A Bayesian Analysis of Small n Sequential Multiple Assignment Randomized Trials (snSMARTs)

Boxian Wei * †[1], Thomas M. Braun[1], Roy N. Tamura[2], and Kelley M. Kidwell[1]

[1]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, U.S.A.
[2]Health Informatics Institute, University of South Florida, Tampa, FL 33612, U.S.A.

Abstract

Designing clinical trials to study treatments for rare diseases is challenging because of the limited number of available patients. A suggested design is known as the small-n Sequential Multiple Assignment Randomized Trial (snSMART), in which patients are first randomized to one of multiple treatments (stage 1). Patients who respond to their initial treatment continue the same treatment for another stage, while those who fail to respond are re-randomized to one of the remaining treatments (stage 2). The data from both stages are used to compare the efficacy between treatments. Analysis approaches for snSMARTs are limited, and we propose a Bayesian approach that allows for borrowing of information across both stages. Through simulation, we compare the bias, root mean-square error (rMSE), width and coverage rate of 95% confidence/credible interval (CI) of estimators from of our approach to estimators produced from (a) standard approaches that only use the data from stage 1, and (b) a log-Poisson model using data from both stages whose parameters are estimated via generalized estimating equations. We demonstrate the rMSE and width of 95% CIs of our estimators are smaller than the other approaches in realistic settings, so that the collection and use of stage 2 data in snSMARTs provide improved inference for treatments of rare diseases.

Keywords: Joint Model; Rare Disease; Clinical Trial; Bias; Mean-Square Error

*Correspondence to: Boxian Wei, Department of Biostatistics, University of Michigan, School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.
†E-mail: boxian@umich.edu

1 Introduction

The Orphan Drug Act defines rare diseases as disorders affecting fewer than 200,000 individuals in the United States [1]. More than 8,000 recognized rare diseases affect almost 30 million individuals and their families in the United States [2]. Identifying optimal treatment options for patients living with rare diseases is challenging due to the low number of individuals affected. Randomized clinical trials (RCTs) are generally regarded as providing the strongest scientific evidence for the efficacy of a treatment. RCTs attempt to minimize bias and balance confounders across treatments by employing randomization [3]. However, confirmatory RCTs often require a large number of subjects, which is difficult to attain in rare disease trials. Thus, RCTs studying treatment for rare diseases commonly have reduced power compared to studies of non-rare diseases. As a result, rare disease trials are more likely to be single arm (63.0% vs. 29.6% for non-rare disease trials) and non-randomized (64.5% vs. 36.1% for non-rare disease trials) [4]. Small sample trials of rare diseases that are randomized and multi-arm are most likely crossover, n-of-1, or adaptive designs [5].

There are disadvantages of the trial designs currently used in the rare disease landscape. For example, single arm studies are employed when the objective of the trial is to obtain preliminary evidence of the treatment efficacy and to collect additional safety data. As a result, single arm trials are not generally used to confirm efficacy [6]. In a crossover study, all subjects receive all experimental treatments. By design, each subject is their own control so that confounding and between-subject variance is reduced, leading to the need of fewer subjects than a parallel design. However, since all participants receive all candidate treatments by crossover, the treatments may expose participants to additional toxicities than a standard RCT or switch participants from an efficacious treatment to a non-efficacious treatment. These challenges have inspired the design of alternative crossover trials so that recruitment might be enhanced by offering patients the ability to switch treatment if they received no benefit from the first [7].

An n-of-1 trial is conducted in a single participant with multiple crossover treatment assign-

ments. In an n-of-1 trial, a participant is their own control so that confounding is reduced and the data can suggest which treatment produces a satisfactory outcome for the participant. However, an n-of-1 trial usually requires multiple crossover treatment assignments to defend against the effect of treatment across time, measurement error, and error from the participant's condition differing across time. The multiple crossover treatment assignments potentially prolong the duration of the trial which may be burdensome for the participant and requires a well developed trial protocol to keep the participant engaged.

Adaptive designs allow for design parameters, such as the sample size, randomization fraction, population recruited, or doses, to be altered during the trial after interim data evaluation [6]. The adaptiveness may reduce the number of subjects recruited to inferior treatment, increase efficiency, improve recruitment and take advantage of accumulating data to enable early stopping of the trial. Alternatively, an adaptive trial is often more complex to design and analyze than other standard clinical trial designs and is susceptible to bias due to temporal drift in participant characteristics.

Although the designs described above may be useful to study treatments in rare diseases, many have called for more innovative trial designs [5]. An example of such a design was proposed by Honkanen, et al [8]. This design consists of an initial randomized placebo-controlled stage, a randomized withdrawal stage for subjects who responded, and a third randomized stage for placebo non-responders who subsequently respond to treatment. As an alternative way for re-randomizing subjects, we propose design and methodological improvements for a small n sequential multiple assignment randomized trial, the snSMART [9].

An snSMART is an application of a SMART design [10, 11, 12] in small samples (e.g., rare diseases). In SMARTs, and hence snSMARTs, patients may be sequentially randomized to treatments where second-stage treatment depends on response to first-stage treatment. For example, in the ARAMIS trial, a two-stage snSMART design to compare three active treatments (Figure 1), patients are randomized equally to one of the treatment arms and followed for six months. The responders continue the same treatment for another six months, while the non-responders are

re-randomized to one of the remaining treatments for an additional six months. The outcome of interest is a binary indicator of response to treatment as defined by a combination of participant and physician measures.

Although SMARTs and snSMARTs have similar designs, we have differentiated their names because each has a distinctly different inferential goal. In a SMART, often the goal is to identify effective treatment sequences or dynamic treatment regimens that define a personalized treatment guideline for each patient that consists of a first stage treatment followed by a second stage treatment [13, 14]. In contrast, the goal of an snSMART here is in finding one superior treatment among several that would be used by itself to treat patients. Thus, in this manuscript, we are not interested in identifying dynamic treatment regimens within an snSMART.

We also highlight that some existing designs allow for absorbing endpoints, at which point all patients are assigned to a different treatment [7, 15]. In contrast, SMARTs and snSMARTs typically do not deal with absorbing endpoints because a treatment is assigned at each stage sequentially depending on the observed outcome of response at the end of current treatment stage in SMARTs and snSMARTs.

Compared to a traditional multi-stage design, such as a crossover design, the snSMART is attractive because it allows participants who have a satisfactory outcome from their treatment to continue to receive that treatment and who do not have a satisfactory outcome from the treatment to switch to another treatment. Hence, an snSMART design may help to improve participant recuitment and retention. However, analytic methods for an snSMART are not fully established, so that the efficiency gains of an snSMART design compared to other designs in rare disease research have not yet been confirmed.

Our methods are motivated by the snSMART ARAMIS (A RAndomized Multi-center study for Isolated Skin vasculitis trial), the design of which mimics the SMART design in metastatic renal cancer [16, 17]. ARAMIS (NCT02939573) is a multi-national trial to evaluate different treatment options for patients with skin vasculitis. Vasculitides are uncommon diseases which can

affect almost any organ, although vasculitis frequently involves the skin as an isolated process or as part of systemic vasculitis. Without high quality studies to guide the management of skin vasculitis, treatment decisions are made based on anecdotal experience and expert opinions. This uncertainty is reflected in variation between providers, leading to patients being treated with agents of uncertain efficacy and unknown relative merit. ARAMIS compares the efficacy of three of the most commonly used treatments for the treatment of skin vasculitis: colchicine, dapsone and azathioprine (Figure 1). Eligible patients are randomized with equal chance of receiving one of the three treatments under investigation for six months. Those who do not respond after the first stage (i.e., six months) are re-randomized equally between the other two treatments. Responders in stage 1 remain on their treatment in stage 2. The outcome of interest is response to treatment at six months as defined by a combination of participant and physician measures.

In Section 2, we present a method to analyze data from an snSMART by sharing information across stages to evaluate the efficacy of each single treatment. The efficacy of a treatment is defined as the response rate at 6 months after initiating that treatment. In Section 3, we present simulation studies to illustrate our model's properties under various scenarios. Our manuscript concludes with a discussion in Section 4.

2 Method

The outcome of interest after each stage is a dicotomized variable, where 1 denotes response and 0 denotes non-response to the assigned treatment. We propose a Bayesian approach that borrows information across both stages to estimate the individual response rate of each treatment. The individual response rate we are interested in is a permanent feature of treatment similar to the treatment effect in a large parallel group trial. We model the first stage outcome as the probability of having a response to the first stage treatment. The second stage outcome is modeled conditionally on the first stage outcome linking the first and second stage response probabilities through linkage parameters.

We compare the estimator of the response rate using the proposed method to estimators produced from three other methods: (a) a log-Poisson model using data from both stages whose parameters are estimated via generalized estimating equations (GEE), (b) a Bayesian method using only the first stage data, and (c) a maximum likelihood method (MLE) using only the first stage data. The details of our proposed model and the log-Poisson model will be discussed next and simulation results for the comparison of estimators produced from the four methods will be shown in the Section 3.

2.1 Bayesian Joint Stage Modeling

For each subject $i = 1, \ldots, N$, stage $j = 1, 2$, and treatment $k = A, B, \ldots$ K, where N denotes the total sample size and K denotes the number of arms, let Y_{ijk} denote the observed response outcome. We model the first stage outcome and the second stage outcome given the first stage outcome each as a Bernoulli random variable. The first stage response rate is denoted as π_k for treatment k. The second stage response rate for first stage responders is equal to $\beta_1 \pi_k$. For non-responders to treatment k in the first stage who receive treatment k' in the second stage, the second stage response rate in the second stage is equal to $\beta_0 \pi_{k'}$. In practice, we assume: (i) The linkage parameters (β_0, β_1) do not depend on the initial treatment k. (ii) The linkage parameter for non-responders is smaller than 1, i.e., $\beta_0 < 1$. (iii) The linkage parameter for responders is greater than 1, i.e, $\beta_1 > 1$. Via simulation, we examine the violations of the assumptions in section 3.

Our proposed Bayesian joint stage model (BJSM) is as follows:

$$Y_{i1k} | \pi_k \sim Bernoulli(\pi_k)$$

$$Y_{i2k'} | Y_{i1k}, \pi_k \sim Bernoulli((\beta_1 \pi_k)^{Y_{i1k}} (\beta_0 \pi_{k'})^{1 - Y_{i1k}})$$

Prior distributions on the first stage response rates and the linkage parameters are used to incorporate physician beliefs about the treatments. For the ARAMIS trial, we specify priors for the parameters involved in the model: $\pi_k \sim Beta(\zeta_k, \eta_k)$, $\beta_0 \sim Beta(\zeta_0, \eta_0)$, $\beta_1 \sim Pareto(1, \phi)$. For π_k, we have chosen to use the hyperparameter values $\zeta_k = 0.4$ and $\eta_k = 1.6$ for two reasons.

First, these parameters lead to a prior mean of $\zeta_k/(\zeta_k+\eta_k)=0.2$ for each of the arms, which was a reasonable *a priori* setting for the ARAMIS study. Second, the sum of the two parameters of a Beta distribution can be viewed as a prior sample size because the prior variance is inversely proportional to that sum. Thus, we assume our prior information is based upon a sample size of $\zeta_k+\eta_k=2$ patients. For β_0, we have chosen hyperparameter values $\zeta_0=1$ and $\eta_0=1$, which lead to a uniform distribution over the interval $[0,1]$. For β_1, we have assumed a hyperparameter value of $\phi=3$, so that, on average, the second stage response rate is $\phi/(\phi-1)=1.5$ times as large as the first stage response rate.

2.2 Log-Poisson Joint Stage Modeling

The log-Poisson joint stage model, which we refer to LPJSM, is a frequentist way of modeling data from two stages, where we use a log link to model the mean and the Poisson family to model the variance of the outcome. We model the the log of each response rate instead of the logit of each response rate mainly for interpretability. The model is as follows:

$$log(E(Y_{i1k}))=log(\mu_{i1k})=\alpha_A\mathbb{1}\{k=A\}+\alpha_B\mathbb{1}\{k=B\}+\alpha_C\mathbb{1}\{k=C\} \tag{1}$$

$$log(E(Y_{i2k'}))=log(\mu_{i2k'})=\alpha_A\mathbb{1}\{k'=A\}+\alpha_B\mathbb{1}\{k'=B\}+\alpha_C\mathbb{1}\{k'=C\}+\gamma_1 Y_{i1k}+\gamma_0(1-Y_{i1k}) \tag{2}$$

where $\mathbb{1}\{\cdot\}$ is an indicator function. The response rates, π_k, and the linkage parameters β_1 and β_0 from the BJSM are equivalent to the exponentiated values of α_k, γ_1 and γ_0, respectively.

We estimate the parameters via GEE [18]:

$$\sum_{i=1}^{N}\frac{\partial\boldsymbol{\mu}_i^T}{\partial\boldsymbol{\theta}}\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)=0, \tag{3}$$

where $\boldsymbol{Y}_i=(Y_{i1k},Y_{i2k'})^T$, $\boldsymbol{\mu}_i=(\mu_{i1k};\mu_{i2k'})^T$, $\boldsymbol{\theta}$ is the parameter vector with $\boldsymbol{\theta}=(\alpha_A,\alpha_B,\alpha_C,$ $\gamma_1,\gamma_0)^T$; $\boldsymbol{V}_i$ is the working covariance matrix of $\boldsymbol{Y}_i$ with $\boldsymbol{V}_i=\boldsymbol{A}_i^{1/2}\boldsymbol{R}(\alpha)\boldsymbol{A}_i^{1/2}$ [19], where $\boldsymbol{A}_i^{1/2}$ is a diagonal matrix with elements being the square root of $Var(Y_{ijk})$, the variance of the outcome of the i^{th} patient at the j^{th} stage under treatment k. The variance of the outcome of the i^{th} patient at the

j^{th} stage is modeled with a Poisson family variance structure, $Var(Y_{ijk}) = \mu_{ijk}$. We use the Poisson family variance structure to construct V_i in equation (3) to find the estimator of $\boldsymbol{\theta}$ as opposed to the binomial family variance structure, because others have reported that estimation sometimes fails to converge when attempting to fit log-binomial models with a small sample size [20]. In addition, we use an independence working correlation structure $\boldsymbol{R}(\alpha) = \mathbf{I}_{2\times 2}$ in the estimating equation because the independence working correlation structure is recommended when binary responses have less than binomial variation over clusters [21].

In estimating the variance of $\hat{\boldsymbol{\theta}}$, we use the robust 'sandwich' covariance estimator, $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}$, where $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\theta}} \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}_1 = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\theta}} \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\theta}}$. We use the binomial family variance structure, $Var(Y_{ijk}) = \mu_{ijk}(1-\mu_{ijk})$, to construct $\boldsymbol{V}_i$ in $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ to estimate variance of $\hat{\boldsymbol{\theta}}$, because the 'sandwich' estimator is consistent, when $\boldsymbol{V}_i$ is correctly specified and even if $\boldsymbol{R}(\alpha)$ misspecified [22].

3 Simulations

We set up the scenarios for our simulation studies in two subsections: simulations when the assumptions for BJSM are satisfied and simulations when our assumptions are violated. We compare the bias, root mean-square error (rMSE), coverage rate and width of the 95% credible/confidence interval (CI) of the estimated parameters. When the BJSM assumptions are true, we compare the estimators from our proposed method (BJSM) to estimators produced from other three methods: the log-Poisson joint stage model (LPJSM), described in Section 2.2, Bayesian first stage model (BFSM), and a first stage maximum likelihood estimates (FSMLE). For both the BFSM and FSMLE, we only use the first stage data for estimation and for the BFSM, we assume the same prior distribution for π_k as the BJSM. When assumptions are violated, we compare the BJSM to an extension with multiple linkage parameters (BJSMM), and the LPJSM. The BJSMM is the same as that of the BJSM except that we allow the linkage parameters to depend on the initial treatment, i.e., β_0 is now replaced by β_{0k} and β_1 is now replaced by β_{1k}. We use exactly the same values

for the hyperparameters as the BJSM for the prior densities of the BJSMM so that we allow for estimating β_{0k} and β_{1k} values that differ among different treatments k, but we give each the same prior distribution.

Bias is defined as the average of the differences between the true value of π_k and the estimated π_k in all simulations. The rMSE is calculated by taking the square root of the mean-square error of the estimators in all simulations. The simulated coverage rate is the frequency that the true value of the response paramter falls in the 95% CI for all simulations. The 95% CIs for the BJSM, BFSM and BJSMM are the highest posterior density (HPD) credible intervals, which is the narrowest interval that covers the 95% of the posterior distribution of π_k. The 95% CIs for the LPJSM and FSMLE are derived based on the asymptotic normality of the estimator of π_k in these two methods, and calculated by the estimator plus or minus 1.96 times of the standard error of the estimator. The parameters are estimated via the R function jags and gee in the R package rjags and gee respectively. The computer programs used to derive estimates and CIs are available upon request from the primary author.

3.1 Simulation Scenarios

We simulate 2000 realizations per scenario; each scenario is a three-arm snSMART. The true values of the response rates in each arm and the linkage parameters in each stage vary in different scenarios; details are presented in Table 1. We focus on simulation results where the total snSMART sample size is 90 (30 patients per treatment), but we provide results for total sample sizes of 45 and 180 in Section 6.

Scenarios 1, 2 and 3 represent three ideal settings. In these scenarios, the linkage parameters for non-responders and for responders are the same for all three treatments, which means the model specification is the same as the data generating process. In scenarios 4 to 12 we vary the values of the linkage parameters to investigate model properties when assumptions are violated (the assumptions are enumerated in Section 2.1). Assumption (i) is violated in scenarios 4-7 and

10-12. Assumption (ii) is violated in scenarios 8, 10 and 12 and assumption (iii) is violated in scenarios 9, 11 and 12.

3.2 Simulation Results When the BJSM Assumptions are True

For scenarios 1, 2 and 3, the bias and rMSE for estimators of the response rates are shown in Table 2. The response rate estimators of the BJSM have the smallest rMSEs among all four methods. The rMSEs of the estimators from the BJSM and LPJSM are smaller than the rMSEs provided by the BFSM and FSMLE, which only use data from the first stage. In scenario 2, the BJSM provides the estimators with smallest bias compared to other three methods. In scenarios 1 and 3, the bias of the estimators for the BJSM is still small but slightly higher than the bias for the other three methods. This may be because the prior mean for the linkage parameter for non-responders is 0.5 which is closer to 0.6, the setting in scenario 2, than 0.8, the setting in scenarios 1 and 3. These observations suggest that, in settings where the assumptions are satisfied, jointly modeling data from two stages provides improved estimators for treatment due to smaller rMSEs. In particular, the biggest gain in rMSE is given by the BJSM which also produces small to negligible bias. Table 3 presents the 95% CI width and coverage rates. Here we see the average width of the 95% CI of the BJSM is smaller than the other approaches and the coverage rate is around the target 95%.

3.3 Simulation Results When the BJSM Assumptions are Violated

Simulation results when the assumptions for the BJSM are violated are shown in Tables 4 and 5. When only assumption (i) is violated (scenarios 4-7), we see that the bias for the response rate estimators is small for all three methods. The estimators of the BJSMM has the smallest bias in scenario 4-6 and BJSM has the smallest bias in scenario 7. The estimators of the BJSM and BJSMM have smaller rMSEs than the LPJSM approach. When only assumption (ii) is violated, as in scenario 8, we see that the response rates are overestimated for all treatment arms by the BJSM and BJSMM. The response rates are overestimated to balance the effect of the underestimated β_0 on the stage 2 response rates of non-responders. The bias for the estimators is higher for BJSM

and BJSMM but still small. The estimators from the LPJSM have the smallest bias and rMSEs. When only assumption (iii) is violated as in scenario 9, the response rates are underestimated for all arms by the BJSM and BJSMM. The response rates are underestimated to balance the effect of the overestimated β_1 on the stage 2 response rates of responders. The estimators from the LPJSM have smallest bias but the rMSEs are higher compared with those from the BJSM and BJSMM.

When more than one assumption is violated (scenarios 10-12), the bias for the estimators of the response rates is lower in scenario 11, where assumption (i) holds and assumptions (ii) and (iii) are violated. This finding occurs because when assumption (ii) is violated, the BJSM and BJSMM tend to overestimate the response rates, and when assumption (iii) is violated, the BJSM and BJSMM tend to underestimate the response rates. and when both assumptions (ii) and (iii) are violated, two errors cancel each other. The BJSM and BJSMM have smaller rMSEs in all three scenarios.

In general, when the assumptions are violated, the response rate estimators of the BJSM have smaller bias and rMSEs than the LPJSM in most of the settings, and smaller rMSEs than the standard approaches that only use the data in stage 1. When multiple linkage parameters are considered in the BJSMM, we do not see a large reduction of bias or rMSEs.

In Table 5, we can see the average width of the 95% CI of the BJSM is smaller than the other approaches. When only assumption (i) is violated (scenarios 4-7), the coverage rates of 95% CIs for treatments B and C are around the target 95%. The coverage rate is readily below the target for all three approaches in scenarios 4 and 6. When only assumption (ii) (scenario 8) or assumption (iii) is violated (scenario 9), the coverage rate for the treatment C is below the target. This can be explained by the larger bias in the response rate estimator for treatment C in scenarios 8 and 9. In scenario 10, (when assumption (i) and (ii) are violated), and Scenario 12, when all the assumptions are violated, the coverage rates of 95% CIs are below the target 95% for treatment A and C because the response rate estimators have higher bias compared to the estimators of the treatment B. When assumptions (ii) and (iii) are violated at the same time, the coverage rates are greater than the

target 95% for the BJSM and BJSMM and below the target 95% for the LPJSM. Similar trends were observed for sample sizes of n=45 and 180; details are given in Table **??**-**??** in the Section 6.

4 Discussion

In this manuscript, we present a Bayesian method (BJSM) to estimate the response rates of multiple treatments from snSMART with two stages. The BJSM is a novel method that links the response rates from two stages of one clinical trial via linkage parameters. The BJSM provides accurate estimators and straight forward clinical interpretations for the parameters. We compared the proposed method to three other methods via simulation and found that the BJSM provides the most accurate estimators among all four methods in small samples.

The BJSM relies on three key assumptions as outlined in Section 2.1. These assumptions simplify our model and make our model easier to interpret. However, there might be situations where the assumptions are violated. Simulation results suggest that the BJSM is able to estimate treatment arm response rates with small sample sizes even when the linkage parameters actually vary among the treatment arms. Nonetheless, further research is certainly warranted to determine the sample size at which moving from a BJSM to a BJSMM might be warranted.

Assumptions (ii) and (iii) in Section 2.1 that constrain the values of the linkage parameters generally hold when the response rates are similar and low (i.e., less than 50%) for all treatments in the trial. For many rare diseases, these assumptions are realistic and, thus, violations generally do not pose problems. If the assumptions are violated, the BJSM provides more biased estimates than the LPJSM, but is more efficient. Sensitivity analyses using the LPJSM can always be done to compare results for the two methods.

A limitation that develops from our assumptions and corresponding priors is that the posterior distributions of β_1 and π_k can have positive probability for $\beta_1 \pi_k > 1$. In reality, we can not have a response rate greater than 1, but our models allow for this. To circumvent this potential problem we considered a power model formulation of the BJSM. In the power model formulation, the second

stage response rates are defined as $\pi_{k'}^{\beta_0}$ and $\pi_k^{\beta_1}$ for non-responders and responders, respectively. This allows β_0 and β_1 to vary on the positive real line. Ultimately, we decided against the power formulation as the linkage parameters are not clinically interpretable and our simulations for the proposed version of BJSM did not draw any samples such that $\beta_1 \pi_k > 1$, making it unlikely for this limitation to be a problem in practice in similar settings.

For ARAMIS, the binary six-month response is based on an objective measure (number of new lesions) that is further combined with two subjective measures, one from the physician and one from the patient. Thus, it seemed reasonable to collapse this complex set of outcomes into a single binary outcome of response. If ARAMIS had instead had a single objective, continuous measure of efficacy, analysis of a dichotomized endpoint would have been unwise, due to the loss of information that dichotomization creates. We also emphasize that ARAMIS assumes no carry-over effects and no period effects. If carry-over effects and/or period effects exist in other applications, our proposed model would need to be modified.

Future work includes extending the BJSM to non-binary outcomes (i.e. continuous and survival outcomes) and establishing sample size calculations based on the analysis of snSMARTs using the BJSM. We aim to develop sample size calculations and an easy-to-use corresponding applet that can target specific differences between treatment arms. The sample size calculations will lead us to consider alternative designs of snSMARTs where more than three treatments are involved or there is an imbalance in second-stage randomization. We are also working on another manuscript in which we extend our Bayesian joint stage model to estimate the treatment effects of dynamic treatment regimens.

5 Acknowledgements

This work was supported through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1507-31108).

6 Supplementary Materials

Simulation results for sample sizes of 45 and 180 (15 and 60 per arm) are in Table 6-13.

References

[1] 107th Congress. Rare Diseases Act of 2002. *Pub L 107–280.* 2002.

[2] Griggs Robert C, Batshaw Mark, Dunkle Mary, et al. Clinical research for rare disease: Opportunities, challenges, and solutions. *Mol Gen Metab.* 2009;96:20–26.

[3] Levin Kate Ann. Study design VII. Randomised controlled trials. *Evid-Based Dent.* 2007;8:22–23.

[4] Bell Stuart A, Smith Catrin Tudur. A comparison of interventional clinical trials in rare versus non-rare diseases: an analysis of ClinicalTrials.gov. *Orphanet J Rare Dis.* 2014;9:1–11.

[5] Gupta Samir, Faughnan Marie E, Tomlinson George A, Bayoumi Ahmed M. A framework for applying unfamiliar trial designs in studies of rare diseases. *J Clin Epidemiol.* 2011;64:1085–1094.

[6] Evans Scott R. Clinical trial structures. *J Exp Stroke Transl Med.* 2010;3:8–18.

[7] Makubate Boikanyo, Senn Stephen. Planning and analysis of cross-over trials in infertility. *Stat Med.* 2010;29:3203–3210.

[8] Honkanen Visa EA, Siegel Andrew F, Szalai John Paul, Berger Vance, Feldman Brian M, Siegel Jeffrey N. A three-stage clinical trial design for rare disorders *Stat Med.* 2001;20:3009–3021.

[9] Tamura Roy N, Krischer Jeffrey P, Pagnoux Christian, et al. A small n sequential multiple assignment randomized trial design for use in rare disease research. *Contemp Clin Trials.* 2016;46:48–51.

[10] Lavori Philip W, Dawson Ree. A design for testing clinical strategies: Biased adaptive within-subject randomization. *J R Stat Soc Ser A Stat Soc.* 2000;163:29–38.

[11] Murphy Susan A. An experimental design for the development of adaptive treatment strategies. *Stat Med.* 2005;24:1455–1481.

[12] Dawson Ree, Lavori Philip W. Efficient design and inference for multistage randomized trials of individualized treatment policies. *Biostatistics.* 2011;13:142–152.

[13] Robins James. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. *Math Modelling.* 1986;7:1393–1512.

[14] Murphy Susan A. Optimal dynamic treatment regimes. *J R Stat Soc Series B Stat Methodol.* 2003;65:331–355.

[15] Nason Martha, Follmann Dean. Design and analysis of crossover trials for absorbing binary endpoints. *Biometrics.* 2010;66:958–965.

[16] Thall Peter F, Wooten Leiko H, Logothetis Christopher J, Millikan Randall E, Tannir Nizar M. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Stat Med.* 2007;26:4687–4702.

[17] Thall Peter F. SMART design, conduct, and analysis in oncology. in *Adaptive treatment strategies in practice: planning trials and analyzing data for personalized medicine.* (Kosorok Michael R., Moodie Erica E. M., Thall Peter F.. , eds.)ch. 4.Philadelphia, ASA, Alexandria, VA: ASA-SIAM Series on Statistics and Applied Probability, SIAM; 2016:41–54.

[18] Zeger Scott L, Liang Kung-Yee. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986;42:121–130.

[19] Pan Wei, Connett John E. Selecting the working correlation structure in generalized estimating equations with application to the lung health study. *Stat Sinica.* 2002;12:475–490.

[20] Williamson Tyler, Eliasziw Misha, Fick Gordon Hilton. Log-binomial models: exploring failed convergence. *Emerg Themes Epidemiol.* 2013;10:1–10.

[21] Hanley James A, Negassa Abdissa, others . GEE analysis of negatively correlated binary responses: a caution. *Stat Med.* 2000;19:715–722.

[22] Halekoh Ulrich, Højsgaard Søren, Yan Jun, others . The R package geepack for generalized estimating equations. *J of Stat Softw.* 2006;15:1–11.

Table 1: Simulation scenarios. π_k is the response rate at six months for the treatment $k = A, B, C$. β_{0k} is the linkage parameter for the first stage non-responders treated with treatment k. β_{1k} is the linkage parameter for the first stage responders treated with treatment k. The linkage parameters link the second stage response rates with the first stage response rate in our proposed model. The three assumptions are: (i) The linkage parameters do not depend on the initial treatment k, i.e., $\beta_{1k} = \beta_1$ and $\beta_{0k} = \beta_0$. (ii) The linkage parameter for non-responders is smaller than 1, i.e., $\beta_0 < 1$. (iii) The linkage parameter for responders is greater than 1, i.e, $\beta_1 > 1$.

Scenarios	π_A	π_B	π_C	β_{0A}	β_{0B}	β_{0C}	β_{1A}	β_{1B}	β_{1C}	Assumptions Violated
1	0.3	0.3	0.3	0.8	0.8	0.8	1.5	1.5	1.5	none
2	0.2	0.3	0.4	0.6	0.6	0.6	1.5	1.5	1.5	none
3	0.2	0.3	0.4	0.8	0.8	0.8	1.5	1.5	1.5	none
4	0.2	0.3	0.4	0.3	0.6	0.9	1.5	1.5	1.5	(i)
5	0.2	0.3	0.4	0.6	0.6	0.6	1.2	1.5	1.8	(i)
6	0.2	0.3	0.4	0.3	0.6	0.9	1.2	1.5	1.8	(i)
7	0.2	0.3	0.4	0.9	0.6	0.3	1.2	1.5	1.8	(i)
8	0.2	0.3	0.4	1.2	1.2	1.2	1.5	1.5	1.5	(ii)
9	0.2	0.3	0.4	0.6	0.6	0.6	0.8	0.8	0.8	(iii)
10	0.2	0.3	0.4	0.3	0.6	1.2	1.2	1.5	1.8	(i), (ii)
11	0.2	0.3	0.4	1.2	1.2	1.2	0.8	0.8	0.8	(ii), (iii)
12	0.2	0.3	0.4	0.3	0.6	1.2	0.8	1.5	1.8	(i), (ii), (iii)

Table 2: Simulated bias and root mean-square error (rMSE) for the estimators of π_k where the BJSM assumptions are satisfied. Four modeling approaches: Bayesian joint stage modeling (BJSM), log-Poisson joint stage modeling (LPJSM), Bayesian first stage modeling (BFSM) and first stage maximum likelihood estimation (FSMLE) are compared. The sample size per treatment arm is 30. π_k is the response rate at six months for treatment k, $k = A, B, C$.

		BJSM		LPJSM		BFSM		FSMLE	
Scenario		Bias	rMSE	Bias	rMSE	Bias	rMSE	Bias	rMSE
1	π_A	0.008	0.062	-0.001	0.069	-0.008	0.079	-0.002	0.084
	π_B	0.008	0.062	0.002	0.069	-0.006	0.078	0.001	0.083
	π_C	0.008	0.061	-0.002	0.068	-0.008	0.078	-0.001	0.083
2	π_A	-0.001	0.056	-0.001	0.059	-0.002	0.069	-0.002	0.074
	π_B	0.001	0.063	0.000	0.070	-0.006	0.078	0.001	0.083
	π_C	0.000	0.067	0.002	0.077	-0.014	0.085	-0.002	0.089
3	π_A	0.005	0.056	-0.001	0.057	-0.002	0.069	-0.002	0.074
	π_B	0.008	0.062	0.000	0.069	-0.006	0.078	0.001	0.083
	π_C	0.011	0.064	0.002	0.076	-0.014	0.085	-0.002	0.089

Table 3: Simulated width and coverage of 95% CI for the estimators of π_k where the BJSM assumptions are satisfied. Four modeling approaches: Bayesian joint stage modeling (BJSM), log-Poisson joint stage modeling (LPJSM), Bayesian first stage modeling (BFSM) and first stage maximum likelihood estimation (FSMLE) are compared. The sample size per treatment arm is 30. π_k is the true response rate at six months for the treatment k, $k = A, B, C$. CR=Coverage Rate.

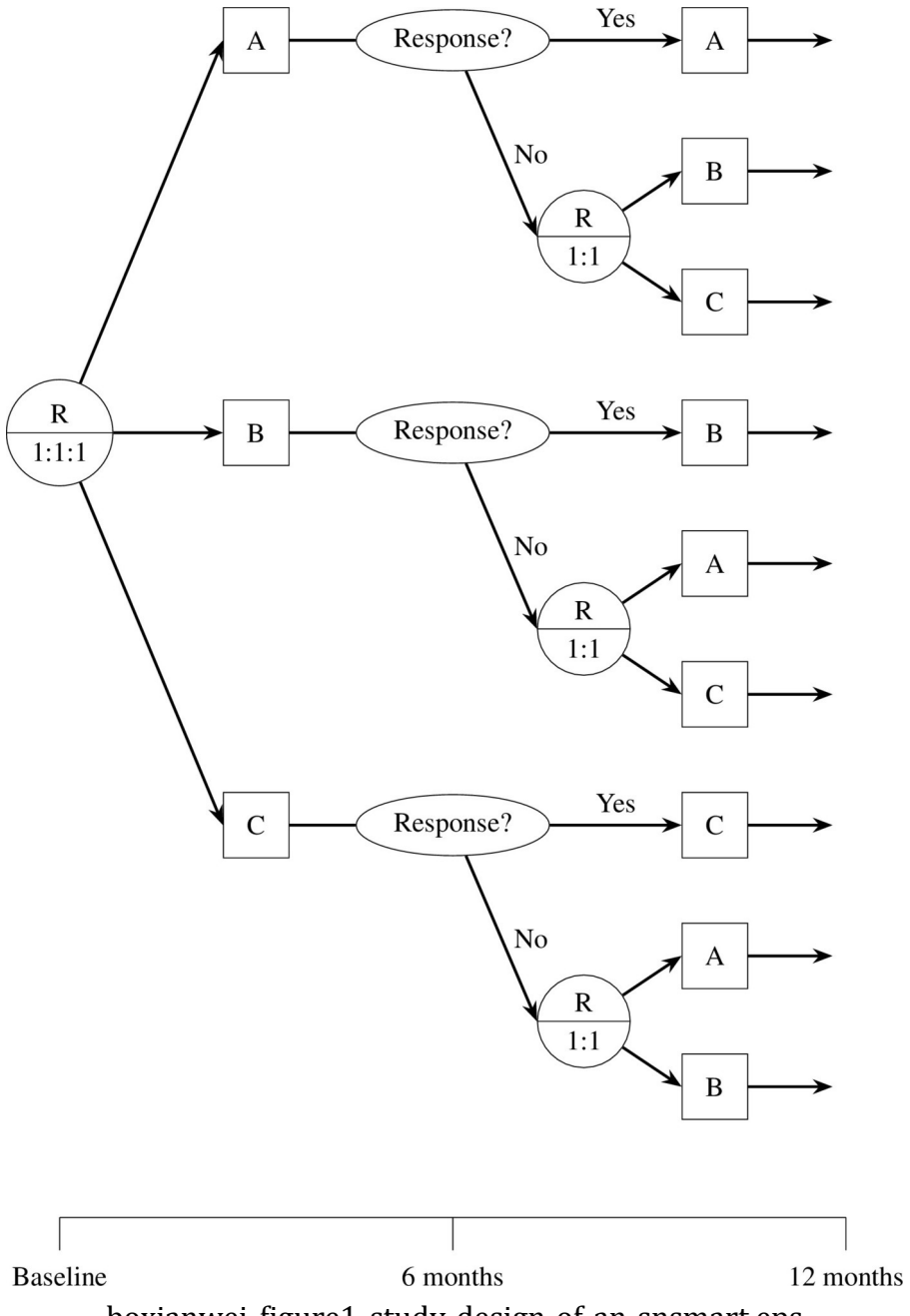
		BJSM		LPJSM		BFSM		FSMLE	
Scenario		Width	CR	Width	CR	Width	CR	Width	CR
1	π_A	0.240	0.944	0.265	0.931	0.299	0.903	0.321	0.950
	π_B	0.240	0.948	0.266	0.936	0.300	0.908	0.322	0.949
	π_C	0.240	0.944	0.265	0.934	0.299	0.908	0.321	0.950
2	π_A	0.213	0.929	0.228	0.932	0.256	0.945	0.277	0.945
	π_B	0.245	0.940	0.269	0.936	0.300	0.908	0.322	0.949
	π_C	0.265	0.948	0.305	0.937	0.323	0.930	0.344	0.930
3	π_A	0.210	0.936	0.222	0.936	0.256	0.945	0.277	0.945
	π_B	0.240	0.942	0.263	0.936	0.300	0.908	0.322	0.949
	π_C	0.258	0.956	0.300	0.937	0.323	0.930	0.344	0.930

Table 4: Simulated bias and root mean-square error (rMSE) for the estimators of π_k when assumptions are violated. Three modeling approaches: Bayesian joint stage modeling (BJSM), Bayesian joint stage modeling with multiple linkage parameters (BJSMM), and log-Poisson joint stage modeling (LPJSM) are compared. The sample size per treatment arm is 30. π_k is the response rate at six months for treatment k, $k = A, B, C$.

		BJSM		BJSMM		LPJSM	
Scenario		Bias	rMSE	Bias	rMSE	Bias	rMSE
4	π_A	-0.024	0.060	-0.021	0.059	-0.029	0.064
	π_B	-0.004	0.063	-0.001	0.062	-0.007	0.070
	π_C	0.027	0.068	0.032	0.071	0.037	0.086
5	π_A	-0.010	0.054	-0.005	0.055	-0.011	0.056
	π_B	-0.003	0.062	0.003	0.063	-0.005	0.068
	π_C	0.022	0.072	0.017	0.070	0.017	0.080
6	π_A	-0.033	0.061	-0.027	0.060	-0.038	0.066
	π_B	-0.008	0.062	0.000	0.062	-0.012	0.069
	π_C	0.048	0.080	0.048	0.081	0.050	0.093
7	π_A	0.014	0.055	0.017	0.058	0.017	0.059
	π_B	0.001	0.062	0.007	0.063	0.001	0.069
	π_C	-0.004	0.073	-0.013	0.073	-0.018	0.080
8	π_A	0.023	0.060	0.029	0.063	-0.001	0.054
	π_B	0.036	0.069	0.042	0.073	0.000	0.065
	π_C	0.047	0.076	0.054	0.081	0.001	0.073
9	π_A	-0.015	0.054	-0.016	0.054	-0.001	0.061
	π_B	-0.029	0.064	-0.030	0.063	0.000	0.073
	π_C	-0.047	0.075	-0.047	0.075	0.002	0.081
10	π_A	-0.038	0.063	-0.030	0.060	-0.045	0.069
	π_B	-0.015	0.063	-0.005	0.062	-0.024	0.070
	π_C	0.073	0.095	0.078	0.099	0.071	0.105
11	π_A	0.011	0.053	0.014	0.054	0.000	0.056
	π_B	0.010	0.055	0.013	0.056	0.000	0.067
	π_C	0.006	0.055	0.010	0.055	0.001	0.076
12	π_A	-0.047	0.066	-0.039	0.063	-0.053	0.073
	π_B	-0.014	0.063	-0.005	0.062	-0.022	0.070
	π_C	0.075	0.097	0.078	0.099	0.076	0.109

Table 5: Simulated width and coverage of 95% CI for the estimators of π_k when assumptions are violated. Three modeling approaches: Bayesian joint stage modeling (BJSM), Bayesian joint stage modeling with multiple linkage parameters (BJSMM), and log-Poisson joint stage modeling (LPJSM) are compared. The sample size per treatment arm is 30. π_k is the true response rate at six months for the treatment k, $k = A, B, C$. CR=Coverage Rate.

		BJSM		BJSMM		LPJSM	
Scenario		Width	CR	Width	CR	Width	CR
4	π_A	0.200	0.868	0.206	0.890	0.220	0.854
	π_B	0.240	0.933	0.247	0.948	0.266	0.929
	π_C	0.262	0.950	0.264	0.938	0.303	0.931
5	π_A	0.208	0.932	0.213	0.936	0.215	0.915
	π_B	0.243	0.942	0.250	0.950	0.262	0.930
	π_C	0.270	0.942	0.278	0.951	0.311	0.942
6	π_A	0.194	0.854	0.203	0.880	0.207	0.824
	π_B	0.238	0.936	0.248	0.946	0.259	0.923
	π_C	0.266	0.912	0.271	0.912	0.310	0.917
7	π_A	0.220	0.957	0.223	0.949	0.222	0.948
	π_B	0.247	0.948	0.253	0.950	0.264	0.936
	π_C	0.272	0.928	0.284	0.935	0.313	0.930
8	π_A	0.211	0.926	0.219	0.927	0.210	0.930
	π_B	0.235	0.922	0.245	0.922	0.252	0.938
	π_C	0.246	0.899	0.255	0.882	0.291	0.944
9	π_A	0.206	0.926	0.207	0.933	0.238	0.934
	π_B	0.235	0.916	0.236	0.918	0.282	0.927
	π_C	0.250	0.892	0.251	0.900	0.316	0.932
10	π_A	0.188	0.834	0.199	0.868	0.202	0.787
	π_B	0.232	0.922	0.244	0.942	0.255	0.902
	π_C	0.260	0.825	0.264	0.811	0.308	0.876
11	π_A	0.207	0.949	0.212	0.950	0.217	0.928
	π_B	0.230	0.960	0.235	0.968	0.260	0.932
	π_C	0.238	0.972	0.243	0.974	0.298	0.944
12	π_A	0.183	0.792	0.194	0.838	0.192	0.735
	π_B	0.232	0.925	0.244	0.942	0.258	0.907
	π_C	0.259	0.814	0.264	0.810	0.312	0.863



boxianwei_figure1_study-design-of-an-snsmart.eps