J. Statist. Res. Iran **15** (2018): 213–236
DOI: 10.29252/jsri.15.2.213

Bayesian Sample Size Determination for Joint Modeling of Longitudinal Measurements and Survival Data

Taban Baghfalaki

Tarbiat Modares University

Received: 10/22/2018 Approved: 8/4/2019

Abstract. A longitudinal study refers to collection of a response variable and possibly some explanatory variables at multiple follow-up times. In many clinical studies with longitudinal measurements, the response variable, for each patient is collected as long as an event of interest, which considered as clinical end point, occurs. Joint modeling of continuous longitudinal measurements and survival time is an approach for accounting association between two outcomes which frequently discussed in the literature, but design aspects of these models have been rarely considered. This paper uses a simulation-based method to determine the sample size from a Bayesian perspective. For this purpose, several Bayesian criteria for sample size determination are used, of which the most important one is the Bayesian power criterion (BPC), where the determined sample sizes are given based on BPC. We determine the sample size based on treatment effect on both outcomes (longitudinal measurements and survival time). The sample size determination is performed based on multiple hypotheses. Using several examples, the proposed Bayesian methods are illustrated and discussed. All the implementations are performed using R2OpenBUGS package and R 3.5.1 software.

Keywords. Bayesian analysis; bayesian power criterion; longitudinal data; sample size determination; survival data; shared parameter model.

MSC 2010: 62N99; 62D99; 62J99.

1 Introduction

A crucial aspect in planning medical studies to test one or more hypotheses tests is sample size determination. The number of subjects in a medical study or clinical trial should be large enough to provide a sufficient power to answer the research question under investigation and should avoid the waste in resources and the risk of exposing the excessive number of individuals to experimental treatments. Choosing the appropriate sample size increases the chance of detecting an effect of treatment and ensures that the study is both ethical and cost-effective.

In many longitudinal studies, longitudinal measurements are gathered alongside time to event (survival) data. In these kinds of studies, an event of interest such as death or disease progression is considered as an important part of the study. Two important characteristics exist in these studies: the first is missingness in the longitudinal measurements and the second is censoring in the survival data. For designing these studies, these two types of properties should be considered. Many studies are performed for modeling these kind of data (Guo and Carlin, 2004; Tsiatis and Davidian, 2004; Baghfalaki and Ganjali, 2015; Baghfalaki et al., 2013, 2014a,b, 2017 among other), but there is rarely studies in sample size determination based on these simultaneous outcomes (Chen et al., 2011).

Often the required sample size in these studies are determined based on those used for longitudinal studies. Sample size determination in classical paradigm are discussed in many studies. The classical approaches can be implemented by some packages such as RMASS2 (Hedeker et al., 1999), PASS (Hintze, 2000), POWERLIB (Johnson et al., 2009) and Optimal Design (Raudenbush, 1997).

Another approach for determining sample size is based on the Monte Carlo simulation (Williams et al., 2007). In this framework, previous information is used to repeatedly draw samples of the type to be employed in the study from a fully specified population model. Then, it is simply calculated for each simulated sample and their properties are evaluated over a large number of repeated samples. This approach might be the only viable alternative when complex analytical methods are to be employed and closed formula are not available (Landau and Stahl, 2013). The above-mentioned approaches are also classified as classical approaches. Another perspective for determining sample size is the Bayesian one which is rarely discussed when calculating sample size in the context of joint modeling of longitudinal

and survival data, but frequently is used for cross-sectional data (Joseph and Belisle, 1997; Pezeshk, 2003; Christensen et al., 2012; Wang and Gelfand, 2002 and other references therein).

Wang and Gelfand (2002) proposed a simulation-based approach to Bayesian sample size determination. Their proposed approach is frequently used for determining sample size in cross-sectional studies (Dendukuri et al., 2004; Gamalo et al., 2014; Mukherjee et al., 2010; Cheng et al., 2010a,b; Dendukuri et al., 2010). However, this approach is not used for determining sample size in joint modeling of longitudinal measurements and survival data. In this paper, Bayesian sample size determination for joint modeling of longitudinal measurements and survival data is discussed based on the proposed method of Wang and Gelfand (2002). The sample sizes are determined based on the variation of the rate of censoring or missingness and the different configurations for dependence between two outcomes. Also, multiple hypothesis testing scenarios are discussed for sample size determination. Some examples are used for illustration of the proposed approach. The use of Bayesian paradigm in joint modeling of longitudinal and survival data provides some flexibility and simplicity for the analysis. As, for analyzing this kind of data set a Bayesian paradigm may be used and so a Bayesian sample size determination (such as the approach mentioned in this paper) may provide a more valid sample size.

This paper is organized as follows: In the next section, joint modeling of longitudinal measurements and survival data from a Bayesian perspective is discussed. Section 3 discusses Bayesian sample size determination for joint modeling of longitudinal data and survival times. Section 4 covers Bayesian sample size determination for joint modeling of longitudinal measurements and survival data using multiple hypotheses testing and Sections 5 contains some examples to investigate the approach. Some conclusions are given in the last section.

2 Joint Modeling of Longitudinal Measurements and Survival Data from a Bayesian Perspective

Let $Y_i(s)$ denote the longitudinal outcome at time point s for the ith individual. The observed times are $s_{ij},\ j : 1, 2, ..., n_i$. Thus, the longitudinal outcomes for the ith individual consist of measurements $\boldsymbol{Y}_i = \{Y_i(s_{ij}),\ j = 1, 2, ..., n_i\}$.

For the longitudinal process, we consider the following linear mixed effects model:

$$Y_i(s_{ij}) = \boldsymbol{x}_i'\boldsymbol{\beta}_1 + \boldsymbol{z}_i'\boldsymbol{b}_i + \varepsilon_i(s_{ij}); \quad i : 1, 2, ..., n, \quad j : 1, 2, ..., n_i, \tag{1}$$

where components of $\boldsymbol{\varepsilon}_i = (\varepsilon_i(s_{i1}), ..., \varepsilon_i(s_{in_i}))'$ are measurement errors, $\boldsymbol{\beta}_1 = (\beta_{11}, ..., \beta_{1p_1})'$ is a p_1-dimensional vector of longitudinal fixed-effect parameters, $\boldsymbol{b}_i = (b_{i1}, ..., b_{iq})' \overset{iid}{\sim} N_q(\boldsymbol{0}, \boldsymbol{D})$ is a q-dimensional vector of random effects and is independent of $\boldsymbol{\varepsilon}_i$. Also, $\boldsymbol{x}_i'$ and $\boldsymbol{z}_i'$ are, respectively, p_1-dimensional and q-dimensional explanatory variables. We assume that $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N_{n_i}(0, \sigma_\varepsilon^2 I_{n_i})$.

Let T_i denote the observed survival time for the ith individual, $i : 1, 2, ..., n$, which is taken as the minimum of the true event time T_i^* and the censoring time C_i, i.e., $T_i = min(T_i^*, C_i)$. We define a censoring indicator, $\delta_i = I(T_i^* \leqslant C_i)$, which is 0 for right-censored individuals and 1 for completely observed individuals. Therefore, the observed data for the time to event outcome consist of the pairs $\{(T_i, \delta_i),\ i = 1, 2, ..., n\}$. For the time to event process, we consider a parametric proportional hazard model with Weibull baseline hazard in a frailty structure. The use of the Weibull proportional hazard model has some advantages: The exponential model is nested in the Weibull model; also, the Weibull model is the only model that is both a proportional hazard model and an accelerated failure time model. The estimated parameters can easily be transformed from one model to the other and back; for more properties of the parametric proportional hazard model, see Wienke (2011). The hazard function in our proposed model is given by

$$h(t_i|\boldsymbol{w}_i, \boldsymbol{\rho}, \boldsymbol{b}_i) = \nu t_i^{\nu-1} \exp\left\{\boldsymbol{w}_i'\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i\right\}. \tag{2}$$

Thus, the density function of the survival time for the ith individual can be written in the form:

$$(\nu t_i^{\nu-1} \exp\left\{\boldsymbol{w}_i'\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i\right\})^{\delta_i} \times \exp\left\{-t_i^{\nu} \exp\left\{\boldsymbol{w}_i'\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i\right\}\right\}$$

where $\boldsymbol{w}_i$ is a p_2-dimensional vector of explanatory variables, $\boldsymbol{\beta}_2 = (\beta_{21}, ..., \beta_{2p_2})'$ is a p_2-dimensional vector of time to event fixed effect parameters, $\boldsymbol{\rho} = (\rho_1, ..., \rho_q)'$ is a q-dimensional vector of parameters of the time to event process and $\boldsymbol{b}_i \overset{iid}{\sim} N_q(\boldsymbol{0}, \boldsymbol{D})$.

If an event occurs at time T_i, that is ith individual experiences death

or drops out, then after that event longitudinal measurements can not be observed. Therefore, $\boldsymbol{Y}_i$ can be partitioned into $\boldsymbol{Y}_{i,obs} = \{Y_i(s_{ij}) : s_{ij} < T_i, j = 1, 2, ..., n_i\}$ which contains all observed longitudinal measurements for the ith individual before the observed event time T_i, and $\boldsymbol{Y}_{i,mis} = \{Y_i(s_{ij}) : s_{ij} \geqslant T_i, j = 2, ..., n_i\}$ which contains the longitudinal measurements that should have been taken until the end of the study. In this context, some individuals are missing due to dropout or death. If the probability of the event occurring is dependent on the unobserved outcome, then missingness is non-ignorable or dropout is non-random.
In our model, $\boldsymbol{\rho}$ is the non-ignorability parameter. This means that if $\boldsymbol{\rho} = \mathbf{0}$ then the longitudinal measurements and the event time are independent and when $\boldsymbol{\rho} \neq \mathbf{0}$, the degree of non-ignorability depends on its value, that is a large significant value of components of $\boldsymbol{\rho}$ cause non-ignorability.

The Bayesian methodology will be implemented using MCMC techniques for the joint modeling of longitudinal and survival data. It can be formulated in a flexible hierarchical representation as follows:

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \overset{ind.}{\sim} N_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}_1 + \boldsymbol{Z}_i\boldsymbol{b}_i, \sigma_\varepsilon^2 I_{n_i}), \tag{3}$$

$$\boldsymbol{b}_i|\boldsymbol{D} \overset{ind.}{\sim} N_q(\mathbf{0}, \boldsymbol{D}), \tag{4}$$

where $\boldsymbol{X}_i = (\boldsymbol{X}'_{i1}, ..., \boldsymbol{X}'_{in_i})$ and $\boldsymbol{Z}_i = (\boldsymbol{Z}'_{i1}, ..., \boldsymbol{Z}'_{in_i})$ are $p_1 \times n_i$ and $q \times n_i$ matrix of explanatory variables. Also, the survival model based on the parametric proportional hazard model is given by:

$$T_i|\boldsymbol{b}_i \overset{ind.}{\sim} Weibull(\boldsymbol{w}'_i\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i, \nu). \tag{5}$$

It is important to note that the components of $\boldsymbol{b}_i$ are shared between two models (models of Equations 3-5) and joint modeling forms using them. Suppose that $\boldsymbol{y}_{mis} = (\boldsymbol{y}'_{1,mis}, \boldsymbol{y}'_{2,mis}, \ldots, \boldsymbol{y}'_{n,mis})'$, $\boldsymbol{y}_{obs} = (\boldsymbol{y}'_{1,obs}, \boldsymbol{y}'_{2,obs}, ..., \boldsymbol{y}'_{n,obs})'$, $\boldsymbol{b} = (\boldsymbol{b}'_1, \boldsymbol{b}'_2, ..., \boldsymbol{b}'_n)'$ and $\boldsymbol{t} = (t_1, t_2, ..., t_n)'$. Therefore, the likelihood function for complete data $\boldsymbol{Data} = (\boldsymbol{y}'_{mis}, \boldsymbol{y}'_{obs}, \boldsymbol{t}', \boldsymbol{b}')'$ is given by

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{Data}) \propto & \prod_{i=1}^{n} \big[\phi_{n_i}(\boldsymbol{y}_i; \boldsymbol{X}_i\boldsymbol{\beta}_1 + \boldsymbol{Z}_i\boldsymbol{b}_i, \sigma_\varepsilon^2 I_{n_i}) \\ \times & \left(\nu t_i^{\nu-1} \exp\left\{\boldsymbol{w}'_i\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i\right\}\right)^{\delta_i} \\ \times & \exp\left\{-t_i^{\nu} \exp\left\{\boldsymbol{w}'_i\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i\right\}\right\} \\ \times & \phi_q(\boldsymbol{b}_i; \mathbf{0}, \boldsymbol{D})]. \end{aligned} \tag{6}$$

Bayesian specification of the model needs to consider prior distribution for all the unknown parameters, in our modeling $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{D}, \boldsymbol{\rho}, \sigma_\varepsilon^2, \nu)'$. Because of, not having prior information from historical data or previous experiment, we shall try to assign non-informative prior distributions for the parameters. Assuming elements of the parameter vector to be independent, the prior distributions are given by:

$$\begin{aligned} &\boldsymbol{\beta}_1 \sim N_{p_1}(\boldsymbol{\beta}_{01}, \boldsymbol{\Sigma}_{01}), \quad \boldsymbol{\beta}_2 \sim N_{p_2}(\boldsymbol{\beta}_{02}, \boldsymbol{\Sigma}_{02}), \\ &\boldsymbol{D} \sim \mathrm{IW}_q(\boldsymbol{\eta}, \boldsymbol{\psi}), \quad \nu \sim \Gamma(\alpha_\nu, \tau_\nu), \\ &\sigma_\varepsilon^2 \sim \mathrm{I}\Gamma(\alpha_0, \tau_0), \quad \rho_k \sim N(\mu_{\rho_k}, \sigma_{\rho_k}^2), \quad k = 1, ..., q. \end{aligned} \tag{7}$$

The hyper parameters of these priors are selected in such way that they lead to the low-informative prior distributions. The joint posterior density of all unobservable components is given by the combining of the likelihood function (6) and the prior distribution (7):

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{y}_{mis} | \boldsymbol{y}_{obs}, \boldsymbol{t}) \quad \propto \quad & \prod_{i=1}^{n} \big[\phi_{n_i}(\boldsymbol{y}_i; \boldsymbol{X}_i\boldsymbol{\beta}_1 + \boldsymbol{Z}_i\boldsymbol{b}_i, \sigma_\varepsilon^2 I_{n_i}) \\ \times \quad & \phi_q(\boldsymbol{b}_i; \boldsymbol{0}, \boldsymbol{D}) \\ \times \quad & \left(\nu t_i^{\nu-1} \exp\left\{\boldsymbol{w}'\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i\right\}\right)^{\delta_i} \\ \times \quad & \exp\left\{-t_i^{\nu} \exp\left\{\boldsymbol{w}'\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i\right\}\right\}\big] \times \pi(\boldsymbol{\theta}). \end{aligned} \tag{8}$$

MCMC methods such as the Gibbs sampler and Metropolis-Hastings algorithm can be used to draw samples, from which features of marginal posterior distribution of interest can be inferred. The Gibbs sampler works by drawing samples iteratively from conditional posterior distributions deriving from (8).

3 Bayesian Sample Size Determination for Joint Modeling

Based on the Bayesian sample size determination approach proposed by Wang and Gelfand (2002), which is a simulation-based approach, two sets of priors are considered. The first set is called the "fitting" or "analysis" priors which are used for analyzing data. The other set of priors is called "design" or "sampling" sets which drawing upon expertise, we may speculate upon a variety of informative scenarios regarding the unknown parameters

and capture each with a suitable sampling prior. In the first step of the simulation-based approach, one generates the parameters from the design priors then a sample from the generated parameters is obtained. Note that the frequentist/classical viewpoint of the sample size determination approach requires a point estimate of the variance and the smallest meaningful difference (Diggle et al., 2002). This information is elicited based on pilot data or the opinion of experts. The use of design priors in the proposed approach of Wang and Gelfand (2002) replaces this part of the classical sample size determination approach. After generating data using the design priors in this stage, they are analyzed by the fitting priors. In this paper, non-informative priors are used as fitting priors [as mentioned in equation (7)], as we assume there is no prior information or any expert opinion to be used to construct a prior distribution. If there is any prior information to be used to elicit an appropriate prior distribution, one may use it in the step of fitting prior and for sure, if the specification of the prior is correct, one will achieve a better result. In general, significance of a regression coefficient is used to obtain the optimal sample size. This approach can be improved by choosing a threshold other than zero or any other function of the parameters of interest. Let $\boldsymbol{Data}^{(n)} = (\boldsymbol{x}_1^{(n)}, \boldsymbol{x}_2^{(n)}, \boldsymbol{y}^{(n)}, \boldsymbol{t}^{(n)})$ be a generated data set using design priors with sample size n. Also, consider models (1) with $\boldsymbol{x}_i'\boldsymbol{\beta}_1 + \boldsymbol{z}_i'\boldsymbol{b}_i = \beta_{10} + \beta_{11}x_{1i} + \sum_{k=2}^{p_1} \beta_{1k}x_{ki} + b_{0i} + b_{1i}t_j$, where x_1 is the treatment/placebo binary factor of interest and β_{11} is its regression coefficient of interest for determining sample size based on the longitudinal model. Again, consider models (2) with $\boldsymbol{w}_i'\boldsymbol{\beta}_2 = \beta_{20} + \beta_{21}x_{1i} + \sum_{k=2}^{p_2} \beta_{2k}w_{ki}$, where x_1 is the treatment/placebo binary factor of interest and β_{21} is its regression coefficient of interest for determining sample size based on the survival model. In this paper, four criteria are used for determining sample size in joint model: Bayesian power criterion (BPC), average length criterion (ALC), average posterior variance criterion (APVC) and average converge criterion (ACC). In the following, at first β_{11} is used for describing the criteria that form the sample size determination of the longitudinal model in detail. This discussion is the same for determining sample size based on survival model.

BPC Consider hypothesis test $H_0 : \beta_{11} \leqslant 0$ versus $H_1 : \beta_{11} > 0$; that is, the relationship of the binary factor and the response variable is expected to be positive. BPC (Wang and Gelfand, 2002; Baghfalaki, 2019)

based on the sample size n and probabilities α and η is given by

$$E\left[I\left(P(\beta_{11} > 0|\mathbf{Data}^{(n)}) > 1-\alpha\right)\right] \geqslant \eta, \tag{9}$$

where $I(\cdot)$ is an indicator function. By this approach, we are looking for n such that the expected of having high posterior probability of a significant β_{11} (at $1-\alpha$ level) is high (at least η). The common choice for α is 0.01, 0.05 or 0.1, also, η is usually considered to be 0.8 or 0.9. Based on (9), the sample size is determined as the optimal value of n such that $P(\beta_{11} > 0|\mathbf{Data}^{(n)}) > 1-\alpha$ at least $\%100(1-\eta)$. If the hypothesis test is considered as $H_0 : \beta_{11} \geqslant 0$ versus $H_1 : \beta < 0$, the BPC is defined as $E\left[I\left(P(\beta_{11} < 0|\mathbf{Data}^{(n)}) > 1-\alpha\right)\right] \geqslant \eta$.

ALC (Wang and Gelfand, 2002; Baghfalaki, 2019) Consider an equal tails $(1-\alpha)$ Bayesian interval estimation

$$C(\mathbf{Data}^{(n)}) = (L_{\alpha/2}(\mathbf{Data}^{(n)}), U_{1-\alpha/2}(\mathbf{Data}^{(n)})).$$

Given $l \geqslant 0$, the ALC seeks n such that

$$E[U_{1-\alpha/2}(\mathbf{Data}^{(n)}) - L_{\alpha/2}(\mathbf{Data}^{(n)})] \leqslant l \tag{10}$$

i.e. to find an optimal sample size that for it the expected value of width of the interval is smaller than l. Note that a $(1-\alpha)$ highest posterior density (HPD) can replace the equal tails interval.

APVC (Wang and Gelfand, 2002; Baghfalaki, 2019) Given $\varepsilon \geqslant 0$, the APVC seeks n such that $E\left(\mathrm{var}(\beta_{11}|\mathbf{Data}^{(n)})\right) \leqslant \varepsilon$. i.e. such that the expected value of the posterior variance of β_{11} to be as small as possible. Instead of the variance, the average posterior interquartile range or other measures of dispersion might be used. This approach may be regarded as an estimation or hypothesis testing approach.

ACC (Wang and Gelfand, 2002; Baghfalaki, 2019) Suppose $C(\mathbf{Data}^{(n)})$ is a Bayesian interval estimation or HPD with length ℓ. ACC seeks n such that

$$E\left(P(\beta_{11} \in C(\mathbf{Data}^{(n)})|\mathbf{Data}^{(n)})\right) \geqslant 1-\alpha. \tag{11}$$

i.e. find n such that the expected value of posterior probabilities of

coverage of β_{11} to be large (e.g. larger than for example $1-\alpha = 0.9$ or 0.95).

The simulation-based approach of Wang and Gelfand (2002) can be summarized as follows:

1. Specify the value of the effect of β_{11} that is of interest and specify the design and fitting priors.
2. For each sample size n, the following steps are repeated M times:
   (a) Generate values of the unknown parameters from their design priors.
   (b) Simulate the values of covariates from continuous or discrete distributions and the response variables from its distribution.
   (c) Analyse the generated data set of step (2b) using the fitting priors.
   (d) Calculate BPC.
3. Fit a curve or surface through Bayesian power values and find an adequate sample size for a desired power using interpolation. In this paper, the curve is fitted using a polynomial regression.

In step 2(d), one can calculate ACC, APVC, or ALC and determine the sample size based on these criteria. ACC is a controlling criterion, that is, a sample size is obtained such that ACC is relatively large. In the following section, the Bayesian sample size determination using multiple testing for joint modeling is also illustrated. We have used the available software packages OpenBUGS (Spiegelhalter et al., 2003) and R ([sec:hello]https://cran.r-project.org/) for implementation of the models and the method. The program codes for determining the sample sizes are available under request from the authors.

4 Bayesian Sample Size Determination Using Multiple Testing for Joint Modeling

For sample size determination by considering the effect of treatment/placebo on both the longitudinal and survival models, we use multiple testing approach. In the following, we discuss multiple testing for this frameworks.

1. Consider models (1) and (2), such that $\boldsymbol{x}_i'\boldsymbol{\beta}_1 + \boldsymbol{z}_i'\boldsymbol{b}_i = \beta_{10} + \beta_{11}x_{1i} + \sum_{k=2}^{p_1} \beta_{1k}x_{ki} + b_{0i} + b_{1i}t_j$ and $\boldsymbol{w}_i'\boldsymbol{\beta}_2 + \boldsymbol{\rho}'\boldsymbol{b}_i = \beta_{20} + \beta_{21}x_{1i} + \sum_{k=2}^{p_2} \beta_{2k}w_{ki} + \rho_1 b_{0i} + \rho_2 b_{1i}$. Also, let x_1 be the treatment/placebo binary factor of interest and β_{11} and β_{21} are two regression coefficients corresponding to longitudinal model and survival model, respectively. Specify the values of the effect of β_{11} and β_{21} that is of interest and specify the design and fitting priors.

2. For each sample size n, the following steps are repeated M times:

   (a) Generate values of the unknown parameters from their design priors.

   (b) Simulate the values of covariates from continuous or discrete distributions and the response variables from its distribution.

   (c) Analyse the generated data set of step (2b) using the fitting priors.

   (d) If the relationship between x_1 and the response variable is expected to be positive, the BPC is given by

$$E\left[I\left(P(\beta_{11} > 0|\mathbf{Data}^{(n)}) > 1-\alpha\right) I\left(P(\beta_{21} > 0|\mathbf{Data}^{(n)}) > 1-\alpha\right)\right] \geqslant \eta.$$

   If the relationship between x_1 and the longitudinal response variable is expected to be positive, but the relationship between x_1 and the survival variable is expected to be negative, the BPC is given by

$$E\left[I\left(P(\beta_{11} > 0|\mathbf{Data}^{(n)}) > 1-\alpha\right) I\left(P(\beta_{21} < 0|\mathbf{Data}^{(n)}) > 1-\alpha\right)\right] \geqslant \eta.$$

   This criterion can be modified by a threshold other than 0 or any other function of the parameters of interest.

3. Fit a curve or surface through Bayesian power values and find an adequate sample size for a desired power using interpolation. In this paper, the curve is fitted using a polynomial regression.

Note that ALC, APVC and ACC can be defined for each parameter of interest separately. However, one can aggregate them using a function such as the mean. In this paper, the first paradigm, i.e. the consideration of each parameter separately, is adopted.

5 Applied Simulated Examples

In this section, the proposed approaches are illustrated using some simulated examples. The effect of different association parameters, different number of repeated measurements of longitudinal data, various rates of censoring of survival times and the use of the separate model instead of joint model on required sample sizes are illustrated.

5.1 The Effect of Different Association Parameters

Consider a continuous response variable with $n_i = m = 3, 8, \quad i = 1, 2, ..., n$ repeated measurements with n individuals. The treatment indicator is denoted by $Treat$, where $Treat = 1$ if the treatment is given and $Treat = 0$ if a placebo is used. We assume that the linear predictor for longitudinal and survival model are, respectively, $\mu^y_{ij} = \beta_{10} + \beta_{11}Treat_i + \beta_{12}t_j + b_{i0} + b_{i1}t_j$ and $\mu^t_i = \beta_{20} + \beta_{21}Treat_i + \rho_1 b_{i0} + \rho_2 b_{i1}$, where $Treat$ is equally allocated to treatment and placebo groups. Also, 10% rate of censoring is considered and t is considered as observed time point of longitudinal measurements and is a vector such that $t = (2, 6, 8, 10, 12, 14, 16, 18)'$, also, for $m = 3$ the three first components are considered. We assume that $\mathbf{b}_i = (b_{i0}, b_{i1})'$ is the vector of random effects with distribution $N_2(\mathbf{0}, \mathbf{D})$. The design and fitting priors for described joint model (1) and (2) with the above-mentioned linear predictors are given in Table 1. After considering this design prior, the rate of missingness are varied in range of 10% to 40%. The purpose is sample size determination based on β_{11} and β_{21}, so we fix $\beta_{11} = 1$ and $\beta_{21} = -1$ and use two degenerated distributions at these points as design priors for them, but one can consider other values for them. Also, we consider five different association design priors for association parameters ρ_1 and ρ_2 (for simplicity we consider $\rho_1 = \rho_2$). In all the results of this paper, we ran two parallel MCMC chains with different starting values for each 15000 iterations. Then we discarded the first 5000 iterations as preconvergence burn-in and retained 10000 as the posterior analysis. Then we have checked convergence of parameter estimates using Gelman and Rubin diagnostic test for the joint model (Gelman and Rubin, 1992).

Figure 1 shows BPC for different sample sizes for $1 - \alpha = 0.9$ and 0.95. This figure shows that increasing the absolute value of association parameters decreases the BPC. Also, the higher the number of repeated measures is, the lower the needed sample size is for a given level of BPC. Table 2 shows the required sample size for attaining the values of 0.8, 0.9 and 0.95 for BPC

Table 1. Design and fitting priors for sample size determination.

Parameter	Design prior	Fitting prior
β_{10}	N(5,0.3)	N(0,1000)
β_{11}	Degenerated at 1	N(0,1000)
β_{12}	N(1,0.2)	N(0,1000)
σ_ε^2	U(1,3)	IΓ(0.01,0.01)
$\boldsymbol{D}$	IW$(I_2, 2)$	IW$(10I_2, 2)$
β_{20}	N(1,0.25)	N(0,1000)
β_{21}	Degenerated at -1	N(0,1000)
	N(-4,0.2)	N(0,1000)
	N(-2,0.2)	N(0,1000)
ρ_1, ρ_2	N(0,0.2)	N(0,1000)
	N(2,0.2)	N(0,1000)
	N(4,0.2)	N(0,1000)
ν	$N(2, \sqrt{0.4})$	$\Gamma(0.1, 0.1)$

and for $1-\alpha = 0.9$ and 0.95. This table shows that the higher the BPC and $1-\alpha$ are, the higher is the sample size. Also, the results shows that the larger number of repeated measurements lead to the larger values of BPC and therefore, a smaller number of required sample sizes. Figures 2, 3 and 4 show ACC, ALC and APVC for different sample sizes, respectively. Note that in this paper, ACC is used as a controlling criterion, that is, a sample size is obtained such that ACC is relatively large. Figure 2 shows that the values of ACC for all different sample sizes are more than 0.8. Also, Figures 3 and 4 show that the larger sample sizes and the larger number of repeated measurements lead to the smaller values of ALC and APVC. For example, the required sample size for an ALC of β_{11}, when $\rho_1 = \rho_2 = -2$, equal to 2 is 207 for $m = 3$ and 96 for $m = 8$ and for an ALC of β_{21} equal to 2 is 316 for $m = 3$ and 165 for $m = 8$. Also, a level of APVC of β_{11}, when $\rho_1 = \rho_2 = -2$, equal to 0.5 is required sample sizes 364 and 40 for $m = 3$ and $m = 8$, respectively, also, a level of APVC of β_{21}, when $\rho_1 = \rho_2 = -2$, equal to 0.5 is required sample sizes 189 and 153 for $m = 3$ and $m = 8$, respectively.

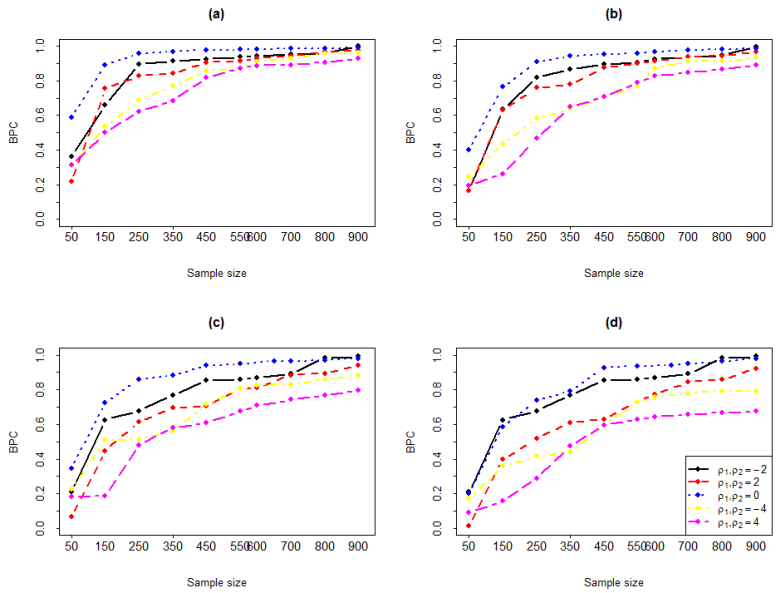

Figure 1. The BPC for different sample sizes, different association parameters and different number of repeated measurements. The panel (a) for $1-\alpha=0.9$ and $m=3$, the panel (b) for $1-\alpha=0.95$ and $m=3$, the panel (c) for $1-\alpha=0.9$ and $m=8$ and the panel (d) for $1-\alpha=0.95$ and $m=8$.

Table 2. Required sample sizes to make a BPC of 0.8, 0.9 and 0.95 using joint modeling.

		$m=3$			$m=8$		
		BPC					
$1-\alpha$	$\rho_1=\rho_2$	0.8	0.9	0.95	0.8	0.9	0.95
	-2	346	730	825	221	305	395
	2	531	829	887	239	329	617
0.9	0	225	311	393	134	204	257
	-4	553	905	924	466	783	885
	4	894	973	998	449	905	973
	-2	453	812	881	256	348	832
	2	705	859	907	298	461	850
0.95	0	297	466	677	171	241	328
	-4	724	947	961	535	746	947
	4	983	1023	1039	538	902	949

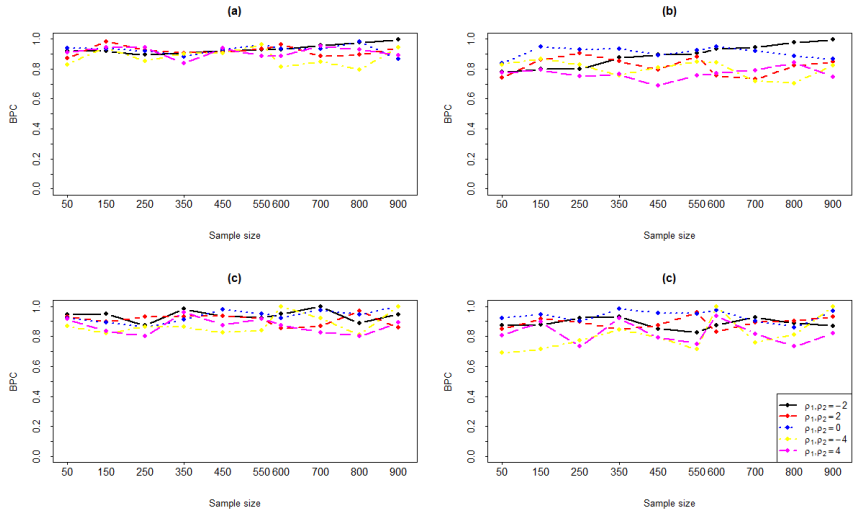

Figure 2. The ACC for different sample sizes, different association parameters and different number of repeated measurements. The panel (a) for β_{11} and $m = 3$, the panel (b) for β_{21} and $m = 3$, the panel (c) for β_{11} and $m = 8$ and the panel (d) for β_{21} and $m = 8$.

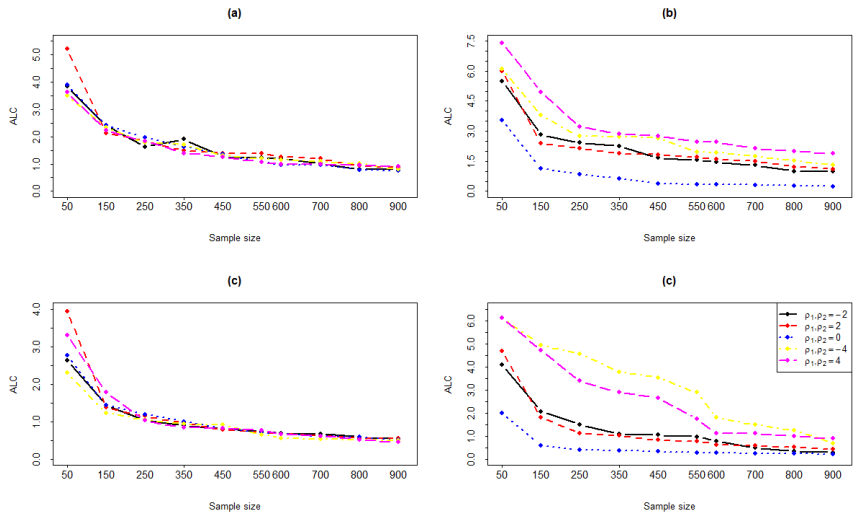

Figure 3. The ALC for different sample sizes, different association parameters and different number of repeated measurements. The panel (a) for β_{11} and $m = 3$, the panel (b) for β_{21} and $m = 3$, the panel (c) for β_{11} and $m = 8$ and the panel (d) for β_{21} and $m = 8$.

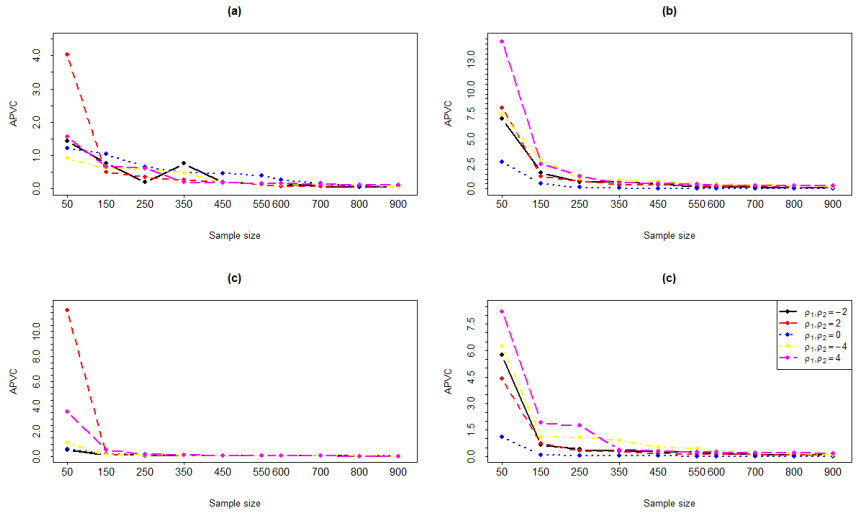

Figure 4. The APVC for different sample sizes, different association parameters and different number of repeated measurements. The panel (a) for β_{11} and $m = 3$, the panel (b) for β_{21} and $m = 3$, the panel (c) for β_{11} and $m = 8$ and the panel (d) for β_{21} and $m = 8$.

5.2 The Effect of Separate Models Instead of Joint Model

In this subsection, for investigating the effect of the use of separate models instead of joint model, we fixed some status in the previous section: $\rho_1 = \rho_2 = -2$ and $m = 8$. Separate models contain a mixed effects model for analysing longitudinal data and a parametric proportional hazard model with Weibull baseline hazard for analysing survival times. These models are fitted separately on each outcome. The separate models and joint model are used for analyzing the data and the mentioned criteria for determining sample sizes are calculated. Table 3 shows the required sample sizes for attaining the values of 0.8, 0.9 and 0.95 for BPC and for $1 - \alpha = 0.9$ and 0.95. This table shows that for attaining a pre-defined BPC a very larger sample size is required in the use of separate models instead of the joint model. Also, Figure 5 shows BPC for $1 - \alpha = 0.9$ and 0.95 for this example.

5.3 The Effect of Different Rates of Censoring

This subsection is performed for investigating the effect of different rates of censoring on the required sample size for attaining a pre-defined level of BPC. For this purpose, the example of the previous subsection with three

Table 3. Required sample sizes to make a BPC of 0.8, 0.9 and 0.95 using joint modeling with $m = 8$ and $\rho_1 = \rho_2 = -2$ for separate and joint models.

		BPC		
$1-\alpha$	model	0.8	0.9	0.95
0.9	joint model	221	305	395
	separate model	940	1006	1032
0.95	joint model	256	348	832
	separate model	1076	1157	1190

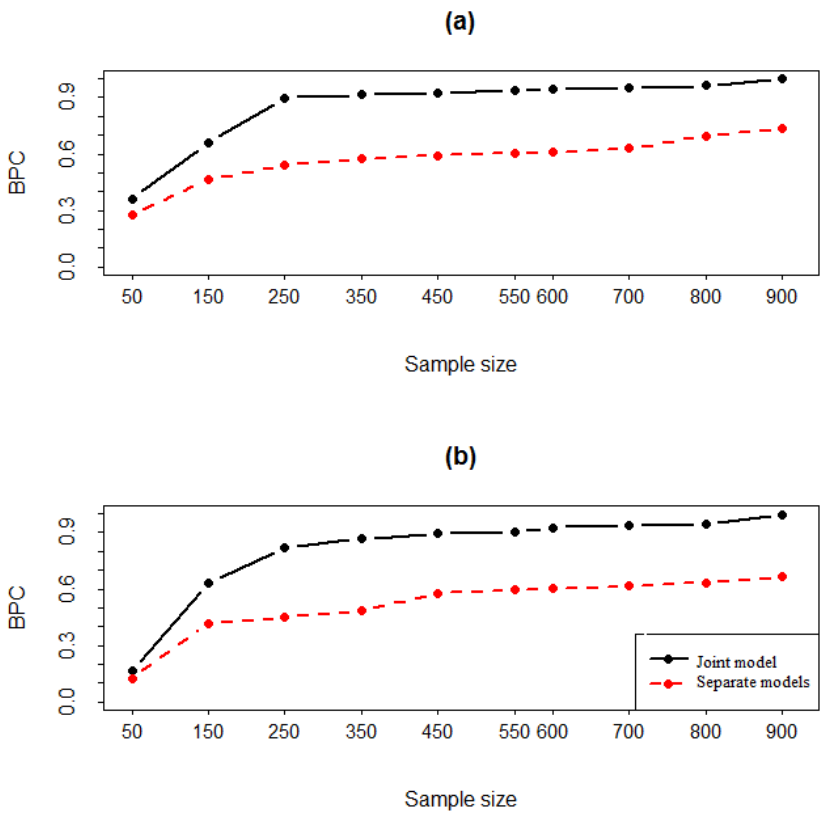

Figure 5. The BPC for different sample sizes, and different number of repeated measurements for $\rho_1 = \rho_2 = -2$ and $m = 8$. The panel (a) for $1-\alpha = 0.9$ and the panel (b) for $1-\alpha = 0.95$.

Table 4. Needed sample sizes to make a BPC of 0.8, 0.9 and 0.95 using joint modeling with $m = 8$ and $\rho_1 = \rho_2 = -2$ for different rates of censoring.

		BPC		
$1-\alpha$	censoring rate	0.8	0.9	0.95
0.9	10	221	305	395
	30	654	778	883
	50	706	853	906
0.95	10	256	348	832
	30	687	802	902
	50	759	883	930

rates of censoring $10\%, 30\%$ and 50% are considered. The required sample sizes for attaining BPC of 0.8, 0.9 and 0.95 and for $1-\alpha = 0.9$ and 0.95 are given in Table 4. The results of this table show that the higher the rate of censoring and the higher the value of $1-\alpha$ the smaller and the values of BPC. Therefore, the required sample sizes for attaining a level of BPC for higher rates of censoring is larger than those of a lower rate of censoring.

Figure 6 shows BPC for different sample sizes for $1-\alpha = 0.9$ and 0.95. This figure shows that increasing of the rate of censoring decreases the BPC.

6 Conclusions

In this paper, a simulation-based approach is proposed for determining sample size in joint model of longitudinal measurements and survival times. For this purpose, the sample size is determined based on Bayesian power criterion. Also, some other criteria such as average length, average coverage and average posterior variance are evaluated to obtain the desired sample size. Some tables containing the desirable sample size are provided for BPC values equal to 0.8, 0.9 and 0.95. Also, the sample sizes based on the interpolation of ALC and APVC are obtained and in all the sample size determinations the AC criterion is checked to be a large value.

In this paper, multiple hypotheses testing are used for sample size determination. In this framework, the effect of an event of interest on both the longitudinal model and survival model is investigated.

In comparison with other criteria which discussed in this paper, sample size determination based on BPC is more reasonable, because some pre-

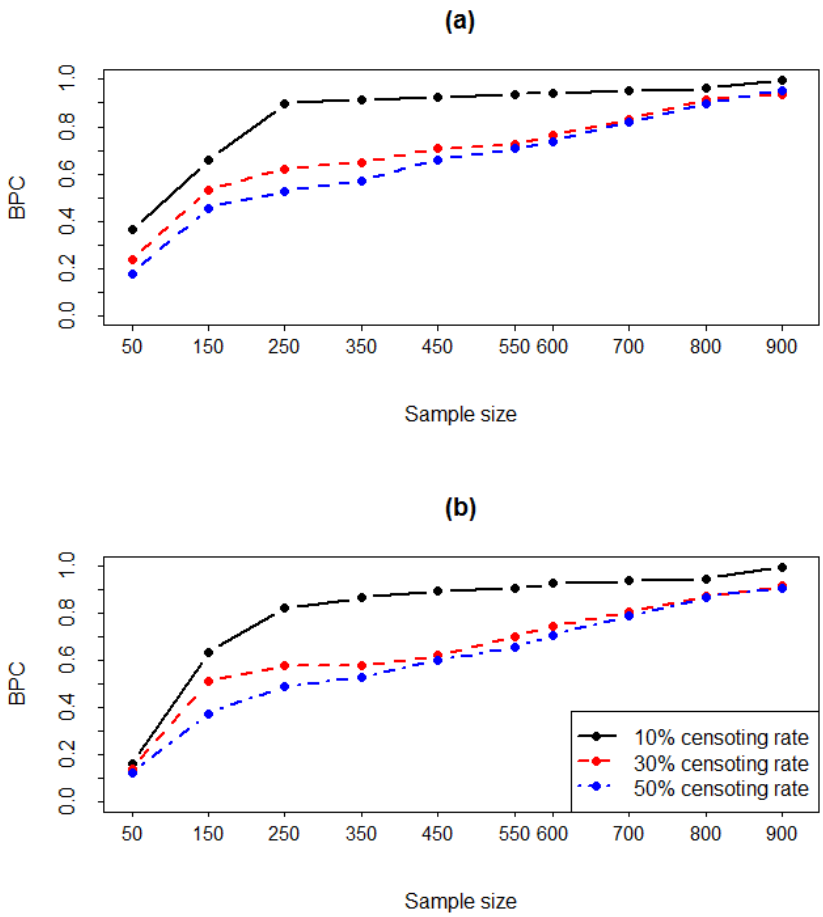

Figure 6. The BPC for different sample sizes, and different rates of censoring for $\rho_1 = \rho_2 = -2$ and $m = 8$. The panel (a) for $1 - \alpha = 0.9$ and the panel (b) for $1 - \alpha = 0.95$.

defined thresholds exist for this criterion and one can try to have a sample size with large value of BPC. In usage of ALC and ACC, we use the simultaneous confidence intervals with Bonferroni correction, as we consider, the Bonferroni corrected simultaneous confidence intervals is at level α and the family wise error rate is controlled at level 0.1.

In this paper, random right censored survival times are considered, but one can investigate the effect of other type of censoring in future studies.

As future work, one can use another multiple comparison correction or define the algorithm for defining Bayesian credible region for multiple testing such as those defined by Breth (1978), Turkkan and Pham-Gia (1997), Shalloway (2014) among others. The APVC is a criteria for controlling expected value of the posterior variance and one should control the posterior variances of both parameters by controlling the posterior variance of each parameter. As another note, one can conclude that the BPC is more a "making decision" type of criteria, while the others are "how accurate the conclusion would be" type of criteria. Because of this the final conclusion by using the BPC is based on BPC probabilities which has a "type I error" type of interpretation but the other criteria do not seem to do the same role. In this paper, the effect of the censoring rate is also investigated. As expected the smaller rate of censoring leads us to more accurate parameter estimation. Based on the results, one can conclude that this accuracy leads to smaller sample size for a pre-defined level of BPC. Also, as many authors pointed in the literature of joint modeling of longitudinal data and survival times, the use of the separate models instead of joint model leads to biased parameter estimates. This inaccuracy is reflected by large sample sizes.

In this paper, we use non-informative priors for fitting priors, if there exist any prior information to be used to help the elicitation of an appropriate prior distribution we may use it in our fitting priors. Also, ACC is used as a controlling criterion, that is, a sample size is obtained such that ACC is relatively large. The other criteria such as ALC and APVC have not any threshold and the sample size determination based on them required a pre-chosen or pre-determined value. One would consider a multi-objective criterion for determining sample size based on more than one criterion.

In this paper, we discussed sample size determination for continuous longitudinal data in the joint modeling. As a future work, one can consider sample size determination based on the approach proposed in this paper for binary or ordinal longitudinal data.

Another point which should be discussed in this paper is the use of

Bayesian paradigm for sample size determination instead of classical approach. When Bayesian approach is used for statistical inference, one can apply the Bayseian sample size determination for detecting required sample size, but an analysis with a classical paradigm needs classical sample size determination. In most of the classical approach for determining sample size a term cost is considered but in our approach this term can not be considered and considering it is an open problem in longitudinal data setting. However, by any word obtaining more accurate sample size in a study reduce the cost of the study.

One of the most important points in using the Bayesian sample size determination is the use of Bayesian paradigm and its benefits. The simplicity and flexibility of MCMC approach in determining sample size is an important point which should be referred. A disadvantage of using this approach is in the choice of the prior distributions which a discussion about it can be considered as future work.

In this paper, the effects of association parameter, censoring rate and the usual separate model are investigated. In a similar way, the effect of different covariance matrix of the random effects, for considering within-subject correlation, can be investigated. Also, a linear mixed effect model and a linear frailty model are, respectively, used for longitudinal and survival modeling. As a future work the nonlinear counterparts of these models can be investigated for sample size determination.

References

Baghfalaki, T. (2019). Bayesian Sample Size Determination for Longitudinal Studies with Continuous Response based on Different Scientific Questions of Interest. *Journal of Biopharmaceutical Statistics*, **29**, 244-270.

Baghfalaki, T. and Ganjali, M. (2015). A Bayesian Approach for Joint Modeling of Skew-normal Longitudinal Measurements and Time to Event Data. *REVSTAT–Statistical Journal*, **13**, 169-191.

Baghfalaki, T., Ganjali, M. and Berridge, D. (2013). Robust Joint Modeling of Longitudinal Measurements and Time to Event Data using Normal/independent Distributions: a Bayesian Approach. *Biometrical Journal*, **55**, 844-865.

Baghfalaki, T., Ganjali, M. and Berridge, D. (2014a). Joint Modeling of Multivariate Longitudinal Mixed Measurements and Time to Event Data using a Bayesian Approach. *Journal of Applied Statistics*, **41**, 1934-1955.

Baghfalaki, T., Ganjali, M. and Hashemi, R. (2014b). Bayesian Joint Modeling of Longitudinal Measurements and Time-to-event Data using Robust Distributions. *Journal of biopharmaceutical statistics*, **24**, 834-855.

Baghfalaki, T., Ganjali, M. and Verbeke, G. (2017). A Shared Parameter Model of Longitudinal Measurements and Survival Time with Heterogeneous Random-effects Distribution. *Journal of Applied Statistics*, **44**, 2813-2836.

Breth, M. (1978). Bayesian Confidence Bands for a Distribution Function. *The Annals of Statistics*, 649-657.

Brutti, P., De Santis, F. and Gubbiotti, S. (2008). Robust Bayesian Sample Size Determination in Clinical Trials. *Statistics in Medicine*, **27**, 2290-2306.

Chen, L.M., Ibrahim, J.G. and Chu, H. (2011). Sample Size and Power Determination in Joint Modeling of Longitudinal and Survival Data. *Statistics in medicine*, **30**, 2295–2309.

Cheng, D., Branscum, A.J. and Stamey, J.D. (2010a). A Bayesian Approach to Sample Size Determination for Studies Designed to Evaluate Continuous Medical Tests. *Computational Statistics and Data Analysis*, **54**, 298-307.

Cheng, J., Edwards, L.J., Maldonado, Molina, M.M., Komro, K.A. and Muller, K.E. (2010b). Real Longitudinal Data Nnalysis for Real People: Building a Good Enough Mixed Model. *Statistics in Medicine*, **29**, 504-520.

Christensen, R., Johnson, W., Branscum, A. and Hanson, T. (2012). *Bayesian Ideas and Data Analysis*, CHAPMAN and HALL/CRC, Boca Raton.

Dendukuri, N., Belisle, P. and Joseph, L. (2010). Bayesian Sample Size for Diagnostic Test Studies in the Absence of a Gold Standard: Comparing Identifiable with Non-identifiable Models. *Statistics in Medicine*, **29**, 2688-2697.

Dendukuri, N., Rahme, E., Bélisle, P. and Joseph, L. (2004). Bayesian Sample Size Determination for Prevalence and Diagnostic Test Studies in the Absence of a Gold Standard Test. *Biometrics*, **60**, 388-397.

Gamalo, M.A., Tiwari, R.C. and LaVange, L.M. (2014). Bayesian Approach to the Design and Analysis of Non-inferiority Trials for Anti-infective Products. *Pharmaceutical Statistics*, **13**, 25-40.

Gelman, A., Rubin, D.B. (1992). Inference from Iterative Simulation using Multiple Sequences. *Statistical Science*, **7**, 457–511.

Guo, X, and Carlin, B.P. (2004). Separate and Joint Modeling of Longitudinal and Event Time Data using Standard Computer Packages. *The American Statistician*, **58**, 16-24.

Hedeker, D., Gibbons, R.D. and Waternaux, C. (1999). Sample Size Estimation for Longi-

tudinal Designs with Attrition: Comparing Time-related Contrasts between Two Groups. *Journal of Educational and Behavioral Statistics*, **24**, 70-93.

Hintze, J.L. (2000). *Power Analysis and Sample Size (PASS) for Windows User's Guide*. Kaysville, Utah, NCSS.

Johnson, J.L., Muller, K.E., Slaughter, J.C., Gurka, M.J., Gribbin, M.J. and Simpson, S.L. (2009). POWERLIB: SAS/IML Software for Computing Power in Multivariate Linear Models. *Journal of Statistical Software*, **30**.

Joseph, L. and Belisle, P. (1997). Bayesian Sample Size Determination for Normal Means and Differences between Normal Means. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **46**, 209-226.

Landau, S., and Stahl, D. (2013). Sample Size and power Calculations for Medical Studies by Simulation when Closed form Expressions are not Available. *Statistical Methods in Medical Research*, **22**, 324-345.

Pezeshk, H. (2003). Bayesian Techniques for Sample Size Determination in Clinical Trials: a Short Review. *Statistical Methods in Medical Research*, **12**, 489-504.

Raudenbush, S.W. (1997). Statistical Analysis and Optimal Design for Cluster Randomized Trials. *Psychological Methods*, **2**, 173.

Shalloway, D. (2014). The Evidentiary Credible Region. *Bayesian Analysis*, **9**, 909-922.

Spiegelhalter, D.J., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS Examples, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK.

Tsiatis, A.A. and Davidian, M. (2004). Joint Modeling of Longitudinal and Time-to-event Data: an Overview. *Statistica Sinica*, 809-834.

Turkkan, N. and Pham-Gia, T. (1997). Algorithm AS 308: Highest Posterior Density Credible Region and Minimum Area Confidence Region: the Bivariate Case. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **46**, 131-140.

Wang, F. and Gelfand, A.E. (2002). A Simulation-based Approach to Bayesian Sample Size Determination for Performance under a Given Model and for Separating Models. *Statistical Science*, 193-208.

Wienke, A. (2011). *Frailty Models in Survival Analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Williams, M.S., Ebel, E.D. and Wagner, B.A. (2007). Monte Carlo Approaches for Determining Power and Sample Size in Low-prevalence Applications. *Preventive Veterinary Medicine*, **82**, 151-158.

Taban Baghfalak
Department of Statistics,
Faculty of Mathematical Sciences,
Tarbiat Modares University,
Tehran, Iran.
email: *t.baghfalaki@modares.ac.ir*