Supplementary Content for Web-Based Statistical Tools for the Analysis and Design of Clinical Trials that Incorporate Historical Controls

The supplementary content provides detailed instructions for using interface "Dynamic borrowing for right-censored time-to-event data using piecewise constant hazards," http://research.mdacc.tmc.edu/SmeeactWeb/ASurv.aspx.

The software's hierarchical Bayesian statistical approach uses either a simple exponential (see Section 4.1 of reference [1] below) or piecewise exponential (reference [2]) likelihood for time-to event data, and is formulated to borrow strength dynamically from the historical control data. The procedure facilitates more partial pooling of the control information in the absence of evidence for heterogeneity. The hierarchical model employs a sparsity-inducing "spike-and-slab" hyperprior distribution whose two components can be specified to vary between a uniform distribution to a two-point mixture distribution between roughly no borrowing and near exchangeability.

The web interface enables the user to analyze data uploaded to the compute server. To determine the appropriate number of pieces in the piecewise exponential likelihood [2], the software uses the AIC-optimal time-axis partition to conduct joint posterior inference on the historical and concurrent data via implementation of an AIC search (using the current data alone) among equidistant quantiles of width determined by the parameter Min.Event. Additionally, an AIC score is computed for the simple exponential model used in [1] having constant baseline hazard over the entire follow-up period. If the AIC-optimal partition in the first step contains at least two intervals, the piecewise exponential model [2] is used; otherwise, the simple exponential model is used.

Program output includes information about the treatment effect parameter ξ, which is essentially a log-acceleration factor used in a proportional hazards adjustment of the baseline hazard function to capture the effect of the novel treatment. Posterior summaries are provided for the treatment effect, corresponding hazard rate ratio (HR), and median survival for treatment and control. In addition, we provide the effective historical sample size (EHSS, the effective number of controls being borrowed from the historical data) and the novel treatment randomization probability one should use for the next patient given the results of the current interim analysis, a function of EHSS and the number of future patients remaining to enroll into the study, R.

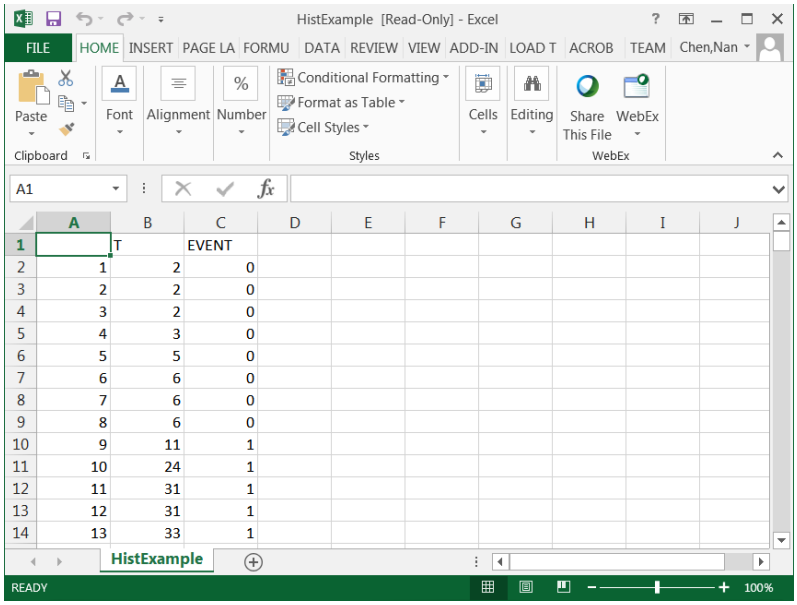
Analysis of the program-supplied example data uses the "Default data" option, while analysis of user-supplied external data instead uses the "Upload data" option. Finally, the "Quick demo" option outputs results for the default data obtained from analysis using the default design and hyperparameters, with no user input permitted.

With the "Upload data" option, the user uploads to the compute server. In order to use the software for interim analysis of an adaptive design, the user must specify the planned number of future patients that remain to enroll into the trial (R). Default values are specified for the other 3 inputs with the option for the user to modify.

More details about the data, inputs, functionality, and output can be found below.

Data preparation

Users can create their own data files for the historical and concurrent data. To check the data structure of the input files, users can download the example data files from the website and the example data files contains two files for the current (currExample.csv) and historical (histExample.csv) data. The data structure of two files are:



	A	B	C
1		T	EVENT
2	1	2	0
3	2	2	0
4	3	2	0
5	4	3	0
6	5	5	0
7	6	6	0
8	7	6	0
9	8	6	0
10	9	11	1
11	10	24	1
12	11	31	1
13	12	31	1
14	13	33	1

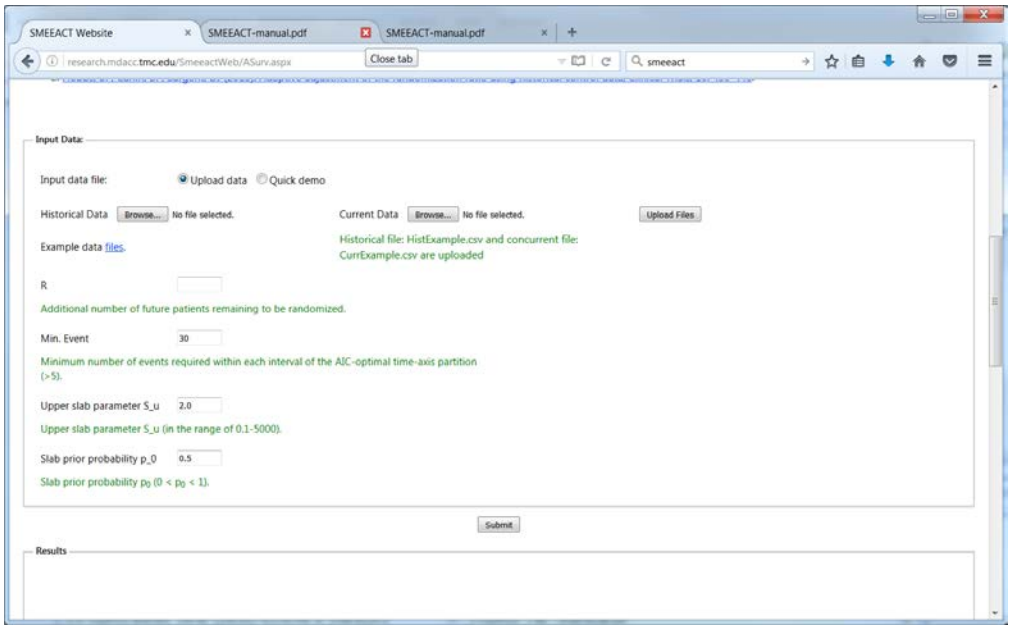
Inputs are described in detail as follows:

- "Historical Data" must contain
  - a single column labeled "T" containing time-to-event duration for each patient (assumed unique by row)
  - a single column labeled "EVENT" corresponding indicator that the event was observed (1=event, 0=right-censored)
  - example data file "`histExample.csv`"

- "Current Data" must contain
  - a single column labeled "T" containing time-to-event duration for each patient (assumed unique by row)
  - a single column labeled "EVENT" containing indicators that the event was observed (1=event, 0=right-censored)
  - a single column labeled "TREATMENT" containing indicators that the patient received the novel treatment (1=treatment, 0=control)
  - example data file "`currExample.csv`"

Application usage

With the "Upload data" option, the user uploads to the compute server using "Browse…" to select the proper local directory path. The data must consist of separate .csv files containing the historical and concurrent datasets. After choose the file paths for the historical and current data files, click "upload" button and upload files to the server. After successfully uploading the files to the server, a green text message will appears under the path selection box, as shown in the following picture.



Additionally, the user must specify the planned number of future patients that remain to enroll into the trial (R). Default values are specified for the other inputs with the option for the user to modify.

- R = Additional number of future patients remaining to be randomized
    - Used to calculate the adaptive randomization probability for assigning the next patient to enroll to novel treatment. (acceptable values: R>0) .
    - Set by default to 100 in the example analysis.

- Min.Event denotes the minimum number of events required within each interval of the AIC-optimal time-axis partition used in the piecewise exponential model likelihood. It is set by default to 30 with the option for the user to modify (acceptable values: $5 \leq$ Min.Event $\leq$ N).

- S_u is a hyperparameter in Bayesian hierarchical model characterizing the upper bound of the slab specified in the hyperprior for τ. It is set by default to 2 (the value specified in reference [2]) with the option for the user to modify (acceptable values: $0.1 <$ S_u $\leq$ 5000). Setting S_u equal to the location of the "spike" (which is fixed at 5000, again per reference [2]) yields a simple Uniform(0.1, 5000) hyperprior distribution for τ.

- p_0 is a hyperparameter in the Bayesian hierarchical model characterizing one minus the prior probability mass assigned to the spike (again, with location fixed at 5000 per [2]). It is set by default to 0.5 (the value specified in [2]) with the option for the user to modify (acceptable values: $0 \leq p_0 \leq 1$).

After providing values for the inputs, click the "Submit" button to submit the job to the server for job running. The entire computation will take several minutes to complete, depending on the size of datasets.

Functionality: The web application performs the following:

1) Computes Kaplan-Meier curves for both the historical controls and current time to event data, generates corresponding figures.

2) Performs search for AIC-optimal time-axis partition among observed equidistant quantiles of width determined by parameter Min.Event using the concurrent data alone. Also considers AIC in the absence of a partition of the time-axis, corresponding to a simple exponential model having constant baseline hazard over the entire follow-up period.

3) Fits the joint Bayesian hierarchical model with AIC-optimal time-axis partition. If the partition contains at least two intervals then the piecewise exponential model described in [2] used. Otherwise, the exponential model described in Section 4.1 of [1] is used.

4) Fits the corresponding reference models to the concurrent data alone.

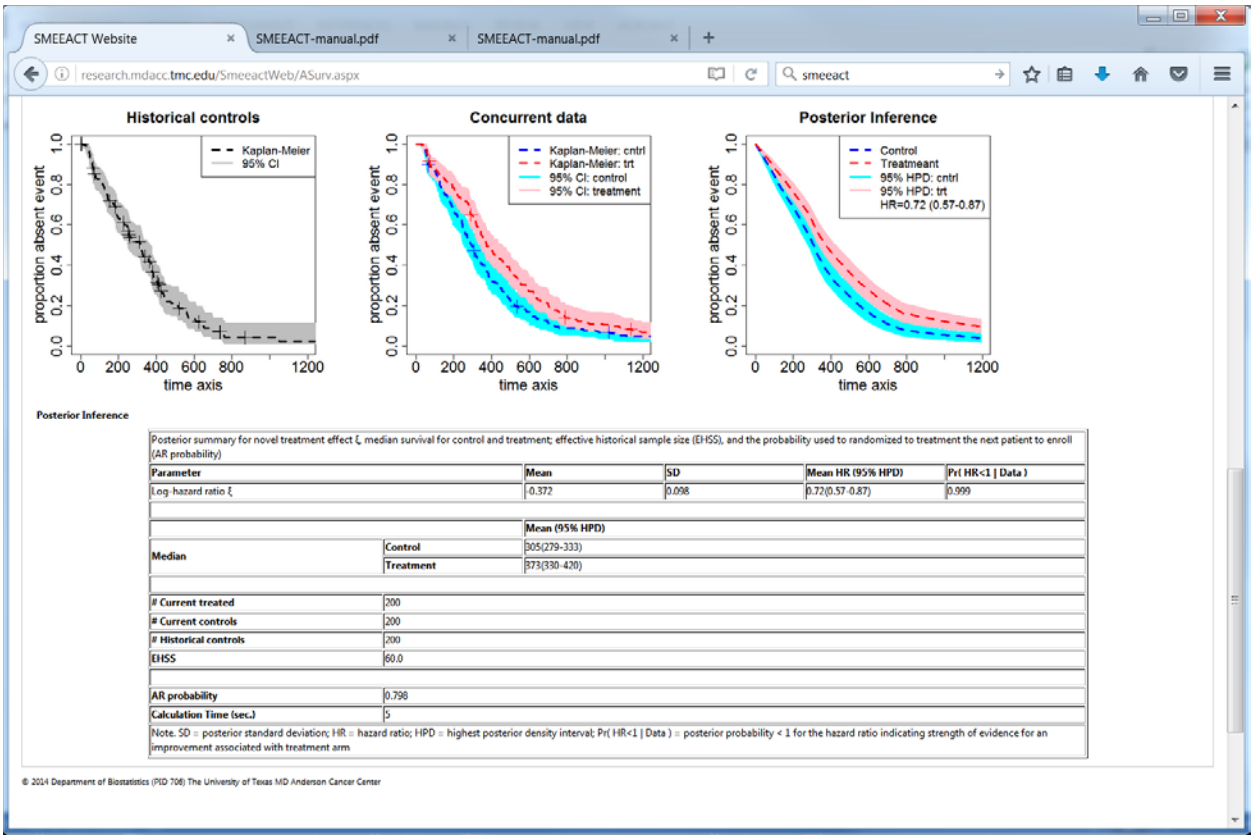
5) Computes the effective historical sample size (EHSS) (equation (9) in reference [2]), and the treatment randomization probability appropriate for the next patient to enroll after an interim analysis (equation (11) in reference [2]), a function of the allocation of patients to treatment and control arms, EHSS, and the number of additional future enrollments R.

6) Computes posterior summaries for the novel treatment effect (ξ), corresponding hazard rate ratio (HR), and median survival for treatment and control from the MCMC samples.

7) Generates a figure depicting posterior survival curves corresponding to control and treatment cohorts.

Application output

After the computation is finished, the calculation results will appear on the web browser. These include the Kaplan-Meier curves for both the historical controls and concurrent time to event data, the posterior inference, and the summary table (see the following figure).



Note that in this example, the treatment effect is significant, with 95% highest posterior density (HPD) credible interval for the hazard rate ratio (HR) excluding the null value (1.0). Also, the evidence for commensurability is very strong, leading to a high EHSS and correspondingly large adaptive randomization (AR) probabilities (shown assuming that R=100 future patients remain to be enroll into the trial), the result of the procedure having decided to use the historical controls and thus leading to a desire for more future patients to be assigned to novel treatment in order to achieve balance at the study's completion. Additional information on the adaptive randomization probability function can be found in the Appendix.

References:

1. Hobbs, B.P., Sargent, D.J., and Carlin, B.P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, 7: 639–674.

2. Hobbs, B.P., Carlin, B.P., and Sargent, D.J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials*, 10: 430-440.

3. Hobbs, B.P., Carlin, B.P., Mandrekar, S., and Sargent, D.J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67: 1047--1056.

Appendix: Adaptive randomization probability

Figure 4 below illustrates the adaptive randomization probability (11 in reference [2]) as a function of R (the number of future patients remaining to enroll into the trial) for our example analysis summarized in Table 1. Let n^{cntrl} and n^{trt} denote the numbers of patients assigned to control and treatment arms in the current trial. The adaptive randomization probability for the next patient will be 1 if EHSS $\geq$ R + $n^{cntrl} - n^{trt}$. In our example an equal number of patients have been assigned to the control and treatment arms ($n^{cntrl} = n^{trt} = 200$) at the time of the example interim analysis. In this scenario, if EHSS is larger than the number of patients remaining to enroll into the trial then maximal possible balance is achieved by assigning all remaining patients to the novel treatment. Thus, the AR probability is equal to 1 for R $\leq$ EHSS. For R > EHSS, the AR probability decreases from 1 in proportion to the number of future patients that could be assigned to control while still achieving balance after the trial's final enrollment (relative to the current value of EHSS). The AR probability of 0.798 reported in Table 1 corresponds to the case where R=100 additional patients remain to enroll into the trial.

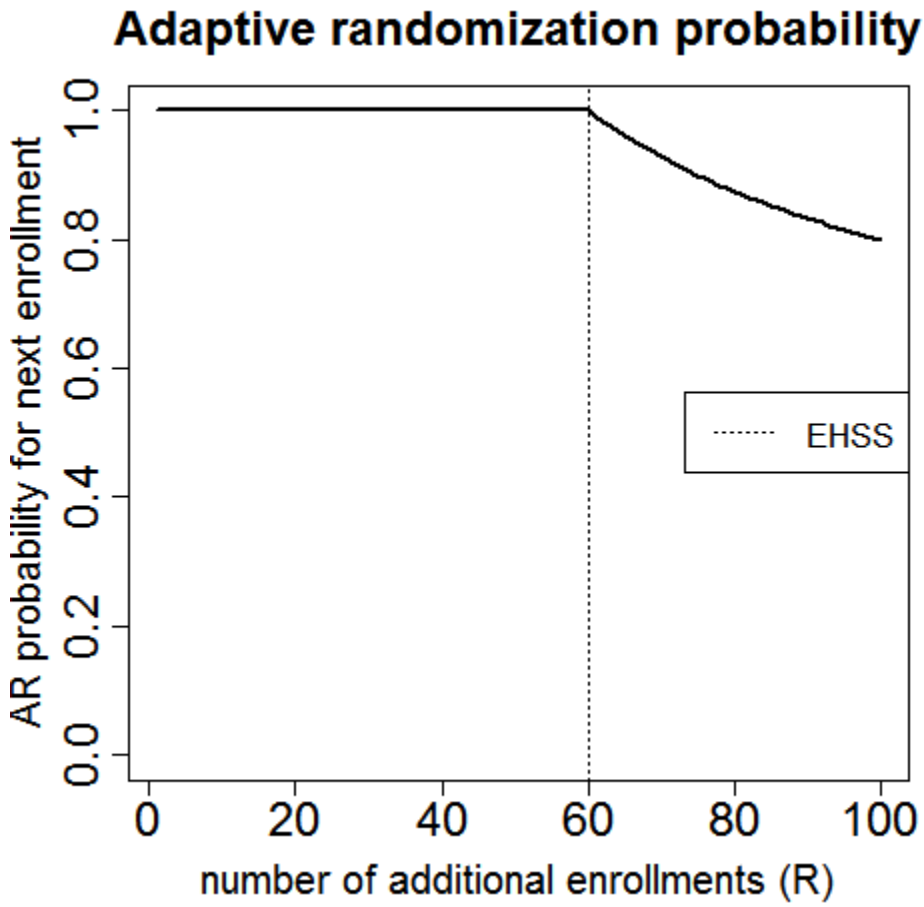


Figure 4: Adaptive randomization probability (11 in reference [2]) as a function of the number of patients remaining to enroll into the trial derived from analysis using the example data.