2014VP Clinical Study Proposal

Executive Summary

This proposal is part of a mock submission activity that builds on work including a draft manuscript and multiple previous mock pre-submission meetings (Q150804, Q150804/S001).

MDIC proposes a prospective, non-randomized, single-arm multi-center clinical study to evaluate the safety of the hypothetical Model 2014 ICD lead. The primary objective of this study is to evaluate the lead fracture survival of the Model 2014 lead. The study incorporates virtual patient data from an engineering model as an informative prior, using a loss function to control the amount of borrowing from the virtual patient model in the event of discordance between prior and current data.

The clinical trial design we propose will enroll between 200 and 410 patients. The maximum percentage of virtual patients will not exceed 45%. Type I error is controlled at 0.10. The study power is 85% at the expected lower bound of virtual patient model performance. The frequentist operating characteristics of the proposed design under various scenarios are provided in Table 3.

As with the previous discussions, the scope of this study is for failure due to conductor fracture inside the intracardiac anatomical zone. In an actual submission, efficacy of the device, as well as all potential failure modes and anatomical zones would be addressed.

The outcome of the previous discussions is summarized here:

- Clinical performance of the model 2014VP ICD lead has been simulated using a Virtual Patient Model (VPM). The VPM incorporates *in-vitro* test data, *in-vivo* use condition measurements, and a statistical reliability projection methodology.

- The context of use for the VPM is to generate clinical outcome data that could be used with prospective data from the proposed clinical study to evaluate the safety of the 2014VP ICD lead for intracardiac fractures. The model input data and algorithm have been shown to be credible for this intended use. Validation of the model using a market released lead with known long-term performance was a valuable step in establishing credibility.

- The simulated lower 5% confidence bound on intracardiac fracture survival for the 2014VP patient population at 5 years is 99.75%, which compares favorably to the results of the same methodology applied to the predicate product model 2005 lead, which had simulated 5 year survival of 99.68%, and has had acceptable performance in the field.

- We proposed the following for consideration of a clinical study design: 1) the maximum % of virtual patients would be in the range of 30-50%, and 2) the type I error would be in the range of 0.15-0.20, as supported by the influence, risk, and credibility of the virtual patient model. Note that these numbers are given as ranges, because the clinical study had not been designed at the time of the earlier discussions.

This document contains the following sections:

1. Study Design

2. Incorporation of Virtual Patient Data

3. Simulation of Clinical Trial Operating Characteristics

MDIC is requesting feedback from FDA on the following two questions:

1. Is the study design acceptable for the specific context of lead failure due to intracardiac conductor fractures, given the assumptions associated with this mock submission framework?

2. Is the virtual patient model discussed in Q150804/S001 suitable as an informative prior for the proposed study design?

1. Study Design

1.1. Overview

The proposed study is a prospective, single-arm, multicenter objective performance criteria (OPC) study to assess 18-month safety performance for US market approval of the Model 2014VP ICD lead. Because of the similarities the Model 2014VP lead to the predicate Model 2005 market released lead, we propose to utilize a Bayesian adaptive design with informative prior.

The maximum number of subjects who will be enrolled in this trial is 410. Interim monitoring will be conducted starting after the first 200 patients are enrolled, and then at 30 patient increments until maximum enrollment or early stopping due to expected success or futility. The amount of borrowing will be limited to 160 virtual patients. In addition, our proposed Bayesian methodology adds the protection of a loss function which will down-weight the strength of the prior in cases where there is disagreement between virtual patient data and the current study.

All enrolled subjects will continue their follow-up. The study primary objective will be analyzed when all implanted subjects have completed their 18 months follow-up visit. The marketing application may be submitted early if the predictive probability of success at interim analyses is larger than 90%.

1.2 Study Rationale

The 2014VP is a good candidate for an adaptive Bayesian trial with an informative prior for the following reasons (also summarized in Q150804/S001):

- The Model 2014VP lead has similar or identical features to every part of the predecessor Model 2005 lead. The portions that are different (insulation, conductor material) have been characterized with the virtual patient model, and the expected fracture performance is comparable to the predecessor acceptable.

- 2014VP is intended to be used in the same patient population, with the same indications, as the Model 2005.

- Advancements in predictive engineering through activities such as the AAMI (Assocation for the Advancement of Medical Instrumentation) Transvenous Leads Working Group have allowed MDIC to gain understanding about potential failure modes, giving confidence that the virtual patient data will be exchangeable

- The proposed methodology incorporates a loss function to control the amount of borrowing from the virtual patient model in the event of discordance between prior and current data.

1.3 Study Objective

Note specific to this mock submission: in this document, we restrict our analysis to lead fracture in the intracardiac anatomical zone. In an actual study, all anatomical zones and all failure modes would likely be included. Additionally, an actual study would typically contain efficacy objectives.

The goal of this trial is to show that the safety of the Model 2014VP lead meets the objective performance criterion (OPC) which have been established based on the historical failure rate for market released leads. The primary endpoint of this trial is the fraction of subjects (θ) experiencing a lead fracture in the intracardiac anatomical zone during the 18 months post-implant. Subjects who reach 18 months post-implant (documented by a follow-up Case Report Form on or after 18 months post-implant) without experiencing a lead fracture in the intracardiac anatomical zone are considered event-free.

Primary Objective: 18-month primary safety objective

> The 2014VP will be considered safe if the probability of subjects experiencing an intracardiac lead fracture at 18 months post-implant is less than 3%.
>
> Null Hypothesis (H_0): $\theta \geq 0.03$
>
> Alternative Hypothesis (H_a): $\theta < 0.03$
>
> The null hypothesis will be rejected with a type I error rate of 10%.

1.4 Informative Prior

1.4.1 Engineering model

This section is a summary of material presented in the previous mock submission document Q150804/S001. It is intended to provide context for the reader, but not specific details of the implementation of the engineering model.

The design of the 2014VP lead results in a change in handling characteristics as compared to the predicate product model 2005 lead, which has acceptable field performance since its market release in 2005. Additionally, manufacturing changes that improve fatigue performance[1] of the MP35N[2] conductor material have been implemented. Together, these changes are expected to result in equal or better fatigue performance in the patient population. The 2005 and 2014VP lead cross sections are shown in Figure 1.

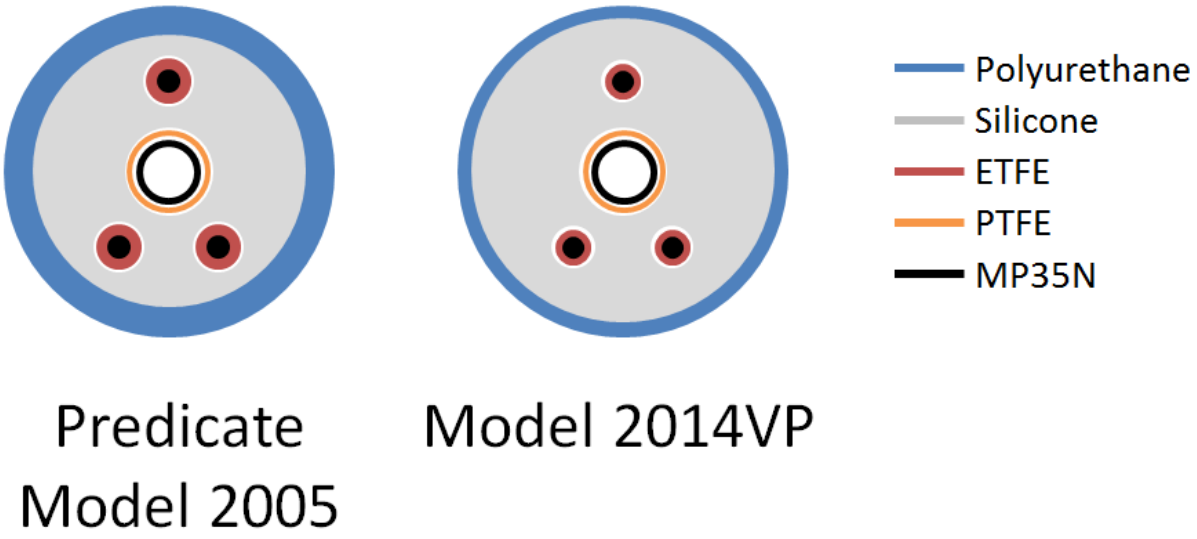

Figure 1: Cross sections of the predicate product model 2005 and 2014VP

The expected field performance for fracture survival for the 2014VP lead is estimated by a virtual patient model (VPM). The VPM is an algorithm that predicts the fracture survival for patient cohorts based on distributions of the following input data as depicted on the left in Figure 2:

- Lead curvatures measured from imaging,

[1] The manufacturing changes to the MP35N result in a reduced number and size of titanium nitride (TiN) particles. These particles commonly act as fatigue crack initiation sites [5].
[2] MP35N is a trademark of SPS Technologies, Jenkintown, PA, USA

- Activity levels for patients in different age categories, and
- Fatigue life (S-N) curves of manufactured leads

Cohorts of virtual patients can be simulated using the VPM algorithm, generating families of survival curves as the output, which are shown on the right side of Figure 2. The VPM algorithm is presented in this section, and is described in greater detail in [1]. As noted above, we are only considering lead fracture inside the heart. In an actual submission for a cardiac lead, all components and all anatomical zones would be accounted for.

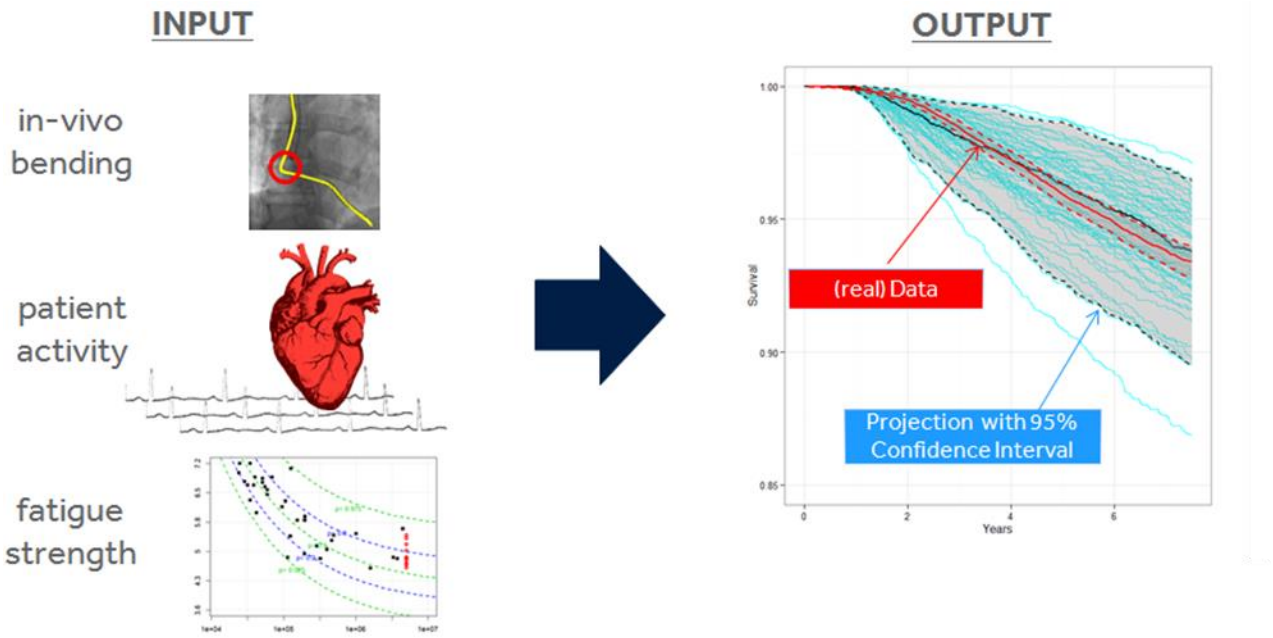

Figure 2: Illustration of the virtual patient model.

The VPM simulates lead fracture rate in many cohorts of patients, as discussed in [7,8][3]. Each individual patient cohort is the result of a Monte Carlo simulation expressing variability due to patient anatomy, physician implant factors, material properties and manufacturing tolerances, captured by the VPM inputs. However, because the variability is not known exactly (e.g., the mean and standard deviation for fatigue strength are estimates from the available data), we simulate multiple patient cohorts with different possible levels of variability, i.e., uncertainty for each input. This can be thought of as a Monte Carlo simulation of Monte Carlo simulations.

[3] The algorithms in references [7] and [8] have slight differences. The approach discussed in [8] had the intent of matching field performance with modeled results, and contained analysis elements such as healing time, and load ratio effects that improved the quantitative projection ability. The algorithm in [7] has been simplified, with the sole goal of distinguishing the field performance of leads with good field performance from those that do not have good field performance. This objective does not require the output to match absolute field performance, but rather to preserve the relative ranking between products.

A high level version of the algorithm to generate a single virtual patient is as follows:

a. Generate random patient (age), based on patient demographics from predicate device registry, to capture mortality for competing risk calculation. Note that age may also be used as a surrogate for activity level in many other cases. We are assuming a constant heart rate of 79 beats per minute for all patients, regardless of age.

b. Generate random implant curvature (i.e., bending) (see section 2.2 and 3.1) to represent how the lead was implanted during the surgery and the patient's anatomy.

c. Generate random fatigue strength using curvature from steps a & b as inputs. (see section 2.3 and 3.3)

d. Convert cycles to years assuming 79 beats per minute.

e. Based on patient age and time to fracture, determine whether the lead will fracture during the life of the patient.

This process is repeated many (e.g. 1000) times to simulate an entire cohort of virtual patients, using a Monte Carlo simulation.

In order to generate multiple patient cohorts, we generate new values for the parameters (e.g., location parameter, scale parameter, mean, standard deviation) of the statistical distributions for curvature and fatigue strength and repeat the process. The reason there is variability in these parameters is because the original data sets have a finite sample size, resulting in uncertainty. Examples are given in section 4. Note that some parameters (e.g. patient age, mortality and heart rate) may have negligible uncertainty due to the large sample size for input data. Additionally, if the virtual patient model is relatively insensitive to variation in that parameter, the resulting uncertainty in the virtual patient model will be negligible.

Each virtual patient is represented by a time to event (e.g. months) and a corresponding event indicator (e.g. failure or censor). From this time to event, we can construct a binary variable indicating failure occurring before a particular time point (e.g. 18 months). This binary variable will be called y_0, and is used for constructing the binomial prior for the primary analysis discussed below. Note that the prior of the piecewise exponential analysis described below will utilize the time to event of the virtual patient model.

The distribution of simulated implant survival for the multiple patient cohorts at particular time points will be the output of interest that is compared to clinical trial results.

1.4.2 Use condition collection

For the 2014VP virtual patient model, one of the inputs is *in-vivo* bending, see Figure 2. The initial approximation for *in-vivo* bending was determined with hypothetical stiffness of the 2014VP lead, and interpolating from a transfer function based on the AAMI Lead Imaging

Study, as discussed in Q150804/S001. The hypothetical results were used to guide the design of the clinical study discussed here.

To obtain accurate data for *in-vivo* bending on implanted 2014VP leads, fluoroscopic imaging will performed on the first 40 patients[4]. This will yield a distribution of actual *in-vivo* bending for the 2014VP that can be used to update the virtual patient model.

For example purposes in this document, we assume the imaging is completed, and the updated *in-vivo* bending distribution shows a higher curvature amplitude than the original assumption, resulting in an increased predicted failure rate. The previous and updated *in-vivo* data and failure rate are summarized in the Table 1 below.

Table 1: Original and updated use condition inputs and virtual patient failure rate

	original	**after update**
curvature amplitude (lognormal location, scale / mean, standard deviation) in^{-1}	-1, 0.44 / 0.41, 0.19	-1, 0.65 / 0.45, 0.33
failure rate (lower 95^{th} confidence bound)	0.19%	1.22%

Although the updated virtual patient failure rate is higher than the original model (1.22% vs. 0.19%), it is still within the safety objective (3%). We use the updated data for *in-vivo* bending for all further analysis, presented in section 3.

Note that the updated model could affect all operating characteristics of the study. Therefore, all operating characteristics should be analyzed after collection of use conditions. The sponsor would be expected to fully pre-specify the procedure for model updating and analysis.

The patients that were imaged continue to be enrolled in the study, as their clinical outcome was not measured or affected.

Note specific to mock submission: in an actual application, the use condition collection procedure could be part of the study and the results might not be available for analysis in a pre-IDE submission setting. In such a case, the sponsor would likely evaluate various scenarios of updated virtual patient performance when selecting a loss function.

1.4.3 Loss function overview

Simulated fracture survival of the 2014VP lead in virtual patients will be used as an informative prior, in an adaptation of the power prior method [1, 5, 6]. The weight of the virtual data will be adjusted using a loss function [2], which scales from 0 to 1 according to the similarity of the virtual patient and observed data. This loss function adjusts the amount of weight the prior

[4] Following the same procedure used for the AAMI Lead Imaging Study

receives thereby controlling the amount of borrowing from the model in the event of discordance between VPM and observed results.

If there is a high level of agreement between observed data and the VPM or the failure rate is lower for the observed data, the virtual patient data will be weighted at or near 100% . If the observed data performs worse than the virtual patient data, the loss function will assign very little or zero weight to the virtual patient data. By coupling the power prior method with a flexible loss function our proposed methodology allows an automated approach for determining the strength of the prior.

1.5 Adaptive Bayesian Design

The Bayesian adaptive design will 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 200 to 410 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 200th 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. Table 2 illustrates a potential enrollment scenario.

Table 2: Scenario for trial enrollment, assuming 0.5 subjects per day, 410 subjects enrolled, no subjects with events, and no early stopping

Stage	Subjects Enrolled	Expected Days into Trial	Subjects with 6 month follow-up completed	Subjects with 12 month follow-up completed	Subjects with 18 month follow-up completed
1	200	400	110	20	0
2	230	460	140	50	0
3	260	520	170	80	0
4	290	580	200	110	20
5	320	640	230	140	50
6	350	700	260	170	80
7	380	760	290	200	110
8	410	820	320	230	140
Close trial /complete		1360	410	410	410

Adaptive Bayesian sample size algorithm, also illustrated in Figure 3:

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 410 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 interim analysis, or if the maximum sample size of 410 is reached

4. The decision rule for trial success will be defined as the posterior probability of $P(\theta < 0.03|Data) \geq 0.95$. Operating characteristics for this design are provided in Section 3.

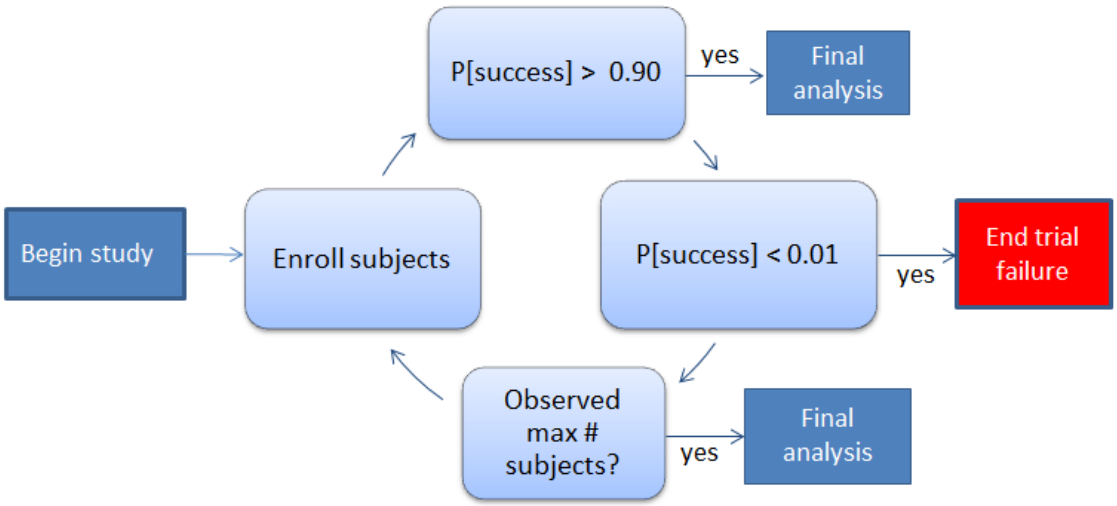

Figure 3: Adaptive Trial Algorithm

1.5.1 Final Analysis

Let *y* be an indicator for a safety event (subject experiences fracture) at 18 months for patients enrolled in the trial, where *y*=1 means subject experiences a lead fracture. The outcomes are modeled as

$y \sim Binomial(n, \theta)$

where, θ is modeled using an informative prior based on the simulated outcomes using the VPM (see section 2 for additional details). The informative prior can be expressed as $Beta(\theta_0^i * n_0 + 1, n_0 - \theta_0^i * n_0 + 1)$ for the i^{th} virtual patient cohort. The effective virtual patient sample size n_0 is not fixed and is used to control the amount of borrowing from the virtual patient model (see section 2.5). Note that all enrolled patients will have complete follow up at the time of final analysis. The safety rate θ will be compared to an OPC of 3% and study success will be claimed if there is at least 95% posterior probability of exceeding the OPC.

1.5.2 Interim Analysis

In this trial interim analysis will be conducted for sample size adaptation. The study may be stopped for futility based on the results of an interim analysis. Study success will only be determined at the final analysis and will not be claimed at an interim analysis. At the time of each interim analysis, some patients will not have completed the full evaluation period. The primary outcome of lead fracture is continually monitored remotely. Therefore, a longitudinal time-to-fracture model is 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 to impute final outcomes. The predictive probability model that will be used is a piecewise exponential. We will use data from types 1 and 2 to estimate the hazard rates for the piecewise exponential function. There are three segments, with knot points at 6 and 12 months. The use of three segments models the virtual patients well, and allows independent estimates for early, random, and wear-out failures.

We assume a gamma distribution prior for each of the exponential segments, which allows us to estimate the hazard rate using closed form conjugacy. Therefore, the posterior is also a gamma distribution, see section 2.6 for specific details. Weight given to the virtual patients will be determined by the loss function discussed in section 2.7.

2. Incorporation of Virtual Patient Data

2.1 Maximum weight from virtual patient data

The virtual patient sample size will be less than or equal to 160, (e.g. $n_0 \leq 160$). Because the virtual patient simulations can generate an extremely high number of simulated patients, the data will be scaled within the power prior to create an effective sample size of ≤ 160 patients. With a maximum of 160 virtual patients and a minimum of 200 real patients, the maximum percentage of virtual patients is 45% (160/(160+200)). This was determined based on the influence, risk, and credibility of the virtual patient model.

2.2 Definitions for the virtual patients used in the informative prior:

We define the simulated outcomes for the virtual patients as a vector y_0. To account for uncertainty in the engineering model, we make m_0 sets of virtual patients y_0^i. This is similar to a bootstrap procedure, where each dataset $y_0^1 \ldots y_0^{m_0}$ allows for inference about the uncertainty in

the parameters of the virtual patient model. The matrix representation below represents the entire set of virtual patients from the engineering model.

$$\begin{bmatrix} y_0^1 \\ \dots \\ y_0^i \\ \dots \\ y_0^{m_0} \end{bmatrix} = \begin{bmatrix} y_{01}^1 \dots y_{0N_0}^1 \\ \dots \\ y_{01}^i \dots y_{0N_0}^i \\ \dots \\ y_{01}^{m_0} \dots y_{0N_0}^{m_0} \end{bmatrix}$$

Each of the y_0^i vectors have N_0 elements. Each element represents an individual virtual patient outcome, in this case a binary outcome of success or failure after some time. Note that the virtual patient model actually produces a time to event, t_0 and corresponding event indicator, d_0. With t_0 and d_0, we can construct y_0 at any given time. We will use t_0 and d_0 for imputation purposes. A comprehensive overview of the methodology used for simulating virtual patients was provided in Q-sub Q150804/S001.

2.3 Incorporation of virtual patients using the power prior method

We assume the engineering model accurately predicts clinical outcomes and that the outcomes of virtual and clinical patients will be similar for the failure mode in the engineering model. Later in this document we discuss the use of a loss function to discount the number of virtual patients if there is disagreement between virtual and clinical data.

The prior dataset of virtual patient cohorts $y_0^1 \dots y_0^{m_0}$ discussed in section 3.2 represents a total of 500,000 simulated virtual patients ($m_0 = 500, N_0 = 1000$). Clearly, the number of virtual patients is substantially larger than the enrollment of real patients in the clinical study. Therefore it is necessary to downweight the number of virtual patients while still allowing the expression of uncertainty in the engineering model. In order to accomplish this, we employ a modification of the power prior as follows.

2.4 Constructing distribution of virtual patient fracture rate θ_0

From the virtual patient cohorts, $y_0^1 \dots y_0^m$, we can construct estimated failure rate at eighteen months as follows:

$$\theta_0^1 = \frac{1}{N_0} \sum_{i=1}^{N_0} y_{0i}^1 \qquad (1)$$

Performing the operation in (1) for each of the m_0 virtual patient cohorts generates the vector $\{\theta_0^1 \dots \theta_0^{m_0}\}$. This vector gives the distribution of θ_0. Note that m_0 will typically be large, therefore the empirical distribution is a good representation of the actual distribution.

2.5 Power prior implementation for binomial distribution

Let y be a vector of n outcomes for the current data, $\{y_1, y_2, \ldots, y_n\}$, where each individual y is a 0 or 1, with 1 indicating an event. Note that at an interim analysis, y will include imputed outcomes. Similarly, y_0^i is a vector of outcomes for the i^{th} virtual patient cohort.

$n_0 = 160 \times a_0$ is the effective virtual patient sample size. It can vary between 0 (no virtual patients) and 160 (maximum weight of virtual patients) based on the results of the a_0 loss function, discussed in section 3.3. When n_0 is 0, the evaluation above converts to traditional beta-binomial with a non-informative uniform prior.

Following the work of Ibrahim and Chen [1], the posterior probability for event rate θ using the power prior is given as follows:

$for\ i = 1 \ldots m_0$

$$\alpha = \sum y + \theta_0^i * n_0 + 1$$

$$\beta = n - \sum y + n_0 - \theta_0^i * n_0 + 1$$

draw one $\theta_i \sim beta(\alpha, \beta)$

This generates a vector of posterior values $\{\theta_1 \ldots \theta_{m_0}\}$, to be used for evaluating endpoints and constructing credible intervals.

2.6 Longitudinal modeling using piecewise exponential distributions

We assume a gamma distribution prior for each of the exponential segments, which allows us to estimate the hazard rate using closed form conjugacy. Therefore, the posterior is also a gamma distribution.

We partition duration time into 3 intervals and assume the hazard is constant within each interval. The event times are therefore modeled as:

$$T_{event} \sim PE(\lambda_1, \lambda_2, \lambda_3)$$

where, λ_1,λ_2 and λ_3 and the hazards for the three segments.

Let t_{0p}^i be the event or censor times in the p^{th} interval for the i^{th} cohort of virtual patients. Similarly, let t_p be the event or censor times in the p^{th} interval for the real patient. Let $d_{0p}^i = \{1,0\}$ where 1 indicates an event at time t_{0p}^i for virtual patients in the ith cohort. Similarly, let $d_p = \{1,0\}$ where 1 indicates an event at time t_p for real patients. Then the posterior is as follows:

$for\ i = 1 \ldots m_0$

$$\text{draw one } \lambda_p^i \sim gamma\left(\sum_{j=1}^{n} d_j + \frac{n_0}{N_0}\sum_{j=1}^{N_0} d_{0j}^i + .01\ ,\ \sum_{j=1}^{n} t_i + \frac{n_0}{N_0}\sum_{j=1}^{N_0} t_{0j}^i + .01\right)$$

2.7 Loss function: determining n_0

A loss function is used to adjust the strength of the prior according to the agreement between current and virtual patient data. This loss function approach was proposed by the MDIC working group and is also detailed in a draft publication [2]. At the final analysis, the loss function is evaluated using the pre-specified shape and scale parameters given below. This evaluation determines the number of virtual patients, n_0.

If we put a minimally informative prior on θ_c for the current data (say a uniform(0,1) prior) , then θ_c has a beta posterior distribution:

$$\theta_c \sim beta(y + 1, n - y + 1) \qquad (2)$$

From (2), we can draw m_0 instances of θ_c. Given the large m_0, we can stochastically compare the distribution of θ_c to θ_0 using a posterior Bayesian p-value [3] as:

$$p = P(\theta_c \leq \theta_0) = \frac{1}{m_0}\sum I\{\theta_c^i < \theta_0^i\} \qquad (3)$$

The desired characteristics of the loss function utilized here are:

1. If there is a high level of agreement between current and virtual patient data, the loss function should allow n_0 to provide close to full weight to the virtual patient data.

2. Conversely, if evidence of disagreement between current and virtual patient data emerges, then a_0 should start to down-weight the prior, i.e. n_0 approaches zero as p approaches zero.

There are many loss functions which can be constructed to down weight the prior as a function of p. The Weibull cumulative distribution function can meet these criteria:

$$F(p) = 1 - e^{-(p/\lambda)^k} \qquad (4)$$

We evaluate two candidate loss functions:

$\lambda = 0.05, k = 1.5$ (loss function #1) and $\lambda = 0.10, k = 1.5$ (loss function #2)

To convert the loss function to number of virtual patients (n_0), we multiple by n_{max} as follows:

$$n_0 = 160 * F(p)$$

Recall that 160 is the maximum number of virtual patients. Sample values are listed in Table 3 below and illustrated in Figure 4.

Table 3: Candidate loss functions

p	$1 - e^{-(p/.1)^{1.5}}$ loss function #1	$1 - e^{-(p/.05)^{1.5}}$ loss function #2
0.01	0.03	0.09
0.05	0.30	0.63
0.1	0.63	0.94
0.2	0.94	1.00
0.5	1.00	1.00

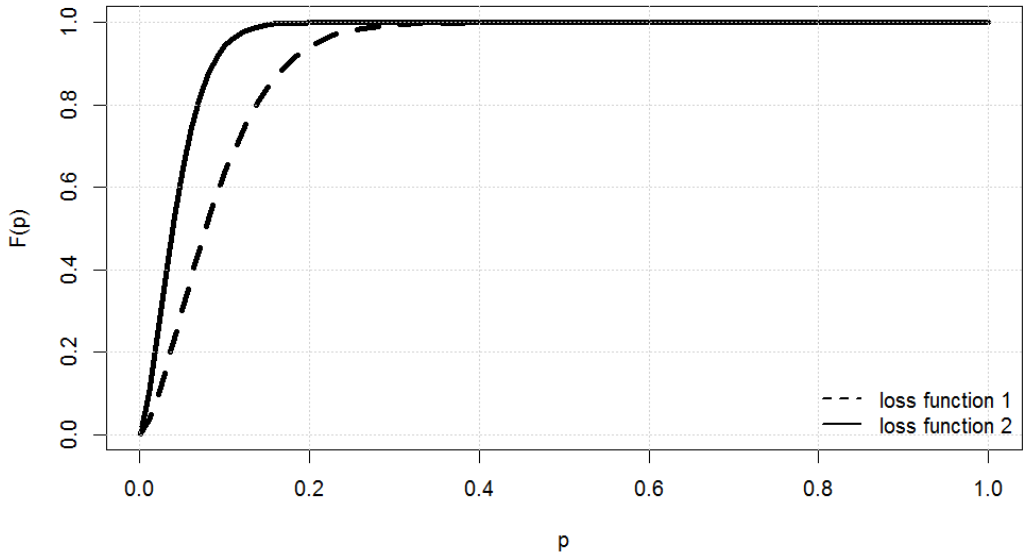

Figure 4: Candidate loss functions

Note that the loss functions proposed here do not reduce the strength of the prior when the current study outperforms the virtual patient 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. We propose to use loss function #2 ($\lambda = 0.1$ $k = 1.5$) in this trial based on the operating characteristics discussed below.

3. Clinical trial operating characteristics

The choice of a loss function is a major consideration for this methodology. We evaluate the two loss functions discussed above. Our criteria for selecting loss function parameters λ and k were twofold:

1. Power of $\geqslant$ 80% at a 1% failure rate, which was the upper 95th confidence bound of our virtual patient model[5].

2. Type I error $\leqslant$ 10% at the performance criteria of 3%.

Simulations were performed with true failure rate θ_t of 0.001, 0.01, 0.015, 0.02, 0.025, and 0.03. Recall that 0.03 is the null hypothesis, and a trial success with a true failure rate of 0.03 is a type I error. 800 simulated trials were run at each value of true failure rate. Enrollment was assumed to follow a Poisson distribution with an expected rate of 0.5 patients / day, with hazard rate increasing by factors of two and three for the second and third intervals respectively. This allows characterization of power and type I error, as well as number of enrolled patients (n) and the number of virtual patients (n_0), all as a function of the true failure rate. Table 4 and Figure 5 show the outcome with two Weibull loss functions as well as the two extreme cases of a fixed prior ($n_0 = 160$) and an uninformative prior ($n_0 = 0$).

Table 4: Operating characteristics for fixed, non-informative, and loss function weighted priors

Prior	power at 1% rate	type 1 error rate	Average number of VP (n_0)	average number of enrolled (n) at 1% rate	rate of early stop for futility at 3%	rate of early stop for success at 1%
Fixed prior (N0=160)	0.96	0.29	160	209	9%	98%
Non-informative prior	0.64	0.03	0	248	44%	85%
Loss function #1	0.78	0.05	89	244	41%	89%
Loss function #2	0.85	0.10	121	240	34%	89%

From Table 4, we see that the fixed prior has a type I error, 29%, which does not meet our criteria. The non-informative prior meets the type I error objective, but has a low power at a failure rate of 1%, 64%. Loss function #1has attractive properties of 78% power at the 1% failure rate, with very little change in type I error compared to the non-informative prior. However, loss function #1 does not meet the criteria of $\geq$ 80% power. Finally, loss function #2 satisfies both type I error and power criteria. Therefore we propose using loss function #2 for the clinical study.

[5] From table 1, upper 95th percentile failure rate was 1.2% after updated use conditions.

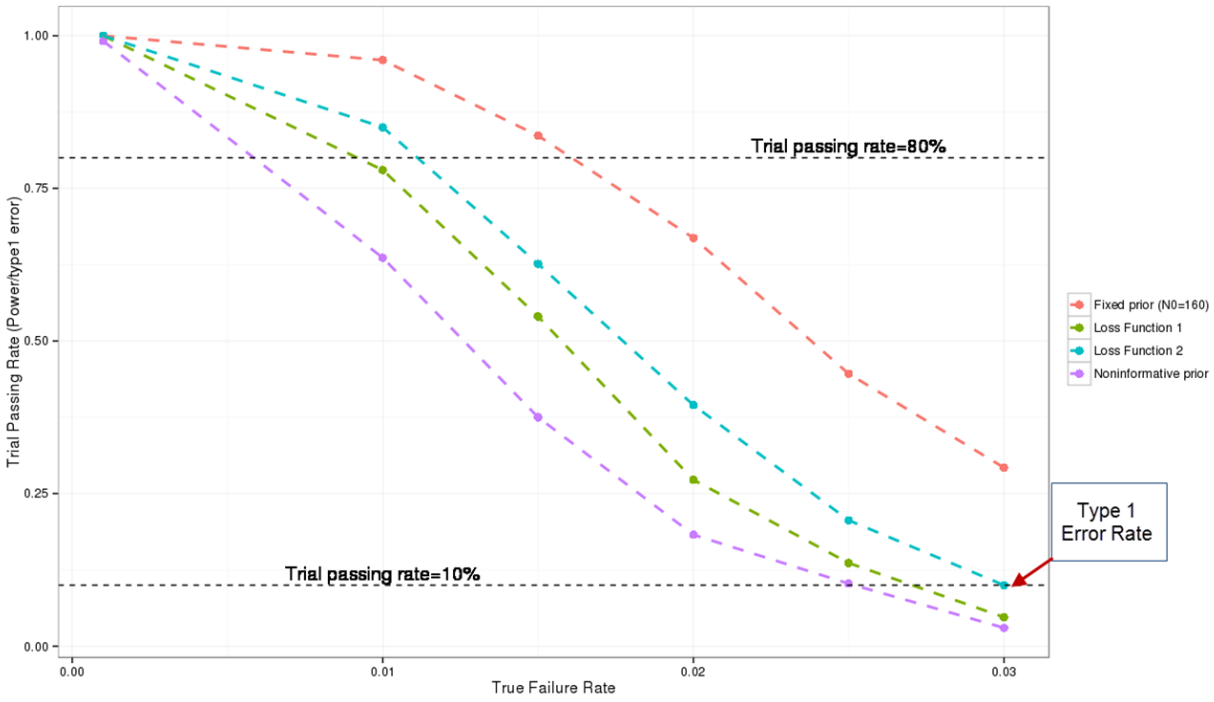

Figure 5: Clinical trial operating characteristics

The desirable characteristics of the loss function approach can be seen apparent in Figure 5. In the ideal case, at low failure rates (e.g. ≤ 1%), we wish to have the high power of the fixed prior. At the null hypothesis (3%), we wish to have the low type I error of the non-informative prior. The loss function curves display similar characteristics to the fixed prior at low failure rates (i.e. high power) and are more similar to the non-informative prior at the null hypothesis (i.e. low type I error).

Additional characteristics of the study, using loss function #2, are shown in Tables 5, 6, and 7. Table 5 shows the characteristics for a range of true failure rates. Tables 6 and 7 show a range of enrollment scenarios and operating characteristics when the true failure rate is 1% and 3%.

Table 5: Operating Characteristics with varying true failure rates, using loss function #2

Failure rate (θ_t)	**Probability of trial success**	**Average Enrollment**	**Average Virtual Patient Sample Size**	**Trial Duration (months)**	**Probability of early stop for futility**
0.001	1.00	204	160	19.7	0.000
0.01	0.85	240	121	22.1	0.004
0.015	0.63	256	84	23.2	0.026
0.02	0.40	271	54	24.2	0.100
0.025	0.21	280	32	24.8	0.218
0.03	0.10	289	18	25.4	0.348

Table 6: Enrollment scenarios and operating characteristics with true failure rate of 1%

Subjects enrolled	Proportion of Trials	Probability of futility	Stop enrollment early for expected success	Probability of success	Average Virtual Patient Sample Size	Stop and win	Stop and lose
200	0.67	0.00	1.00	0.86	129.61	0.86	0.14
230	0.09	0.00	1.00	0.82	122.07	0.82	0.18
260	0.04	0.00	1.00	0.80	104.71	0.80	0.20
290	0.04	0.03	0.97	0.81	115.88	0.84	0.16
320	0.03	0.05	0.95	0.77	90.23	0.81	0.19
350	0.01	0.00	1.00	1.00	115.75	1.00	0.00
410	0.12	N/A	N/A	0.85	84.67	N/A	N/A

Table 7: Enrollment scenarios and operating characteristics with true failure rate of 3%

Subjects enrolled	Proportion of Trials	Probability of futility	Stop enrollment early for expected success	Probability of success	Average Virtual Patient Sample Size	Stop and win	Stop and lose
200	0.37	0.06	0.94	0.19	33.91	0.21	0.79
230	0.07	0.48	0.52	0.12	22.41	0.23	0.77
260	0.07	0.64	0.36	0.09	14.05	0.24	0.76
290	0.06	0.92	0.08	0.04	5.42	0.50	0.50
320	0.08	0.89	0.11	0.03	4.27	0.29	0.71
350	0.07	0.98	0.02	0.00	3.38	0.00	1.00
410	0.27	0.19	N/A	0.03	6.96	N/A	N/A

Figure 6 illustrates the effect of the loss function on the number of virtual patients given the number of observed events in the study. Number of observed events is correlated to true underlying failure rate. For example, with a true failure rate in the range of 0.02 - 0.025, the expected number of events is approximately 4, and the expected number of virtual patients is approximately 40. For a true failure rate in the range of 0.005 - 0.01, the expected number of events is approximately 2 and the expected number of virtual patients is approximately 135. Note that the relationship between events and true failure rate is a probabilistic relationship that also depends on enrollment rate.

From Figure 6, loss function #1 penalizes more severely for observed events compared to loss function #2. For example, after one observed event, loss function #1 begins to decrement the weight of virtual patients, while loss function #2 continues to apply full weight. After two or three events, the virtual patient weight is approximately 50% of the maximum weight (for loss

functions #1 and #2 respectively). Note that after approximately seven observed events, the virtual patient weight is negligible for both loss functions.

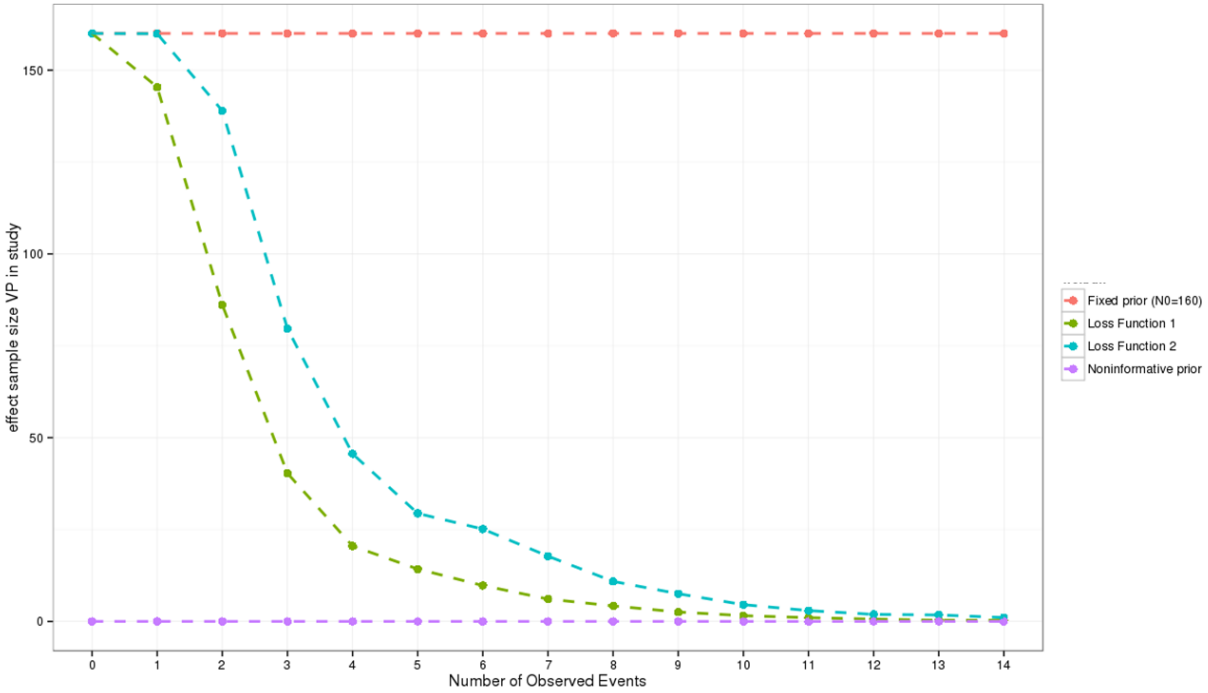

Figure 6: Number of virtual patients vs. number of observed events

Figure 7 shows average number of enrolled and virtual patients 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 virtual patient data. The maximum total enrollment occurs near or at a failure rate of 3%. The difference between the two loss functions can be seen in the right panel of Figure 7. Loss function #1 begins to penalize the virtual patients at a lower failure rate than loss function #2. By comparing Figures 6 and 7 one can make a correspondence between number of events in the study and true failure rate. For example, 2 observed events roughly corresponds to a 1% failure rate.

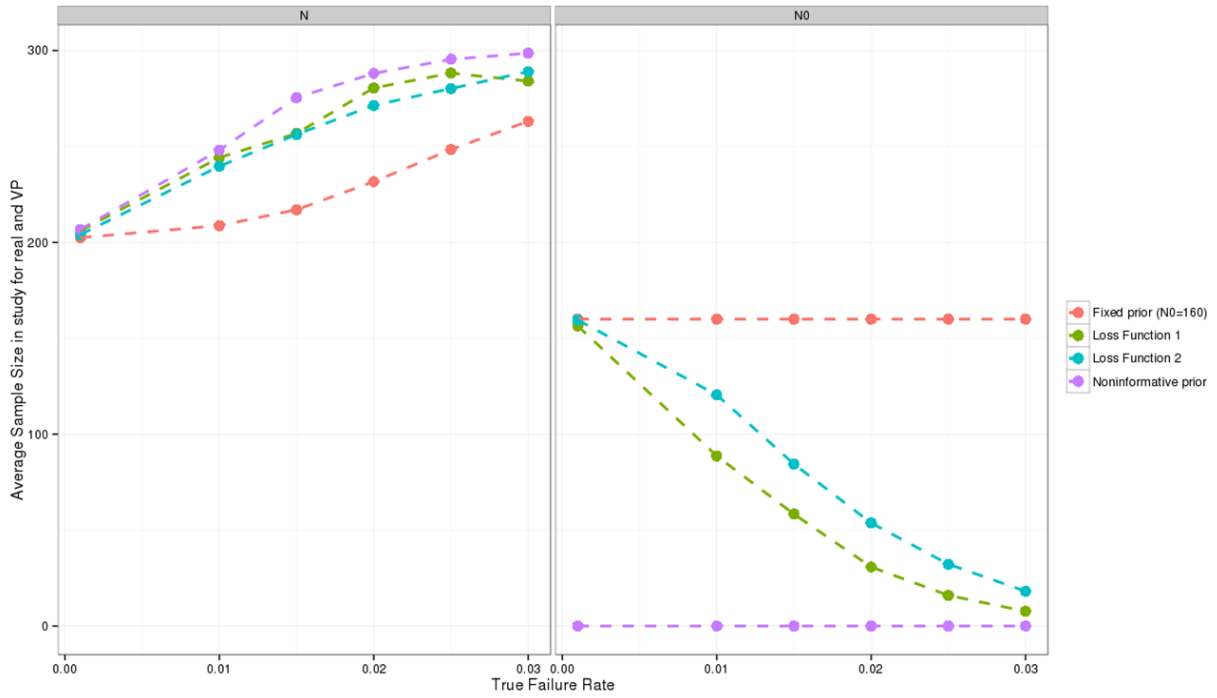

Figure 7: Number of enrolled (N, left panel) and virtual (N_0, right panel) patients vs. true failure rate

References:

[1] J. G. Ibrahim and M.-H. Chen, "Power prior distributions for regression models," *Statistical Science*, vol. 15, no. 1, pp. 46-60, 2000.

[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*

[3] A. Gelman, J. Carlin, H. Stern and D. Rubin, Bayesian Data Analysis, Boca Raton: Chapman & Hall / CRC, 2004.

[4] Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2010). Bayesian adaptive methods for clinical trials. CRC press.

[5] De Santis, Fulvio. "Using historical data for Bayesian sample size determination." Journal of the Royal Statistical Society: Series A (Statistics in Society) 170.1 (2007): 95-113.

[6] Ibrahim, Joseph G., et al. "The power prior: theory and applications." Statistics in medicine 34.28 (2015): 3724-3749.

[7] AAMI_PC_WG1_N169_E, "Outline of Requirements for Fatigue Performance of Cardiac Rhythm Management Leads"

[8] Haddad, Himes, Campbell, "Fracture prediction of cardiac lead medical devices using Bayesian networks", Reliability Engineering and System Safety, 123(2014); 145-157.