An implementation example of Incorporation of stochastic engineering models as prior information in Bayesian medical device trials

**Model XXXXX** Study Design
MDIC Computer Modeling and Simulation project

Medical Device Company proposes a clinical study modeled after the methods developed as a collaborative effort between FDA and industry through the MDIC consortium\(^1\). This ongoing work includes a draft manuscript and multiple mock pre-submission meetings (Q150804). The study proposed here incorporates historical safety data as an informative prior, with a loss function to protect against discordance between prior and current data.

**Sections:**

1. Study Design
2. Study Objectives
3. Incorporation of Historical Data
4. Simulation of Clinical Trial Operating Characteristics

1. Study Design

1.1. Overview

Prospective, one arm, multicenter OPC study aimed at generating 6-month safety performance results for US market approval. Because of the similarities new medical device has to predecessor market released products, we propose a modest (less than or equal to 10% weight, effectively 40 patients) Bayesian informative prior. In addition, we will add the protection of a loss function which will down-weight the strength of the prior in cases where there is disagreement between historical data and the current study.

The maximum number of enrollments for this study is 400. Interim analyses will be conducted to determine the enrollment sample size needed to meet that objectives. The first interim analysis will be conducted at XX enrolled patients, and then at YY patient increments until maximum enrollment or early stopping due to expected success or futility.

\(^1\) Medical Device Innovation Consortium (mdic.org), Clinical Trials Powered by Bench and Simulation Working Group.
All enrolled subjects will continue their follow-up. The study primary objective will be analyzed when all implanted subjects have completed their 6 months follow-up visit. Conditional approval may be requested depending on results of interim analyses.

The **new medical device** is a good candidate for an adaptive Bayesian trial with an informative prior for the following reasons:

- There are predecessor products with identical or highly similar features to every part of the new medical device. Also, these predecessors are used in the same patient population as the new device.
- Advancements in predictive engineering through activities such as the AAMI (Association for the Advancement of Medical Instrumentation) Transvenous Leads Working Group have allowed **Medical Device Company** to gain understanding about potential failure modes, giving confidence that the historical data will be exchangeable.
- Collaborative initiatives between industry and FDA are currently underway to develop a methodology for incorporating informative priors into medical device clinical trials.
- Predecessor products with similar features have had very good field performance.

### 1.2 Informative Prior

We also note that the MDIC working group has focused on the use of engineering model data, however this method is also readily applicable to historical data sets.

Safety data from the 401 patients from the **predecessor device** study will be used as an informative prior, in an adaptation of the power prior method [1]. The weight of the historical data will be adjusted using a loss function [2], which scales from 0 to 1 according to the similarity of the historical and observed data. This loss function adjusts the amount of weight the prior receives. This prevents the use of an informative prior where exchangeability issues are present. The loss function will be discussed below.

If the analyses show a high level of agreement or better performance for new medical device compared to predecessor device, the predecessor device data will be weighted at or near 100%. If the new medical device performs worse than predecessor data, the new medical device data will receive very little or zero weight. The advantage of using a loss function with the power prior method is protection against using an informative prior where there are issues with exchangeability.

Figure 1 illustrates the effect of the power prior coupled with the loss function. The figure shows credible intervals in scenarios where the prior is optimistic (better performance than current study), in agreement (similar performance to the current study), and pessimistic (worse performance than the clinical study). As will be discussed later, the data from the predecessor device study are scaled to represent a maximum of 40 patients (i.e. 10% weight).
The prior data set in Figure 1 has 401 samples. However, the maximum effective historical data sample size is $n_H = 40$, for a maximum weight of 10%. Therefore, the prior will be scaled by a factor of $(40/401)$. If the effective sample size is 2 out of 40, the prior has received 5% of the maximum weight, or 0.5%.

The panels in Figure 1 can be interpreted as follows:

- **Top panel**: The current data shows lower performance than the prior. The loss function produces a substantial penalty resulting in almost no weight to the prior (0.5%). The posterior is essentially the same as the current study.

- **Middle panel**: The current data is very similar to the prior. The loss function penalty is small, resulting in a prior weight of 7% (recall that the maximum weight is 10%). Because the agreement is good, the posterior is similar to both the prior and current study.

- **Bottom panel**: The current data is very similar, with slightly better performance than the prior. The loss function produces a weight very close to the maximum of 10%. The posterior is a balance between the prior and current study.
1.3 Adaptive Bayesian Design

A Bayesian adaptive design is set up to enroll patients until a sufficient sample size is achieved to have high probability of meeting the endpoint, or high probability of a futile study. The sample size of the study may vary from **XX** to **YY** subjects due to the adaptations to the trial. This study follows methods from Berry, et.al. [4].

The first interim analysis will take place after the 100th subject is enrolled in the study with additional analyses at 30 subject increments. At each interim analysis enrollment may be stopped for expected success or futility.

Adaptive Bayesian sample size algorithm, also illustrated in Figure 2:

1. If the predictive probability of trial success with the current sample size is larger than 90% then enrollment will stop for expected success ($P[suc] > 0.90$).
2. If the predictive probability of trial success for the maximum sample size of 400 subjects is less than 1% then enrollment will stop for futility ($P[suc] < 0.01$).
3. If neither step 1 or step 2 above holds then enrollment will continue until the next sample size analysis, or if the maximum sample size of 400 is reached.

![Figure 2: Adaptive Trial Algorithm](image)

At the time of each interim analysis, some patients will not have completed the full evaluation period. A longitudinal model will be employed to enable final observations to be imputed for those subjects with incomplete information.

There are 3 types of subjects at a given interim analysis:

1. subjects that have complete data
2. subjects that have partial data (censored value at a particular time)
3. subjects that have no information (subjects that have not been enrolled)

We will need to construct predictive probabilities for types 2 and 3. The predictive probability model that will be used is a piecewise exponential. This will allow us to simulate the final outcomes for the subjects who have not had an event and have not completed 6 month follow up.

2. Study Objective

Primary Objective: 6-month primary safety objective (predecessor device as an informative prior)

The new medical device will be considered safe if the probability of subjects free of severe adverse device-related complications at 6 months post-implant is greater than \( ZZ\% \) (i.e., the two-sided 95% CI lower confidence bound must be greater than \( ZZ\% \))

Null Hypothesis (H0): \( \theta \geq QQ \)

Alternative Hypothesis (Ha): \( \theta < QQ \), where \( \theta \) is probability of experiencing a primary safety event.

The null hypothesis is rejected when at least 97.5% of the posterior distribution is below \( QQ\% \), i.e. \( Pr(\theta < QQ \mid Data) \geq 0.975 \)

3. Incorporation of Historical Data

3.1 Maximum weight from historical data

A maximum weight of \( n_h = 40 \) effective patients will be given to the historical data. Because the actual studies enrolled more than 40 patients (401 in this case), the data will be scaled within the power prior to create an effective weight of 40 patients.

3.2 Power prior implementation for binomial distribution

Let \( y_c \) and \( y_h \) be the number of events that occur for the current data and historical data respectively and let \( n_c \) and \( n_h \) be the total number of subjects for the current and historical studies that reach the 6 month time point or had an event. Following the work of Ibrahim and Chen [1], the posterior probability for event rate \( \theta \) using the power prior is given as:

\[
P(\theta \mid y_c, y_h, n_c, n_h, a_0) = beta(\alpha, \beta)
\]

where

\[
\alpha = y_c + a_0 y_h + 1
\]

\[
\beta = (n_c - y_c) + a_0 (n_h - y_h) + 1
\]
$a_0$ is a scalar parameter between 0 and 1 governing the weight of the historical data. When $a_0$ is 1, the historical data is fully used. When $a_0$ is 0, the historical data has no effect. We use a loss function to determine $a_0$, discussed in the next section.

### 3.3 Loss function: determining $a_0$

A loss function is used to adjust the strength of the prior according to the agreement between current and historical data. This loss function approach was proposed by the MDIC working group and is also detailed in a draft publication [2].

Suppose we develop the posterior distribution of $\theta_c$ and $\theta_h$ for the current data and historical data, respectively, both with minimally informative priors. Calculate $\theta$ as:

$$\theta_c = beta(y_c + 1, n_c - y_c + 1)$$
$$\theta_h = beta(y_h + 1, n_h - y_h + 1)$$

We can stochastically compare the distribution of $\theta_c$ to $\theta_h$ using a posterior Bayesian $p$-value [3] as:

$$p = P(\theta_c \leq \theta_h)$$

The desired characteristics of the loss function utilized here are:

1. For $p \geq 0.5$, there is a high level of agreement between current and historical data, therefore the loss function should allow $a_0$ to be close to 1, allowing for full weight of historical data.

2. Conversely, for $p < 0.5$, there begins to be evidence of disagreement between current and historical data, and $a_0$ should start to down-weight the prior, i.e. $a_0$ approaches zero as $p$ approaches zero.

The Weibull cumulative distribution function meets these criteria:

$$a_0 = 1 - e^{-(p/3)^2}$$

Note that for cases where the number of samples in the prior is different than the effective number, a scaling factor will be applied, where $n_h$ is the desired effective number of prior samples and $N_h$ is the actual number of prior samples:

$$a_0 = \frac{n_h}{N_h} \left[ 1 - e^{-(p/3)^2} \right]$$

Sample values are listed in the table below and illustrated in Figure 3:
Note that the loss function proposed here does not reduce the strength of the prior when the current study outperforms the historical data. This implementation of the loss function is only concerned with negative impacts to patients, i.e. it penalizes an optimistic prior while not penalizing a pessimistic prior.

Also, note that this function is selected for the shape of the CDF rather than due to conventional statistical properties of the Weibull distribution.

### 4. Clinical trial operating characteristics

Simulations were performed with true failure rate of 0.XX, 0.YY, and 0.ZZ. 5,000 simulations were run at each value of true failure rate. This allows characterization of power and type I error, as well as number of enrolled patients and the amount of borrowing from the historical data ($a_0$), all as a function of the true failure rate.

Figure 4 describes the relationship between power, type I error, and true failure rate in the current study. When the true failure rate is less than XX, the chance of trial success (power) is at least YY%. As the true failure rate approaches the null hypothesis (0.QQ failure rate), the chance of trial success decreases to ZZ.

**From the simulations, in order to keep the type I error below 0.05, the posterior decision rule probability was set to XX (0.98??).** This is inflation is due to repeated looks at the data.
Figure 4: Power vs. true failure rate and type I error at the null hypothesis

Figure 5 shows number of enrolled patients vs. true failure rate. Error bars indicate 5th and 95th percentile from the 5,000 simulations at each true failure rate. Additionally, Figure 5 shows $a_0$, the weight of the historical data vs. true failure rate. This plot shows that when the true failure rate is very low relative to the null hypothesis, enrollment will stop early and nearly full weight is given to the historical data. The total number of enrollments peaks at a true failure rate of 0.00. The number of enrollments drops as the true failure rate approaches the null hypothesis, due to early stopping for futility.

Figure 5: Number of enrolled patients and weight of historical data vs. true failure rate
References:


[2] T. Haddad, A. Himes, L. Thompson, T. Irony, R. Nair, "Incorporation of stochastic engineering models as prior information in Bayesian medical device trials", *Draft manuscript*
