Supplementary Online Materials for "Bayesian Design of a Survival Trial with a Cured Fraction using Historical Data"

Matthew A. Psioda and Joseph G. Ibrahim

Web Appendix A: Further Evaluation of the Bayesian Type I Error Rate

In this appendix we provide a comparison of designs based on the Bayesian type I error rate using two historical data sets with different levels of informativeness. We let the E1690 dataset used in the paper serve as an example highly-informative historical dataset. To obtain a historical dataset with a lower degree of informativeness, we artificially reduced the information in the E1690 dataset to 30% of the actual dataset. This was done by exponentiating the likelihood. Figure 1 presents the historical trial posterior distributions for the treatment effect using each dataset as well as the corresponding default null sampling priors. Each density has been normalized to have a maximum value of one to faciliate comparison.

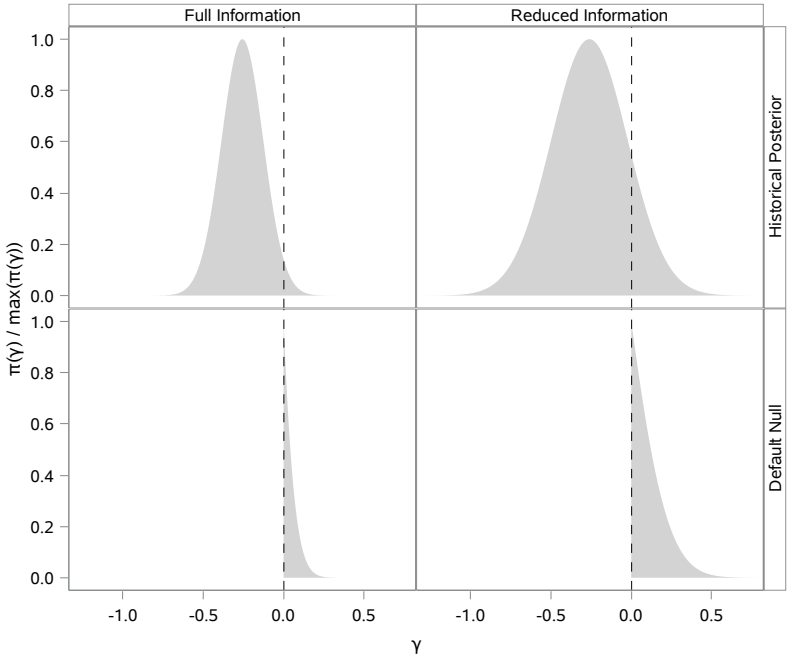

Figure 1: $\pi\left(\gamma \mid \boldsymbol{D}_0\right)$ and default null sampling priors.

Note that the default null sampling prior for the reduced information dataset supports noticeably worse treatment affects than the corresponding null sampling prior for the full information dataset. Thus, a Bayesian type I error rate defined using the reduced information null sampling prior would allow more liberal borrowing compared to a Bayesian type I error rate defined using the full information null sampling prior. However, the reduced information dataset simply has less information to borrow and so it is not clear from Figure 1 exactly how the overall sample size reduction would compare between the two cases. Table 1 presents power analyses based on the two datasets using a point-mass (PM) alternative sampling prior based on the full information dataset. Using the same alternative sampling prior ensures an apples-to-apples comparison of the sample sizes required to achieve a specified level of power. We evaluated using

Bayesian type I error rates based on the default null sampling prior (DN) as well as the partially-elicited null sampling prior (EN) as described in the paper.

Table 1: Bayesian power estimates for select sample sizes

	DN + PM				EN + PM			
	Reduced		Full		Reduced		Full	
n	Power	a_0	Power	a_0	Power	a_0	Power	a_0
600	0.80	1.00	0.81	0.33	0.78	0.70	0.77	0.20
610	0.81	1.00	0.82	0.34	0.78	0.71	0.78	0.21
620	0.81	1.00	0.82	0.34	0.79	0.72	0.79	0.21
630	0.82	1.00	0.83	0.34	0.80	0.73	0.80	0.22
640	0.82	1.00	0.84	0.34	0.80	0.73	0.80	0.22
650	0.83	1.00	0.84	0.35	0.81	0.74	0.81	0.22
660	0.83	1.00	0.85	0.35	0.81	0.75	0.82	0.23
670	0.84	1.00	0.85	0.35	0.82	0.76	0.82	0.23
680	0.84	1.00	0.86	0.36	0.83	0.77	0.83	0.23
690	0.85	1.00	0.86	0.36	0.83	0.77	0.83	0.23
700	0.85	1.00	0.87	0.37	0.84	0.78	0.84	0.24
710	0.86	1.00	0.87	0.37	0.84	0.79	0.84	0.24
720	0.86	1.00	0.88	0.37	0.85	0.79	0.85	0.24
730	0.86	1.00	0.88	0.38	0.85	0.80	0.85	0.24
740	0.87	1.00	0.88	0.38	0.86	0.81	0.86	0.24
750	0.87	1.00	0.89	0.38	0.86	0.81	0.86	0.24
760	0.88	1.00	0.89	0.39	0.87	0.82	0.86	0.25
770	0.88	1.00	0.90	0.39	0.87	0.83	0.87	0.25
780	0.88	1.00	0.90	0.39	0.87	0.83	0.87	0.25
790	0.89	1.00	0.90	0.40	0.88	0.84	0.88	0.25
800	0.89	1.00	0.91	0.40	0.88	0.84	0.88	0.25

It is quite clear from the table that the overall efficiency gain resulting from controlling the Bayesian type I error rate based on the default null sampling prior is very similar for the two historical datasets. Thus, one can conclude that the Bayesian type I error rate based on the default null sampling prior naturally penalizes highly informative prior information more than weakly informative prior information. In fact, we can see that for the reduced information dataset, all the information can be borrowed without surpassing the Bayesian type I error rate threshold when using the DN sampling prior. Lastly, the power analyses associated with using the EN sampling priors to define the Bayesian type I error rate are essentially identical. This should not be surprising as this null sampling prior was specified independently of the prior information for the treatment effect.

Web Appendix B: Robustness to Null Sampling Prior Misspecification

In this appendix we describe a simulation study that illustrates the robustness of Bayesian type I error control under null sampling prior misspecification in the scenario where both the historical and new trial have approximately balanced sample size across the treatment groups. Such balance was present in the E1690 trial and we note that most trial designs will use 1:1 randomization (when borrowing information on all subjects from a balanced historical trial) and so the assumption of balanced sample size is not artificial. When both studies have balanced sample size, the combined dataset will have an approximately equal amount of information on both the treated and control subjects. As a result, any systematic bias associated with null sampling prior misspecification for the nuisance parameters will impact both treatment groups equally leading to minimal bias in the posterior distribution of the treatment effect γ. The following simulation study illustrates this phenomenon.

We focused on a scenario where the DN sampling prior was used to identify a_0 and considered estimating the Bayesian type I error rate assuming the *correct* null sampling prior was a perturbed version of the LN sampling prior discussed in Section 5 of the paper. As noted in Section 4.3 of the paper, conditioning on the null essentially shifts the posterior distribution for the intercept parameter in the cure rate regression model (β_1) in the negative direction. We considered further shifting the mode value for the null sampling prior for β_1 in the negative direction an additional 1-3 standard deviations (SD). We then evaluated the Bayesian Type I error rate over a range of sample sizes for the new trial (each sample size paired with the a_0 value noted in Table 3 of the paper). All other simulation settings matched those described in Section 5 of the paper (e.g., distribution for enrollment times). Note from Figure 2 that the Bayesian type I error rate is only modestly effected by the this misspecification. Greater misspecification of the null sampling prior appears to result in a minimal decrease in the Bayesian type I error rate.

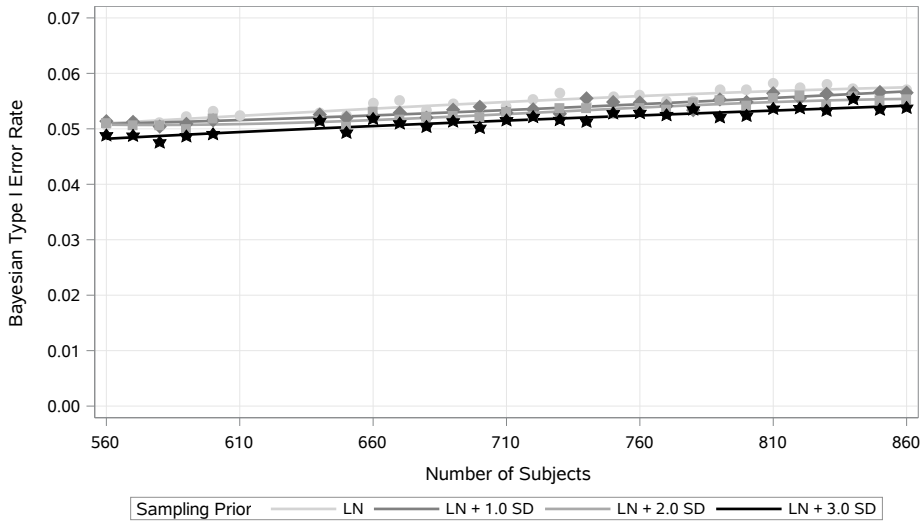

Figure 2: Bayesian type I error rate under null sampling prior misspecification.

Next, we restricted our focus to the smallest and largest sample sizes considered for the design presented in Section 5 of the paper ($n = 560$ and $n = 860$) and evaluated the type I error rate over a large discrete set of values for $\boldsymbol{\xi}$ corresponding to different perturbations to the LN sampling prior means. Our goal here is to illustrate that the type I error rate associated with any fixed value of $\boldsymbol{\xi}$ is largely robust to any systematic bias resulting from borrowing information on nuisance parameters when the amount of

information borrowed from the historical data is balanced across treatment groups. We considered both a 1.0 SD decrease and a 1.0 SD increase to the cure rate regression model parameters and the Weibull rate parameters and evaluated the type I error rate at every possible value of $\boldsymbol{\xi}$ defined through combination of those perturbations. Figure 3 presents a histogram of the estimated type I error rates at each perturbed value of $\boldsymbol{\xi}$. Note that, regardless of the perturbation, the estimated type I error rate is nearly identical to the worst-case Bayesian type I error rate described in the paper (which was based on the assumption that the LN sampling prior was correct).

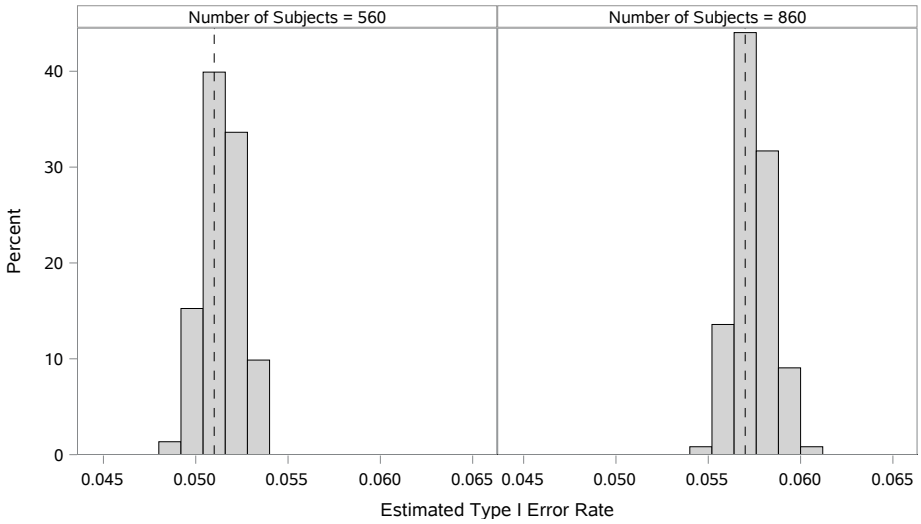

Figure 3: Estimated type I error rate based on perturbed values of $\boldsymbol{\xi}$. The dashed line indicates the Bayesian type I error rate associated with the LN sampling prior from Section 5 of the paper.

Web Appendix C: Comparison of Posterior Probability Estimation: MCMC versus the Laplace Approximation

For the simulation studies in this appendix we used the E1690 data set to serve as the historical data. To demonstrate that the asymptotic approximation developed in the paper is accurate enough for many design problems, we simulated a large number of datasets and fit the cure rate regression model to each of them using the power prior. For each dataset, we computed the posterior probability of the alternative hypothesis using MCMC as well as using the Laplace approximation. We considered sample sizes ranging from $n = 200$ to $n = 500$ and borrowing parameters ranging from $a_0 = 0$ to $a_0 = 1$. Datasets were simulated from the default null and default alternative prior predictive distributions for the data. For posterior probabilities estimated using MCMC, we used 10,000 samples. Scatter plots of the posterior probabilities are presented in Figure 4.

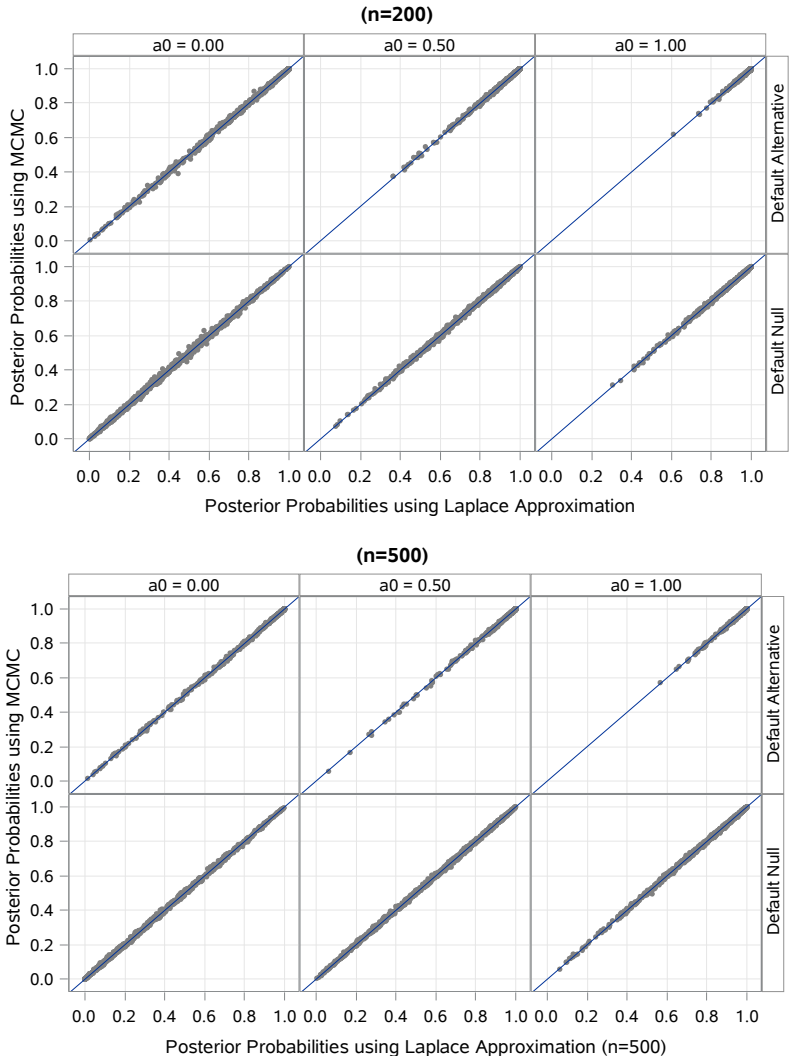

Figure 4: Comparison of posterior probabilities using MCMC versus Laplace approximation.

It is clear from Figure 4 that the two methods of estimation provide very similar results regarding posterior probabilities for the treatment effect. As noted in the paper, this is all that is needed from the approximation since accurately characterizing the posterior distribution for the nuisance parameters during design is not of interest. Of note, the minimum R^2 value associated with regressing the MCMC-estimated posterior probabilities onto those estimated using the Laplace approximation was > 0.997 for every scenario we considered. Considering the fact that the MCMC-estimated probabilities are subject to Monte Carlo error, these results strongly support use of the Laplace approximation; even for sample sizes far less than what was considered in the paper.