted.

14

15 **Keywords:** missing data, missingness, oncology, prostate

16 **1. Introduction**

17 Longitudinal studies have a wide range of applications in medicine, where measurements are
18 repeatedly observed for individuals over time. These measurements are correlated even for
19 different types of responses and collected for the same individuals. For example, in oncology
20 research, more than one biomarker of mixed types is measured repeatedly over time for the same
21 patient. These biomarkers are inherently associated – and this association is accounted for valid
22 conclusions. Simultaneous modelling of mixed biomarkers is an optimal approach to account for
23 correlation among multiple biomarkers [1].

24 Missing data is an unavoidable issue in longitudinal studies because patients may not come for a
25 pre-specified follow-up visit at a particular time point for many reasons. In this regard, missingness
26 exists in collected data and creates hindrances to extracting inference without making assumptions
27 about the missingness mechanism. According to Rubin [2], missing completely at random
28 (MCAR) and missing at random (MAR) mechanisms being not dependent upon unobserved
29 measurements are ignorable. Not missing at random (NMAR) mechanism may depend upon both
30 observed and unobserved responses and is considered non-ignorable. Non-ignorable missingness
31 mechanism is considered in data analysis with longitudinal measurements to get unbiased
32 estimates.

33 In literature, selection models, pattern-mixture models (PMMs), and shared parameter models
34 (SPMs) are mostly applied to tackle non-ignorable missing data. These models use different
35 factorizations for missingness and measured processes to jointly model longitudinal responses
36 incorporating missingness. The SPM approach is widely applied to model MAR data in
37 longitudinal responses. Still, there exists a gap to employ SPM for NMAR. Non-ignorable missing
38 data is addressed by employing logistic or probit models, which are specifically designed to handle

39 situations where the missing data mechanism is influenced by unobserved variables or responses,
40 resulting in NMAR mechanisms. By integrating the logistic or probit model into the analysis,
41 researchers can effectively mitigate potential biases caused by non-ignorable missing data that
42 consequently makes precise inferences based on the available data. Logistic and probit models
43 provide valuable insights into the underlying patterns of missingness, contributing to obtaining
44 reliable results in longitudinal studies and other scenarios where missing data is common.

45 Joint modelling terms are applied to simultaneously analyzed more than one outcome. Many
46 methods have been proposed to jointly analyze single or multivariate longitudinal measurements
47 alongside event time data in recent years [3,4]. Another joint modelling term is used to
48 simultaneously analyze more than one longitudinal outcome possibly of different types using
49 random effects, marginal or conditional models [5,6]. Catalano and Ryan [7] analyzed toxicity
50 data proposing bivariate latent variable models to account for the relationship between fetal weight
51 and malformation in live fetuses. Leon and Carriere [8] studied one longitudinal and one binary
52 response to assess maternal smoking's effect on the respiratory illness of children. Liu et al. [9]
53 formulated a joint model to handle longitudinal binary and continuous responses by incorporating
54 an ignorable missing data mechanism. Li et al. [10] applied joint modelling for continuous, binary,
55 and ordinal responses under the Bayesian framework, while Kürüm et al. [11] proposed a joint
56 model to analyze binary and continuous responses under frequentist statistics.

57 Under the umbrella of non-ignorable missing data (NMAR) fall both non-monotone and monotone
58 missingness patterns, which in turn provide incomplete observations. Intermittent missingness
59 comes under non-monotone missingness pattern, where an individual misses any visit during the
60 follow-up time, and returns to show up for subsequent visits. The monotone missingness pattern
61 usually refers to informative drop-out, where an individual leaves the study before completion and

62 never comes back to complete follow-up time. Stubbendick and Ibrahim [12] used a likelihood-
63 based approach to incorporate non-monotone NMAR by proposing a joint likelihood of outcome
64 and missing data indicators. Hogan et al. [13] worked on monotone missingness in longitudinal
65 data. Gaskins et al. [14] worked on non-ignorable drop-out mixed longitudinal responses using
66 PMM under joint modelling methodology.

67 In this article, a joint model is proposed for longitudinal continuous-binary biomarkers using a
68 conditional modelling approach, and two shared parameter models are proposed for non-ignorable
69 missingness under the conditional model. For estimating the model's parameters different
70 approaches are employed: likelihood-based approach, linear programming technique, and
71 Bayesian approach. We work under the Bayesian framework to obtain parameter estimates.

72 The article presents a robust statistical approach to analyze longitudinal mixed-effects biomarker
73 measurements, accounting for non-monotone, non-ignorable missingness in the context of prostate
74 cancer research. By utilizing joint modelling techniques and considering the missingness
75 mechanisms, the researchers aim to obtain more reliable and meaningful results from their data
76 analysis.

77

78 **2. Motivation: Prostate cancer dataset**

79 This research is motivated by a comprehensive dataset obtained from prostate cancer patients who
80 underwent external beam radiotherapy (EBRT), Androgen deprivation therapy (ADT), and a
81 combination of ADT along with other therapies. This dataset spans from 2012 to 2019 and includes
82 information collected during up to 5 follow-up visits after treatment. The data was sourced from
83 Mayo hospital, a renowned referral hospital in Pakistan. Data was received in 2021 by the authors
84 of this study. The authors had no access to information that could identify individual participants.

85 The evaluation of patients took place from the pre-treatment phase to the post-treatment follow-
86 up. During the pre-treatment phase, demographic characteristics and medical history of the patients
87 were recorded. We specifically focused on patients' age, body mass index (BMI), and Gleason
88 score as baseline predictors for our study. Blood tests were also conducted to measure prostate-
89 specific antigen (PSA), alkaline phosphatase (ALT), platelets, and bilirubin. These biomarkers
90 were tracked after treatment at each follow-up visit, with PSA and ALP values considered as
91 indicators of prostate cancer progression after treatment, and endogenous platelets and bilirubin as
92 time-dependent covariates. Our study included a total of 1,504 patients diagnosed with prostate
93 cancer, out of which 1,026 received ADT and combination therapies as part of their treatment.

94

95 **3. Model specification**

96 *3.1 Model for longitudinal measurements*

97 Data analysis is described by the joint model specification; sub-models illustrate such mixed types
98 of complex data modelling. The first sub-model assumes to follow linear mixed-effects model
99 specified for continuous longitudinal biomarker. Binary longitudinal response is assumed to follow
100 mixed-effects logistic regression model.

101 Let Y_{ij} be the continuous longitudinal outcome for i^{th} individual $i = 1, 2, 3, \dots, n$ at time t_{ij}
102 $j = 1, 2, 3, \dots, m^y$, assume linear mixed effects model as

$$103 \quad Y_{ij} = x_{ij}^y \alpha_1 + w_{ij}^y u_{1i} + \varepsilon_{ij} \quad (1)$$

104 where, α_1 is a p^y dimensional vector of fixed effects regression coefficients, x_{ij}^y vector of
105 covariates. w_{ij}^y is q^y dimensional vector of random effects, u_{1i} is a vector of random effects, that
106 is independently and identically distributed as multivariate normal with mean vector 0 and
107 covariance matrix D . $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is the random error term.

108 Let, Z_{ij} be the binary repeated measurements observe for i^{th} individual, $i = 1, 2, 3, \dots, n$ at time s_{ij} .
109 Where $s_{ij} = t_{ij}$, $j = 1, 2, 3, \dots, m^z$, binary response Z_{ij} assumes to follow logistic mixed effects
110 model that is given by

$$111 \quad Z_{ij} | u_{2i} \sim Ber(\lambda_{ij}),$$
$$112 \quad \text{logit}(\lambda_{ij}) = x_{ij}^{z'} \alpha_2 + w_{ij}^{z'} u_{2i} + Y_j Y_{ij}, \quad (2)$$

113
114 where, α_2 is a p^z dimensional vector of fixed effects regression coefficients, x_{ij}^z vector of
115 covariates. w_{ij}^z is q^y dimensional vector of random effects, $u_{2i} \sim MVN(0, D)$ is a vector of random
116 effects, with the addition of associated parameter Y_j to measure the effect of response Y_{ij} at time
117 t_{ij} on response Z_{ij} at the same time s_{ij} .

118

119 ***3.2 Missing data mechanism and shared parameter model***

120 Longitudinal data are not fully observed which leads to incomplete measurements. The most
121 important consideration is to specify an appropriate missing data mechanism based on assumptions
122 related to unobserved and observed data. Little and Schluchter [15] as well as Fitzmaurice and
123 Laird [16] applied general location model described by Olkin and Tate [17] assuming the
124 ignorability assumption. This assumption leads to the usage of observed responses only without
125 considering a model for missingness mechanism. Usually, the expectation-maximization (EM)
126 algorithm is employed for parameter estimation.

127 In this study, we examine two distinct missingness mechanisms to account for the presence of two
128 different missingness patterns in longitudinal outcomes, two SPM are specified for missingness
129 mechanisms.

130 Let $R_i = (R_{i1}, R_{i2}, R_{i3}, \dots, R_{in})'$ be the vector of response indicators, if Y_{ij} observed $R_{ij} = 1$, and R_{ij}
 131 $= 0$ for missing Y_{ij} . In this paper, a SPM based on non-ignorable missingness mechanisms for Y_i
 132 and Z_i consider, assuming logistic mixed effects regression models for R_{ij}^y and R_{ij}^z responses.

133 Let R_{ij}^y be the missingness indicator for continuous response at time t_{ij} , such that $R_{ij}^y = 1$ if Y_{ij} is
 134 not fully observed, the model for R_{ij}^y is given by

$$135 \quad R_{ij}^y | u_{1i} \sim \text{Ber}(\lambda_{ij}^{R^y}),$$

$$136 \quad \text{logit}(\lambda_{ij}^{R^y}) = \psi_{ij}^{y'} \Phi_1 + \pi^{y'} u_{1i}, \quad (3)$$

137
 138 Let R_{ij}^z be the missingness indicator for continuous response at time s_{ij} , such that $R_{ij}^z = 1$ if Z_{ij} is
 139 not fully observed, the model for R_{ij}^z is given by

$$140 \quad R_{ij}^z | u_{2i} \sim \text{Ber}(\lambda_{ij}^{R^z}),$$

$$141 \quad \text{logit}(\lambda_{ij}^{R^z}) = \psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y, \quad (4)$$

142 To formulate joint distribution of responses given the random effects, let Y_{ij} and Z_{ij} are partitioned
 143 into $Y_{ij} = (Y_{ij}^{obs}, Y_{ij}^{mis})$ and $Z_{ij} = (Z_{ij}^{obs}, Z_{ij}^{mis})$, respectively.

144 Joint model given the random effects is given as

$$145 \quad f(Y_{ij}^{obs}, Z_{ij}^{obs}, Y_{ij}^{mis}, Z_{ij}^{mis} | u_{1i}, u_{2i}) = f(Y_{ij}, Z_{ij}, R_{ij}^y, R_{ij}^z | u_{1i}, u_{2i})$$

$$146 \quad f(Y_{ij}, Z_{ij}, R_{ij}^y, R_{ij}^z | u_{1i}, u_{2i}) = f(Z_{ij} | Y_{ij}, R_{ij}^y, R_{ij}^z, u_{1i}, u_{2i}) \times f(Y_{ij} | R_{ij}^y, R_{ij}^z, u_{1i}) \times f(R_{ij}^z | R_{ij}^y, u_{1i}, u_{2i})$$

$$147 \quad \times f(R_{ij}^y | u_{1i}, u_{2i})$$

$$147 \quad = f(Z_{ij} | Y_{ij}, R_{ij}^z, u_{2i}) \times f(Y_{ij} | R_{ij}^y, u_{1i}) \times f(R_{ij}^z | R_{ij}^y, u_{2i}) \times f(R_{ij}^y | u_{1i})$$

$$148 \quad = f(Z_{ij} | Y_{ij}, u_{2i}) \times f(Y_{ij} | u_{1i}) \times f(R_{ij}^z | R_{ij}^y, u_{2i}) \times f(R_{ij}^y | u_{1i}). \quad (5)$$

149 Joint probability distribution function is formulated further as

$$150 \quad f(Y_{ij}, Z_{ij}, R_{ij}^y, R_{ij}^z) = \int_{u_{1i}} \int_{u_{2i}} f(Y_{ij}, Z_{ij}, R_{ij}^y, R_{ij}^z | u_{1i}, u_{2i}) h(u_{1i}, u_{2i}) du_{1i} du_{2i}$$

$$\begin{aligned}
 &= \int_{u_{1i}} \int_{u_{2i}} \prod_{j=1}^{m^y} f(Y_{ij}|u_{1i}) \prod_{j=1}^{m^z} f(Z_{ij}|u_{2i}, Y_{ij}) \times \prod_{j=1}^{m^y} f(R_{ij}^y|u_{1i}) \prod_{j=1}^{m^z} f(R_{ij}^z|u_{2i}, Y_{ij}) du_{1i} du_{2i}. \\
 & \tag{6}
 \end{aligned}$$

3.3 Prior distributions and hierarchical model

The unknown parameters of the proposed model are estimated by employing the Bayesian framework using Markov chain Monte Carlo (MCMC) techniques [18]. MCMC has an advantage over conventional methods, with MCMC conditional distribution of each parameter given others is easily specifying. Priors are chosen to carry out Bayesian inference for unknown parameters, let $\emptyset = \{\alpha_1, \alpha_2, \sigma^2, D_1, D_2, \Phi_1, \Phi_2, \pi^y, \pi^z, Y_l, \kappa_k\}$ be the unknown parameters' vector where $k = 1, 2, 3, \dots, m^y$ and $l = 1, 2, 3, \dots, m^z$. We specify proper prior for each set of unknown parameters: Gaussian distribution for the fixed effects, and Inverse Gamma distribution for random effects. Different levels of variances for each distribution are tried out to make robust choice of fixed effects estimates. Independent prior distributions for unknown parameters are chosen by assigning hyperparameters that lead to low-informative prior distributions.

$$\alpha_1 \sim N'_{p1}(\mu_{\alpha_1}, \Sigma_{\alpha_1}),$$

$$\sigma^2 \sim \text{IG}(a_{\sigma^2}, b_{\sigma^2}),$$

$$\alpha_2 \sim N'_{p2}(\mu_{\alpha_2}, \Sigma_{\alpha_2}),$$

$$D_1 \sim \text{IWishart}(\Psi_{D_1}, \nu_{D_1}),$$

$$D_2 \sim \text{IWishart}(\Psi_{D_2}, \nu_{D_2}),$$

$$\Phi_1 \sim N'_{pk}(\mu_{\Phi_1}, \Sigma_{\Phi_1}),$$

$$\Phi_2 \sim N'_{pk}(\mu_{\Phi_2}, \Sigma_{\Phi_2}),$$

$$\pi^y \sim N'_{q^y}(\mu_{\pi^y}, \Sigma_{\pi^y}),$$

$$\pi^z \sim N'_{q^z}(\mu_{\pi^z}, \Sigma_{\pi^z}),$$

$$\begin{aligned}
 173 \quad & r \sim \Gamma(a_r, b_r), \\
 174 \quad & \gamma_j \sim N(\mu_{\gamma_j}, \sigma_{\gamma_j}^2), j = 1, 2, 3, \dots, m^y, \\
 175 \quad & \kappa_j \sim N(\mu_{\kappa_j}, \sigma_{\kappa_j}^2), j = 1, 2, 3, \dots, m^z, \tag{7}
 \end{aligned}$$

176
 177 where $\Gamma(a, b)$ and $\Gamma(a, b)$ denote inverse gamma distribution and gamma distribution with shape
 178 parameter a and scale parameter b . IWishart (ψ, ν) represents the inverse Wishart distribution
 179 with scale and matrix parameters ν and ψ , respectively. $N(\mu, \Sigma)$ denotes a normal distribution with
 180 mean vector μ and covariance matrix Σ .

181 The joint posterior density is formulated as

$$\begin{aligned}
 182 \quad & \theta(\emptyset, u_1, u_2, Y_{ij}^{mis}, Z_{ij}^{mis} | Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^y} \mathbb{I}(Y_{ij}; x_{ij}^{y'} \alpha_1 + w_{ij}^{y'} u_{1i}, \sigma_\varepsilon^2) \times (\lambda_{ij}^{R^y})^{R_{ij}^y} (1 - \lambda_{ij}^{R^y})^{1 - R_{ij}^y} \\
 & \times \prod_{i=1}^n \prod_{k=1}^{m^z} (\lambda_{ij})^{Z_{ik}} (1 - \lambda_{ij})^{1 - Z_{ik}} \times (\lambda_{ij}^{R^z})^{R_{ik}^z} (1 - \lambda_{ij}^{R^z})^{1 - R_{ik}^z} \times \theta(\emptyset), \tag{8}
 \end{aligned}$$

183
 184
 185 where, $\theta(\emptyset)$ is the joint prior distribution of unknown parameters.

186 Samples are drawn iteratively from conditional posterior distributions derived from (8) using
 187 Gibbs sampler, the full conditional distributions for parameters are given by

$$188 \quad \alpha |_{\mathbf{U}(-\alpha), Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z} \sim N_{p^y}(\Sigma_\alpha | \dots (\prod_{i=1}^n x_i^{y'} \frac{Y_i - w_i^{y'} a_{1i}}{\sigma_\varepsilon^2}), \Sigma_\alpha | \dots) \tag{9}$$

189 Where $\Sigma_\alpha | \dots = (\frac{\sum_{i=1}^n x_i^{y'} x_i^{y'}}{\sigma_\varepsilon^2} + \Sigma_{\alpha_1}^{-1})^{-1}$.

$$190 \quad \sigma_\varepsilon^2 | \mathbf{U}(-\sigma_\varepsilon^2), Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z \sim \Gamma(\frac{m_y n}{2} + a_{\sigma^2}, b_{\sigma^2} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{m^y} (Y_{ij} - x_{ij}^{y'} \alpha_1 + w_{ij}^{y'} u_{1i})^2) \tag{10}$$

191

$$\theta(\alpha_2 | U_{(-\alpha_2)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^z} \frac{\exp(x_{ij}^{z'} \alpha_2 + w_{ij}^{z'} u_{2i} + Y_j Y_{ij})^{z_{ij}}}{1 + \exp(x_{ij}^{z'} \alpha_2 + w_{ij}^{z'} u_{2i} + Y_j Y_{ij})} \times \mathbb{I}_{p^z}(\alpha_2; \mu_{\alpha_2}, \Sigma_{\alpha_2}),$$

$$\theta(\Phi_1 | U_{(-\Phi_1)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^z} \frac{\exp(\psi_{ij}^{y'} \Phi_1 + \pi^{y'} u_{1i})^{R_{ij}^y}}{1 + \exp(\psi_{ij}^{y'} \Phi_1 + \pi^{y'} u_{1i})} \times \mathbb{I}_{p_{R^y}}(\Phi_1; \mu_{\Phi_1}, \Sigma_{\Phi_1}),$$

$$\theta(\Phi_2 | U_{(-\Phi_2)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^z} \frac{\exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)^{R_{ij}^z}}{1 + \exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)} \times \mathbb{I}_{p_{R^z}}(\Phi_2; \mu_{\Phi_2}, \Sigma_{\Phi_2}),$$

$$D_1 | U_{(-D_1)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z \sim IWishart(\pi^y + \sum_{i=1}^n u_{1i} u_{1i}^T v + q^y + n)$$

$$D_2 | U_{(-D_2)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z \sim IWishart(\pi^z + \sum_{i=1}^n u_{2i} u_{2i}^T v + q^z + n)$$

$$\theta(\pi^y | U_{(-\pi^y)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^z} \frac{\exp(\psi_{ij}^{y'} \Phi_1 + \pi^{y'} u_{1i})^{R_{ij}^y}}{1 + \exp(\psi_{ij}^{y'} \Phi_1 + \pi^{y'} u_{1i})} \times \mathbb{I}_{q_{R^y}}(\pi^y; \mu_{\pi^y}, \Sigma_{\pi^y}),$$

$$\theta(\pi^z | U_{(-\pi^z)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^z} \frac{\exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)^{R_{ij}^z}}{1 + \exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)} \times \mathbb{I}_{q_{R^z}}(\pi^z; \mu_{\pi^z}, \Sigma_{\pi^z}),$$

$$\theta(Y_j | U_{(-Y_j)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^z} \frac{\exp(x_{ij}^{z'} \alpha_2 + w_{ij}^{z'} u_{2i} + Y_j Y_{ij})^{z_{ij}}}{1 + \exp(x_{ij}^{z'} \alpha_2 + w_{ij}^{z'} u_{2i} + Y_j Y_{ij})} \times \mathbb{I}(Y_j; \mu_{Y_j}, \sigma_{Y_j}^2),$$

207 (18)

$$208 \quad \theta(\kappa_j | U_{(-\kappa_j)}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^n \prod_{j=1}^{m^z} \frac{\exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)^{R_{ij}^z}}{1 + \exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)} \times \mathbb{I}(\kappa_j; \mu_{\kappa_j}, \sigma_{\kappa_j}^2),$$

209 (19)

210

$$211 \quad \theta(u_{1i} | U_{(-u_{1i})}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{i=1}^{m^y} \mathbb{I}(Y_{ij}; x_{ij}^{y'} \alpha_1 + w_{ij}^{y'} u_{1i}, \sigma_\varepsilon^2) \times \prod_{j=1}^{m^z} \frac{\exp(\psi_{ij}^{y'} \Phi_1 + \pi^{y'} u_{1i})^{R_{ij}^y}}{1 + \exp(\psi_{ij}^{y'} \Phi_1 + \pi^{y'} u_{1i})} \times \mathbb{I}_{q^y}(u_{1i}; 0, D_1),$$

212 (20)

$$213 \quad \theta(u_{2i} | U_{(-u_{2i})}, Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z) \propto \prod_{j=1}^{m^z} \frac{\exp(x_{ij}^{z'} \alpha_2 + w_{ij}^{z'} u_{2i} + Y_j Y_{ij})^{Z_{ij}}}{1 + \exp(x_{ij}^{z'} \alpha_2 + w_{ij}^{z'} u_{2i} + Y_j Y_{ij})} \times \prod_{j=1}^{m^z} \frac{\exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)^{R_{ij}^z}}{1 + \exp(\psi_{ij}^{z'} \Phi_2 + \pi^{z'} u_{2i} + \kappa_j R_{ij}^y)} \times \mathbb{I}_{q^z}(u_{2i}; 0, D_2),$$

214 (21)

$$215 \quad Y_{ij}^{mis} |_{\mathbb{U}(-Y_{ij}^{mis}), Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z} \sim N(x_{ij}^{y'} \alpha_1 + w_{ij}^{y'} u_{1i}, \sigma_\varepsilon^2)$$

216 (22)

$$217 \quad Z_{ij}^{mis} |_{\mathbb{U}(-Z_{ij}^{mis}), Y_{ij}^{obs}, Z_{ij}^{obs}, R_{ij}^y, R_{ij}^z} \sim Ber(\lambda_{ij}).$$

218 (23)

219 4. Simulation studies

220 Some simulation studies are conducted to assess the performance of the proposed model. The first
 221 simulation is designed with assuming no association between two responses, and between
 222 measurements and missingness processes. The second simulation assumes different associations
 223 between biomarkers, also between biomarkers and missingness processes. Each simulation has
 224 100 random samples and each sample includes 500, 1,000, and 3,000 subjects.

225 For each individual, the two longitudinal measurements are generated from the following joint
 226 model,

$$227 \quad Y_{ij} = x_{ij}^y \alpha_1 + u_{1i} + \varepsilon_{1ij}, \quad i = 1, 2, 3, \dots, n, \quad j = 1, 2, 3, \dots, m^y, \quad (24)$$

$$228 \quad Y_{ij}^* = x_{ij}^z \alpha_2 + u_{2i} + Y_j Y_{ij} + \varepsilon_{ij}^*, \quad i = 1, 2, 3, \dots, n, \quad j = 1, 2, 3, \dots, m^z, \quad (25)$$

229 where $\varepsilon_{ij}^* \sim \text{logistic}(0, 1)$, $Z_{ij} = 1$ for $Y_{ij}^* \geq 0$, and $Z_{ij} = 0$ for $Y_{ij}^* < 0$.

230 For missingness processes, following models are assumed to generate data,

$$231 \quad R_{ij}^y | u_{1i} \sim \text{Ber}(\lambda_{ij}^{R^y}), \quad (26)$$

232 where, $\text{logit}(\lambda_{ij}^{R^y}) = \psi_{ij}^y \Phi_1 + \pi^y u_{1i}$.

$$233 \quad R_{ij}^z | u_{2i} \sim \text{Ber}(\lambda_{ij}^{R^z}), \quad (27)$$

234 where, $\text{logit}(\lambda_{ij}^{R^z}) = \psi_{ij}^z \Phi_2 + \pi^z u_{2i} + \kappa_j R_{ij}^y$.

$$235 \quad x_{ij}^y \alpha_1 = \alpha_{10} + \alpha_{11} t_{ij} + \alpha_{12} x_{1i} + \alpha_{13} x_{2i} + \alpha_{14} x_{3i} + \alpha_{15} x_{4i} + \alpha_{16} x_{5i} + \alpha_{17} x_{6i} + \alpha_{18} x_{7i}$$

$$236 \quad x_{ij}^z \alpha_2 = \alpha_{20} + \alpha_{21} s_{ij} + \alpha_{22} x_{1i} + \alpha_{23} x_{2i} + \alpha_{24} x_{3i} + \alpha_{25} x_{4i} + \alpha_{26} x_{5i} + \alpha_{27} x_{6i} + \alpha_{28} x_{7i}$$

$$237 \quad \psi_{ij}^y \Phi_1 = \Phi_{10} + \Phi_{11} X_{3i}, \quad \psi_{ij}^z \Phi_2 = \Phi_{20} + \Phi_{21} X_{4i}, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \quad u_{1i} \sim N(0, \sigma_{u_1}^2), \quad u_{2i} \sim N(0, \sigma_{u_2}^2), \quad t_j = 0, 1, 2, 3, \dots,$$

238 m^y , and $s_j = 0, 1, 2, 3, \dots, m^z$.

239 To make simulation feasible, each dataset included small to large individuals with longitudinal
 240 measurements. Follow-up visits are scheduled at time = $\{0, 1, 2, 3, 4, 5\}$. The true values
 241 considered for $\alpha_1 = \alpha_2 = (1, -1, 0, 2, 1, 1, 1, -1, -1)'$ and $\Phi_1 = \Phi_2 = (2, 1)'$. $x_{11}, x_{12}, x_{13}, x_{14}$ are
 242 generated from $N(0, 1)$, and x_{15}, x_{16}, x_{17} are generated from $\text{binom}(n, 1, 0.1)$. $u_i = (u_{1i}, u_{2i})$ is
 243 generated from $MVN = (0, D)$, where

$$244 \quad D = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

245

246 For the SPM incorporating non-monotone, non-ignorable missing data mechanism, model (26)
247 and model (27) are formulated following Fitzmaurice et al. [19]. We first considered independence
248 $(Y_j, \kappa_j) = (0, 0)'$. Sample sizes are assessed by considering $n = 500$, $n = 1,000$, and $n = 3,000$
249 individuals with different association parameters as $(Y_j, \kappa_j) = (1.000, -1.000)'$, $(Y_j, \kappa_j) =$
250 $(0.500, -0.500)'$, $(Y_j, \kappa_j) = (-0.500, -0.500)'$.

251 Sensitivity analysis of posterior distribution is assessed by selecting different prior distributions.
252 In the case of informative priors, the structure of prior distributions take main focus during the
253 sensitivity analysis. Non-informative priors are assessed based on changes in posterior inference.
254 All our parameters are checked for different parameters' values to assess robustness of posterior
255 means. In addition, sensitivity analysis can also be done for missingness mechanisms, where the
256 assessment for ignorable and non-ignorable missingness can be evaluated. We do not apply
257 sensitivity analysis for different missingness mechanisms, as it is beyond the scope of our study
258 objective.

259 A computational procedure to estimate parameters in proposed joint model is conducted using
260 Gibbs sampling by WinBUGS software. Two parallel MCMC sampling chains run with different
261 starting values, convergence of chains is examined by trace plots and with diagnostic statistics
262 suggested by Gelman et al. [18]. Posterior estimates are based on 10,000 iterations after discarding
263 5,000 of the burn-in period.

264 Data are generated under two scenarios: biomarker variability is not associated with each other
265 ($Y = 0$), and both biomarkers are associated with each other ($Y = \pm 0.5$, or ± 1.0). With that,
266 it is also evaluated whether either missingness process is associated with the measurements process
267 ($\kappa = \pm 0.5$, or ± 1.0) or not. Each dataset is then analyzed using our proposed joint model.

268 Based on the results presented in Table 1, it is to be noted that the performance of the proposed
 269 joint model is good. Consistency of parameter estimates is evaluated by increasing sample size
 270 which causes reduced bias and standard error for respective parameters. The same interpretation
 271 of parameter estimates holds for zero to non-zero association parameters.

272

273 Table 1: Results of the simulation study for binary and continuous longitudinal biomarkers data incorporating NMAR
 274 based on joint modelling for the simulation study. Estimate (Est.), standard error (SE), Bias for $N = 100$ simulated
 275 data with sample size 500, 1,000, and 3,000 considering association $(Y_j, \kappa_j) = (0.5, -0.5)'$.

Parameter	Real	<i>n=500</i>		<i>n=1,000</i>		<i>n=3,000</i>	
		Est(SE)	Bias	Est(SE)	Bias	Est(SE)	Bias
α_{10}	1.000	1.130(0.260)	0.130	0.895(0.161)	-0.105	1.088(0.101)	0.088
α_{11}	-1.000	-1.142(0.196)	-0.142	-0.985(0.119)	0.015	-1.006(0.075)	-0.006
α_{12}	0.000	0.275(0.118)	0.275	0.112(0.073)	0.112	-0.062(0.043)	-0.062
α_{13}	2.000	2.587(0.397)	0.587	1.896(0.242)	-0.104	1.969(0.152)	-0.031
α_{14}	1.000	1.066(0.209)	0.066	0.943(0.135)	-0.057	0.971(0.082)	-0.029
α_{15}	1.000	1.093(0.215)	0.093	1.028(0.139)	0.028	1.016(0.086)	0.016
α_{16}	1.000	1.133(0.366)	0.133	1.586(0.257)	0.586	0.962(0.151)	-0.038
α_{17}	-1.000	-0.622(0.402)	0.378	-1.121(0.288)	-0.121	-0.991(0.162)	0.009
α_{18}	-1.000	-1.169(0.376)	-0.169	-0.863(0.264)	0.137	-0.926(0.168)	0.074
α_{20}	1.000	1.056(0.062)	0.056	1.050(0.043)	0.050	1.044(0.025)	0.044
α_{21}	-1.000	-1.007(0.005)	-0.007	-0.998(0.004)	0.002	-1.000(0.002)	0.000
α_{22}	0.000	-0.045(0.021)	-0.045	-0.033(0.030)	-0.033	-0.005(0.002)	-0.005
α_{23}	2.000	1.976(0.039)	-0.024	1.993(0.028)	-0.007	1.999(0.015)	-0.001
α_{24}	1.000	0.978(0.037)	-0.022	0.980(0.029)	-0.020	1.007(0.015)	0.007
α_{25}	1.000	1.036(0.039)	0.036	1.029(0.027)	0.029	1.006(0.016)	0.006
α_{26}	1.000	1.076(0.122)	0.076	1.069(0.092)	0.069	1.022(0.053)	0.022
α_{27}	-1.000	-1.102(0.132)	-0.102	-1.040(0.095)	-0.040	-1.007(0.050)	-0.007
α_{28}	-1.000	-0.765(0.120)	0.235	-1.159(0.087)	-0.159	-1.020(0.051)	-0.020
Φ_{10}	2.000	1.845(0.136)	-0.155	1.992(0.101)	-0.008	1.997(0.060)	-0.003
Φ_{11}	1.000	1.039(0.077)	0.039	1.030(0.058)	0.030	1.008(0.033)	0.008

Φ_{20}	2.000	1.967(0.080)	-0.031	1.989(0.058)	-0.011	1.991(0.034)	-0.009
Φ_{21}	1.000	0.973(0.078)	-0.027	0.990(0.056)	-0.010	0.999(0.032)	-0.001
σ_{ε}^2	1.000	1.029(0.029)	0.029	1.023(0.020)	0.023	1.006(0.012)	0.006
$\sigma_{u_1}^2$	1.000	0.884(0.074)	-0.116	0.902(0.056)	-0.098	0.989(0.031)	-0.011
$\sigma_{u_2}^2$	1.000	1.562(0.556)	0.562	1.193(0.284)	0.193	1.140(0.191)	0.140
Y_1	0.500	0.341(0.189)	-0.159	0.471(0.118)	-0.029	0.495(0.075)	-0.005
Y_2	0.500	0.276(0.187)	-0.224	0.428(0.119)	-0.072	0.502(0.076)	0.002
Y_3	0.500	0.614(0.239)	0.114	0.559(0.178)	0.059	0.521(0.084)	0.021
Y_4	0.500	0.534(0.249)	0.034	0.526(0.188)	0.026	0.512(0.101)	0.012
Y_5	0.500	0.538(0.425)	0.038	0.532(0.224)	0.032	0.509(0.116)	0.009
Y_6	0.500	0.603(0.353)	0.103	0.554(0.283)	0.054	0.513(0.353)	0.013
κ_1	-0.500	-0.562(0.177)	-0.062	-0.556(0.134)	-0.056	-0.517(0.077)	-0.017
κ_2	-0.500	-0.351(0.182)	0.149	-0.442(0.131)	0.058	-0.490(0.078)	0.010
κ_3	-0.500	-0.428(0.189)	0.072	-0.449(0.132)	0.051	-0.496(0.076)	0.004
κ_4	-0.500	-0.441(0.180)	0.059	-0.450(0.133)	0.050	-0.498(0.077)	0.002
κ_5	-0.500	-0.514(0.181)	-0.014	-0.510(0.132)	-0.010	-0.506(0.078)	-0.006
κ_6	-0.500	-0.521(0.185)	-0.021	-0.516(0.132)	-0.016	-0.507(0.078)	-0.007
π^y	1.000	1.063(0.043)	0.063	0.998(0.035)	-0.002	0.998(0.029)	-0.002
π^z	1.000	1.183(0.711)	0.183	1.058(0.127)	0.058	1.037(0.118)	0.037

276

277

278 5. Prostate cancer data analysis

279 The primary objective of prostate cancer (PC) data analysis is to detect any potential association
 280 between PSA and ALP biomarkers, and to simultaneously analyze both. PC data were collected
 281 from $n = 1504$ patients who had at least 2 measurements of PSA, and ALP. Patients' *Age*,
 282 *Platelets*, *BMI*, *Bilirubin*, *Gleason Score*, *Grade* are recorded along with the prescribed *Drug*.
 283 Following joint model is considered,

284
$$\log PSA_{ij}|u_{1i} \sim N(\mu_{ij}, \sigma^2), \quad (28)$$

285
$$ALP_{ik}|u_{2i} \sim Ber(\lambda_{ik}), \quad (29)$$

286

287
$$\begin{aligned} \mu_{ij} &= \alpha_{10} + \alpha_{11}t_j + \alpha_{12}Age_i + \alpha_{13}Platelets_i + \alpha_{14}BMI_i + \alpha_{15}Bilirubin_i + \alpha_{16}GleasonScore_i \\ &+ \alpha_{17}Drug_i + u_{1i} \end{aligned}$$

288 and

289
$$\begin{aligned} \logit(\lambda_{ij}) &= \alpha_{20} + \alpha_{21}t_j + \alpha_{22}Age_i + \alpha_{23}Platelets_i + \alpha_{24}BMI_i + \alpha_{25}Bilirubin_i + \alpha_{26} \\ &GleasonScore_i + \alpha_{27}Drug_i + \gamma_j \log PSA_{isk} + u_{2i} \end{aligned}$$

290

291 In addition, for non-ignorable missingness mechanisms following models are proposed,

292
$$R_{ij}^{\log PSA} \sim Ber(\lambda_{ij}^{\log PSA}), \quad (30)$$

293 where, $\logit(\lambda_{ij}^{\log PSA}) = \Phi_{10} + \Phi_{11}Platelets_i + \Phi_{12}Bilirubin_i + \pi^{\log PSA}u_{1i}$

294
$$R_{ik}^{ALP} \sim Ber(\lambda_{ik}^{ALP}), \quad (31)$$

295 where, $\logit(\lambda_{ik}^{ALP}) = \Phi_{20} + \Phi_{21}Platelets_i + \Phi_{22}Bilirubin_i + \kappa_k R_{isk}^{\log PSA} + \pi^{ALP}u_{2i}$,

296 $u_{1i} \sim N(0, \sigma_{u_1}^2)$, and $u_{2i} \sim N(0, \sigma_{u_2}^2)$

297 Two parallel MCMC chains runs with different initial values for 10,000 iterations and discarded

298 first 5,000 iterations as pre-convergence burn-in. Convergence of MCMC chains is checked using

299 Gelman-Rubin diagnostic test. In addition, trace plots can plot for unknown parameters to assess

300 convergence. To get posterior inference, prior distributions for unknown parameters are selected

301 as, $u_i = (u_{1i}, u_{2i}) \sim MVN((0,0), D_u)$, $i = 1, \dots, n$, where $D_u \sim IWishart(\psi_{D_u}, \nu_{D_u})$, such that the hyper-

302 parameters of $\psi_{\Sigma_u} = I_2$ and $\nu_{\Sigma_u} = 2$ which lead to low-informative priors. It is also to be assumed

303 that $\sigma_{\epsilon}^2 \sim \Gamma(0.1, 0.1)$, the regression coefficients $(\alpha_{i1}, \alpha_{i2}, \psi_{i1}, \psi_{i2})'$ are fixed effects unknown

304 parameters and the prior distributions for them are $N(0, 1000)$, $\gamma_j, j = 1, \dots, m^y$ is the associated

305 parameter of the continuous longitudinal log(PSA) on the binary longitudinal ALP at time j and
306 the prior distributions for it is $N(0,1000)$. Prior distribution for associated parameter $\kappa_j, j = 1, \dots, m^z$
307 is $N(0,1000), \pi^{\log PSA}, \pi^{ALP} \sim N(0,1000)$.

308 For model selection, we apply the deviance information criterion (DIC), which identifies the most
309 suitable model by balancing between likelihood and parameters' numbers. Posterior means and
310 the number of parameters are used in this criteria to find the best-fitted model [20].

$$311 \quad \quad \quad DIC = \bar{D}(\emptyset) + \bar{P}, \quad (32)$$

312 Where, $\bar{P} = \bar{D}(\emptyset) - D(\bar{\emptyset})$, and $D(\emptyset) = -2\log(f(\frac{w}{\emptyset}))$.

313 $\bar{D}(\emptyset)$ and $\bar{\emptyset}$ are posterior estimates of $D(\emptyset)$ and \emptyset . $w = w_i = (w_1, w_2, w_3, \dots, w_n)$ is full data with
314 marginal density, $f(\frac{w}{\emptyset})$ and \emptyset are vectors of model parameters. The lowest DIC leads us to the best
315 model fit, that is our proposed joint model.

316 The parameter estimates and their 95% posterior intervals are presented in Table 2 and summarized
317 for joint model: PSA level decreases with respect to time after treatment, and it is a good sign to
318 evaluate efficiency of prescribed treatment when this decrease is significant. PSA alone is not an
319 adequate biomarker, ALP measurements are also taken into account. ALP increases with respect
320 to time for individuals, but this increase has not had any significant effect. The results indicate that
321 with increasing age, both PSA and ALP increase. This positive increase is significant for PSA but
322 non-significant for ALP. Platelets are another predictor of increasing PSA significantly, but due to
323 one unit increase in platelets, ALP level decreases non-significantly with 0.026 on average. With
324 increasing BMI, patients' PSA and ALP both tend to significantly decrease. One unit increase in
325 bilirubin measurement makes a significant increase of 0.201, and 0.617 in PSA and ALP
326 measurements, respectively. The results show that PSA and ALP measurements are lower among
327 those patients with a Gleason score greater than or equal to (4+3) as compared to those whose

328 score is lower than and equal to (3+4), it is due to treatment effect. This study's results revealed
 329 that patients who received ALT, prostatectomy, and their combinations have lower PSA and ALP
 330 measurements as compared to those who received EBRT; and this is a good sign to prove drug
 331 efficacy. There exists a positive correlation between PSA and ALP measurements that can be
 332 revealed with associated parameters γ ; while a negative association exists between missingness
 333 and measurement processes. Results revealed that due to the existence of a positive association
 334 both PSA and ALP measurements should be simultaneously taken on PC patients to get insights
 335 about PC progression. In addition, missingness must be incorporated into data to avoid loss of
 336 information.

337

338 Table 2: Parameter estimates (Mean), standard deviation (Sd), and 95% CI (credible interval) of PC data, by applying
 339 our proposed joint SPM, and separate models.

	JM		Separate models	
	Mean(Sd)	95%CI	Mean(Sd)	95%CI
Continuous log(PSA) sub-model				
α_{11} (Intercept)	2.221(0.046)	(2.134, 2.314)	2.167(0.049)	(2.072, 2.261)
α_{12} (Time)	-6.597(0.080)	(-6.752,-6.437)	-6.573(0.081)	(-6.732, -6.415)
α_{13} (Age)	0.290(0.024)	(0.242,0.337)	0.303(0.025)	(0.254, 0.352)
α_{14} (Platelets)	0.251(0.036)	(0.181,0.319)	0.217(0.037)	(0.148, 0.287)
α_{15} (BMI)	-0.206(0.024)	(-0.253,-0.160)	-0.197(0.024)	(-0.246, -0.149)
α_{16} (Bilirubin)	0.201(0.036)	(0.129,0.271)	0.193(0.037)	(0.121, 0.266)
α_{17} (Gleason Score)	0.862(0.161)	(0.543,1.171)	0.862(0.179)	(0.514, 1.213)
α_{18} (Drug)	0.692(0.163)	(0.384,1.009)	0.752(0.182)	(0.399, 1.108)
Φ_{10} (Intercept)	0.475(0.029)	(0.418,0.533)	0.473(0.030)	(0.416, 0.530)
Φ_{11} (Platelets)	-0.003(0.001)	(-0.106,0.101)	0.005(0.003)	(-0.104, 0.101)

Φ_{12} (Bilirubin)	0.041(0.014)	(-0.064,0.149)	0.019(0.010)	(-0.086, 0.124)
σ_{u1}^2	0.334(0.031)	(0.275, 0.396)	0.356(0.032)	(0.295, 0.420)
σ^2	1.399(0.034)	(1.334,1.467)	1.371(0.035)	(1.309, 1.437)
π^{logPSA}	-0.229(0.060)	(-0.346,-0.113)	-0.229(0.062)	(-0.347,-0.113)
Binary ALP sub-model				
α_{21} (Intercept)	-1.741(0.335)	(-2.369,-1.076)	-0.759(0.379)	(-0.912, -0.601)
α_{22} (Time)	0.182(0.077)	(-1.765, 2.014)	-2.652(0.282)	(-3.236, -2.144)
α_{23} (Age)	0.019(0.009)	(-0.108,0.125)	0.062(0.043)	(-0.022, 0.147)
α_{24} (Platelets)	-0.026(0.015)	(-0.173,0.121)	-0.082(0.055)	(-0.19, 0.026)
α_{25} (BMI)	-0.148(0.029)	(-0.244,-0.052)	-0.135(0.039)	(-0.209, -0.058)
α_{26} (Bilirubin)	0.617(0.014)	(0.457,0.782)	0.487(0.047)	(0.397, 0.58)
α_{27} (Gleason Score)	0.490(0.319)	(-0.230,1.095)	0.648(0.328)	(-0.033, 1.263)
α_{28} (Drug)	-0.587(0.313)	(-1.138,0.048)	-0.347(0.335)	(-0.963, 0.343)
Φ_{20} (Intercept)	1.103(0.054)	(0.997, 1.208)	1.100(0.054)	(0.990, 1.212)
Φ_{21} (Platelets)	0.016(0.009)	(-0.100, 0.132)	0.013(0.010)	(-0.100, 0.140)
Φ_{12} (Bilirubin)	0.212(0.063)	(0.090, 0.337)	0.210(0.064)	(0.089, 0.340)
σ_{u2}^2	0.104(0.033)	(0.054, 0.173)	0.101(0.035)	(0.050, 0.180)
π^{ALP}	-0.348(0.126)	(-0.605,-0.107)	-0.343(0.129)	(-0.600, -0.112)
Associated Parameters				
γ_1	0.511(0.147)	(0.219,0.789)		
γ_2	0.029(0.018)	(-0.266,0.308)		
γ_3	0.217(0.150)	(-0.080,0.506)		
γ_4	0.364(0.161)	(0.048,0.677)		
γ_5	1.455(0.335)	(0.829,2.160)		
κ_1	-0.538(0.023)	(-0.585,-0.493)		
κ_2	0.012(0.001)	(-0.048,0.075)		

κ_3	-0.159(0.041)	(-0.239,-0.078)
κ_4	-0.382(0.070)	(-0.520,-0.243)
κ_5	-1.011(0.214)	(-1.451,-0.608)
<i>DIC</i>	33308	53591

340

341

342 **6. Conclusion**

343 The primary goal of this manuscript was to jointly model different types of longitudinal outcomes
344 incorporating non-ignorable missingness. One can directly apply mixed modelling techniques with
345 considering missingness as an ignorable phenomenon. However, follow-up studies could get over
346 the hurdles of loss of information in case of missingness by proposing SPM for mixed-longitudinal
347 outcomes with missingness. For each missing data process, a logit model is proposed via a latent
348 variable to depict the tendency of change for individuals. SPM proposed to join longitudinal and
349 missingness models using random effects. This paper presents a joint modelling approach to
350 analyze longitudinal biomarkers measurements in the presence of non-ignorable missing data due
351 to both intermittent and monotone missingness. In our PC dataset, the non-monotone missingness
352 pattern exists due to patients who missed visits, and monotone missingness data are from those
353 patients who left the follow-up and never came back to complete the treatment process. We use
354 logistic models to describe the missingness patterns. Our proposed model is adequate to account
355 for the association between measurements and missingness processes, and it can be extended to
356 models with more than two mixed longitudinal outcomes incorporating more than two missing
357 outcomes processes.

358 Correlated random effects provide inference for non-ignorable missing data. However, the non-
359 ignorable missingness assumption is untestable for the data at hand. In our collected PC dataset,

360 non-ignorable assumption is verified as missingness occurred due to lack of treatment efficacy.
361 Therefore, we incorporated missingness process in data analysis. However, researchers must put
362 extra care to check assumptions about missingness mechanism, if the cause of missing data is not
363 internally related to responses; local sensitivity analysis should be performed.
364 This study emphasis that PC patients must be monitored for PSA and ALP simultaneously. It is
365 important to take into consideration both PSA and ALP levels as both can influence the health of
366 PC patients. In addition, missing observations must be incorporated in data analysis to get full
367 information about patients' health and wellbeing. An elevated level of PSA shows the non-
368 effectiveness of particular treatment, and an elevated level of ALP depicts the spread of PC tumor
369 cells.

370

371 **Acknowledgements**

372 None.

373

374 **References**

- 375 1. Buhule OD, Wahed AS, Youk AO. Bayesian hierarchical joint modeling of repeatedly
376 measured continuous and ordinal markers of disease severity: Application to Ugandan
377 diabetes data. *Stat Med.* 2017;36(29):4677-91.
- 378 2. Rubin DB. Inference and missing data. *Biometrika.* 1976;63:581-92.
- 379 3. Gao F, Miller JP, Xiong C, Beiser JA, Gordon M. The Ocular Hypertension Treatment
380 Study (OHTS) Group.: A joint-modeling approach to assess the impact of biomarker
381 variability on the risk of developing clinical outcome. *Stat Methods Med.* 2011;20(1):83-
382 100.

- 383 4. Alam K, Maity A, Sinha SK, Rizopoulos D, Sattar A. Joint modeling of longitudinal
384 continuous, longitudinal ordinal, and time-to-event outcomes. *Lifetime Data*
385 *Analysis*. 2021;27:64-90.
- 386 5. Bahrami Samani E, Ganjali M. Bayesian latent variable model for mixed continuous and
387 ordinal responses with possibility of missing responses. *J Appl Stat*. 2011;38(6):1103-16.
- 388 6. Dunson DB, Herring AH. Bayesian latent variable models for mixed discrete
389 outcomes. *Biostatistics*. 2005;6(1):11-25.
- 390 7. Catalano P, Ryan LM. Bivariate latent variable models for clustered discrete and continuous
391 outcomes. *J A Stat Ass*. 1992;87(419):651-8.
- 392 8. Leon AR, Carriere KC. On the one sample location hypothesis for mixed bivariate data.
393 *Commun. Statist. Theory and Meth*. 2000;29(11):2573-81.
- 394 9. Liu X, Daniels MJ, Marcus B. Joint Models for the Association of Longitudinal Binary and
395 Continuous Processes With Application to a Smoking Cessation Trial. *J Am Stat*
396 *Assoc*. 2009;104(486):429-38.
- 397 10. Li Q, Pan J, Belcher J. Bayesian inference for joint modelling of longitudinal continuous,
398 binary and ordinal events. *Stat Methods Med Res*. 2016;25(6):2521-40.
- 399 11. Kürüm E, Hughes J, Li R. A semivarying joint model for longitudinal binary and continuous
400 outcomes. *Canadian J Stat*. 2016;44(1):44-57.
- 401 12. Stubbendick AL, Ibrahim JG. Maximum likelihood methods for nonignorable missing
402 responses and covariates in random effects models. *Biometrics*. 2003;59(4):1140-50.
- 403 13. Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. *Stat*
404 *Med*. 2004;23(9):1455-97.

- 405 14. Gaskins JT, Daniels MJ, Marcus BH. Bayesian methods for nonignorable dropout in joint
406 models in smoking cessation studies. *J Am Stat Assoc.* 2016;111(516):1454-65.
- 407 15. Little RJ, Schluchter M. Maximum likelihood estimation for mixed continuous and
408 categorical data with missing values. *Biometrika.* 1985;72(3):497-512.
- 409 16. Fitzmaurice GM, Laird NM. Regression models for mixed discrete and continuous
410 responses with potentially missing values. *Biometrics.* 1997;53:110-22.
- 411 17. Olkin I. Tate RF. Multivariate correlation models with mixed discrete and continuous
412 variables. *Annals Math Stat.* 1997;32:448-65.
- 413 18. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis.*
414 CRC Press; 2013.
- 415 19. Fitzmaurice GM, Laird NM. Ware JH. *Applied Longitudinal Analysis.* Wiley, Hoboken;
416 2012.
- 417 20. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model
418 complexity and fit. *J R Stat Soc Ser B Stat Methodol.* 2002;64(4):583-639.