

# Robustness Evaluation of Computer-aided Clinical Trials for Medical Devices

Kuk Jin Jang  
The University of Pennsylvania  
jangkj@seas.upenn.edu

Yash Vardhan Pant  
The University of Pennsylvania  
yashpant@seas.upenn.edu

Bo Zhang  
Wharton School, The University of  
Pennsylvania  
bozhan@wharton.edu

James Weimer  
The University of Pennsylvania  
weimerj@seas.upenn.edu

Rahul Mangharam  
The University of Pennsylvania  
rahulm@seas.upenn.edu

## ABSTRACT

Medical cyber-physical systems, such as the implantable cardioverter defibrillator (ICD), require evaluation of safety and efficacy in the context of a patient population in a clinical trial. Advances in computer modeling and simulation allow for generation of a simulated cohort or *virtual cohort* which mimics a patient population and can be used as a source of prior information. A major obstacle to acceptance of simulation results as a source of prior information is the lack of a framework for explicitly modeling sources of uncertainty in simulation results and quantifying the effect on trial outcomes.

In this work, we formulate the Computer-Aided Clinical Trial (CACT) within a Bayesian statistical framework allowing explicit modeling of assumptions and utilization of simulation results at all stages of a clinical trial. To quantify the robustness of the CACT outcome with respect to a simulation assumption, we define  $\delta$ -robustness as the minimum perturbation of the base prior distribution resulting in a change of the CACT outcome and provide a method to estimate the  $\delta$ -robustness.

We demonstrate the utility of the framework and how the results of  $\delta$ -robustness evaluation can be utilized at various stages of a clinical trial through an application to the Rhythm ID Goes Head-to-head Trial (RIGHT), which was a comparative evaluation of the safety and efficacy of specific software algorithms across different implantable cardiac devices. Finally, we introduce a hardware interface that allows for direct interaction with the physical device in order to validate and confirm the results of a CACT for implantable cardiac devices.

## CCS CONCEPTS

• **Computing methodologies** → **Uncertainty quantification;**  
*Simulation evaluation;*

## KEYWORDS

Computer-aided, Clinical trials, Bayesian sensitivity analysis, Robustness, Medical devices, Implantable cardiac devices

## ACM Reference Format:

Kuk Jin Jang, Yash Vardhan Pant, Bo Zhang, James Weimer, and Rahul Mangharam. 2019. Robustness Evaluation of Computer-aided Clinical Trials

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ICCCPS '19, April 16–18, 2019, Montreal, QC, Canada

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6285-6/19/04...\$15.00

<https://doi.org/10.1145/3302509.3311058>

for Medical Devices. In *10th ACM/IEEE International Conference on Cyber-Physical Systems (with CPS-IoT Week 2019) (ICCCPS '19)*, April 16–18, 2019, Montreal, QC, Canada. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3302509.3311058>

## 1 INTRODUCTION

The evaluation of the safety and performance of complex devices and systems is one of the fundamental challenges of cyber-physical systems. For medical cyber-physical systems, such as the pacemaker or implantable cardioverter defibrillator (ICD), the gold standard for evaluating safety and efficacy of a device is in the context of a patient population in a clinical trial. However, clinical trials can cost \$10-20 million and can last anywhere from 4-6 years. Despite this, rigorously planned clinical trials obtain undesirable results [13]. Even with a successful trial, medical devices are still subject to recalls and safety alerts [21].

As medical devices increase in complexity, using clinical trials alone to evaluate newer technologies will be increasingly insufficient. Computer models of physiology and simulation can potentially be utilized in clinical trials by considering them as an alternative source of prior information [11][9]. In such a trial, the data that is measured and analyzed in a clinical trial, called endpoints, are simulated using a physiological model and subsequently applied to either a device model or the physical device itself. However, physics-based models are difficult to develop and even when they are available, incur large variability in the simulated endpoints due to the uncertainty in the parameters [10]. Moreover, these simulations are based on assumptions and can cause large errors in the evaluation of a device when the assumptions differ from reality.

A major obstacle in the usage of simulated endpoints as a source of prior information in clinical trials is the lack of a method to quantify the uncertainty caused by assumptions in a simulation and the effect on clinical trials outcomes.

In [1], a Computer-Aided Clinical Trial (CACT) was defined as a process in which high-level models of the human physiology are used to evaluate device performance on virtual trial populations, and the results of which are used to help plan a real clinical trial. In this work, we expand the definition to include the utilization of simulated endpoints obtained from the simulation of physiological models in the planning, conduct and analysis of a clinical trial. This definition reflects the potential for incorporating simulation results in all stages of a medical device trial.

*Problem:* Given a process to generate a cohort of simulated endpoints and a CACT which utilizes the cohort in the outcome, we wish to estimate the effect of assumptions made during simulation on the outcome of the CACT.

If changes in a simulation assumption have little effect on the overall outcome, then the CACT outcome can be considered 'robust'

with respect to that particular assumption. We aim to quantify this notion of ‘robustness’ of the outcomes of a CACT for a medical device and estimate the value through the following contributions:

- (1) We provide a Bayesian statistical framework for modeling a CACT and its outcome. The framework incorporates results from simulated endpoints through the *pre-clinical prior* and explicitly models simulation assumptions (Section 2).
- (2) We define  $\delta$ -robustness of a CACT to quantify the uncertainty in the CACT outcome due to simulation assumptions and present a method for estimating the value (Section 3).
- (3) We apply the framework in a case study of a previous trial, the Rhythm ID Goes Head-to-head Trial (RIGHT), which evaluated the relative safety and efficacy of two software algorithms in implantable cardiac devices, and confirm that the results of the CACT align with the original trial. Furthermore, we demonstrate how the  $\delta$ -robustness value can be used in both the planning stage of a clinical trial and at trial conclusion through the *robustness plane* (Section 4).
- (4) We introduce a hardware testbed that enables a CACT for ICDs to be automatically conducted through direct interaction with the physical device (Section 5).

## 2 CACT BAYESIAN MODELING FRAMEWORK

### 2.1 Overview of CACTs for medical devices

A clinical trial evaluates a medical device through an inquiry in the form of a hypothesis test about a parameter of interest,  $\theta$ , with a set of observations called *endpoints*,  $y = \{y_1, y_2, \dots, y_N\}$  from a patient cohort of sample size  $N$ . Examples of endpoints include continuous endpoints, such as the amount of blood pressure reduction, and discrete endpoints, such as the number of inappropriate therapies. The typical hypothesis test compares the likelihood of a null hypothesis,  $H_0$ , with one or more alternative hypotheses,  $H_1, H_2, \dots$ , etc. The *outcome* of the trial is either the rejection of the null hypothesis in favor of an alternative hypothesis or the failure to reject the null hypothesis.

As shown in Fig. 1, the clinical trial can be divided into three stages: trial planning, trial conduction, and trial conclusion. Two general approaches to clinical trial evaluation exist. The standard, frequentist approach to a clinical trial considers only the observed endpoints of the trial cohort to answer the primary question. In a Bayesian clinical trial, a prior distribution (e.g. non-informative uniform prior) is placed on the parameter of interest during the planning stage and then updated to form a posterior distribution using observations from the trial. The trial outcome is determined with this posterior distribution.

Similar to the standard clinical trial, a CACT is divided into three stages paralleling the stages of a clinical trial: pre-clinical simulation, interim-trial simulation, and post-trial simulation. During pre-clinical simulation, a physiological model is used to generate a *virtual cohort of physiological signals* (Fig. 1 ②) and is applied to a *device model* in order to generate a cohort of size  $N_o$  of simulated endpoints,  $y_o = \{y_{o1}, y_{o2}, \dots, y_{oN_o}\}$ , called the *virtual endpoint cohort* or *virtual cohort* (Fig. 1 ③).

These endpoints are used to update a non-informative prior distribution to form an informed prior, which we call the *pre-clinical simulation prior distribution* or *pre-clinical prior* (Fig. 1 ④). At the pre-clinical simulation stage, the pre-clinical simulation prior is used to predict the *pre-clinical CACT outcome* of the trial (Fig. 1 ⑤).

The pre-clinical prior is updated when data from real patients become available during the trial conduction and at the conclusion.

The pre-clinical prior is weighted according to the similarity between the virtual cohort and the real patient cohort to form the *pre-clinical power prior* (Fig. 1 ⑥).

During post-trial simulation, the *post-trial simulation posterior distribution* is obtained by updating the pre-clinical power prior (Fig. 1 ⑦). Finally, the overall *post-trial simulation CACT outcome* is computed (Fig. 1 ⑧).

In the case of standard clinical trials, assumptions are made about the target population in order to compute parameters of the trial, such as the sample size. Similarly, in a CACT, simulation assumptions, such as the composition of conditions in a virtual cohort, are modeled within the framework as the CACT prior distribution (Fig. 1 ①). Sec. 3 describes the concept of *robustness of CACT outcomes* and how to evaluate the effect of the CACT prior on the CACT outcome (Fig. 1 ⑨, ⑩).

In the next section, we formally define the concepts presented in this section within a Bayesian statistical framework.

### 2.2 Bayesian modeling framework for CACTs

In order to draw conclusions about the parameter of interest  $\theta$ , we build upon the Bayesian framework for modeling a virtual cohort proposed in [15]. In this work, we have expanded the framework in order to reflect the different stages of a CACT and model sources of uncertainty.

Given a set of simulated endpoints,  $y_o = \{y_{o1}, y_{o2}, \dots, y_{oN_o}\}$ , where  $Y_o \sim f_{Y_o}(y_o | \theta_o; \gamma_o)$  and a set of real patient cohort endpoints  $y = \{y_1, y_2, \dots, y_N\}$ , where  $Y \sim f_Y(y | \theta)$ , the conclusions drawn from the CACT regarding  $\theta$  rely on the following assumption about the exchangeability of the two sets of endpoints:

ASSUMPTION 1 (EXCHANGEABILITY OF SIMULATION ENDPOINTS). *The device model and the physiological model accurately capture the variability of outcomes and generate simulated endpoints,  $y_o$ , such that,*

$$\theta = \theta_o \quad (1a)$$

$$\text{and } f_{Y_o}(y_o | \theta; \gamma_o) = f_Y(y | \theta) \quad (1b)$$

This implies that the information about the parameter of interest obtained through the virtual cohort is the same as that of an actual patient cohort. Moreover, (1b) implies that the distribution of the endpoints is equivalent. (1a) and (1b) must be verified during and after the conduction of the trial when endpoints from real patients,  $y$ , becomes available. Here,  $\gamma_o$  is a parameter related to the simulation assumption which will be defined in the next section.

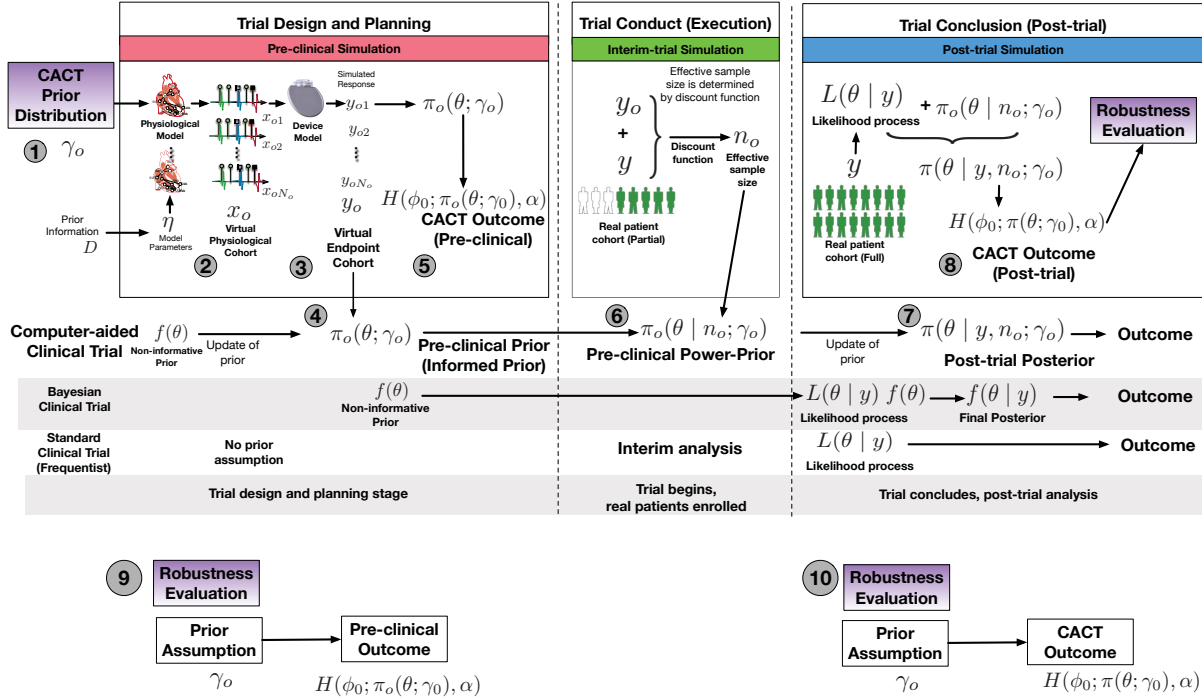
**CACT pre-clinical simulation** We first define the virtual cohort and the relation to the simulation assumptions in the pre-clinical stage and how the assumptions influence the pre-clinical trial outcome. We assume the existence of a device model and an adequate physiological model which generates a cohort of size  $N_o$  of physiological signals,  $x_o = \{x_{o1}, x_{o2}, \dots, x_{oN_o}\}$  (Fig. 1 ②) and define the *virtual physiological cohort*:

DEFINITION 1 (VIRTUAL PHYSIOLOGICAL COHORT). *The virtual physiological cohort or physiological cohort,  $x_o$ , is a set of of size  $N_o$  consisting of I.I.D. instances of the multivariate random variable,  $X_{oi}$ , where,*

$$X_{oi} \sim f_{X_{oi}}(x_{oi} | \psi) \quad (2)$$

*$f_{X_{oi}}$  is the distribution of  $X_{oi}$  indexed by the parameter  $\psi$  and is called the distribution of physiological signals conditioned on  $\psi$ . Therefore, for the set  $X_o = \{X_{o1}, X_{o2}, \dots, X_{oN_o}\}$ ,*

$$f_{X_o}(x_o | \psi) = \prod_{i=1}^{N_o} f_{X_{oi}}(x_{oi} | \psi) \quad (3)$$



**Figure 1: Overview of a Computer-Aided Clinical Trial (CACT) for medical devices and robustness evaluation. A CACT is divided into three phases: pre-clinical simulation (trial design and planning), interim-trial simulation (trial conduct), and post-trial simulation (trial conclusion)**

The physiological cohort represents a set of signals that are used by the device, such as the cardiac electrograms for a pacemaker or the blood glucose level in an insulin pump. Typically, there is no closed form for  $f_{X_o}$ , but samples can be obtained from the distribution by simulating the physiological model. The cohort generation process is equivalent to obtaining samples from the distribution of  $X_o$  for which  $\psi$  is a parameter setting in that process. For example, if the physiological model generates  $k$  types of cardiac rhythms,  $\psi = (\psi_1, \psi_2, \dots, \psi_k)$  can be the number of each rhythm type in the physiological cohort. The proportion of each rhythm type in a physiological cohort is an example of a simulation assumption. This assumption is modeled within a Bayesian framework as a prior distribution on the parameter  $\psi$ , which we define as the *CACT prior distribution*:

**DEFINITION 2 (CACT PRIOR DISTRIBUTION).** The parameter  $\Psi$  of the virtual cohort is a random variable such that,

$$\Psi \sim \pi(\psi | \gamma_o) := \pi_{\gamma_o}(\psi) \quad (4)$$

where,  $\gamma_o$  is the set of parameters for the prior distribution.  $\pi_{\gamma_o}(\psi)$  is defined as the CACT prior distribution.

$\pi_{\gamma_o}(\psi)$  encodes information of simulation assumptions by specifying the form of the function and parameters  $\gamma_o$ . This information may be available from reports of previous clinical trials, otherwise a conservative base prior may be selected, such as the uniform distribution.

Using this prior, we define the marginal distribution of physiological cohorts,  $f_{X_o}(x_o; \gamma_o)$  as:

$$f_{X_o}(x_o; \gamma_o) = E_{\Psi}(X_o, \Psi) = \int f_{X_o}(x_o | \psi) \pi_{\gamma_o}(\psi) dF(\Psi) \quad (5)$$

This can be considered the predictive prior distribution in Bayesian analysis. Note, the integration is over the support of the random variable  $\Psi$ . For simplicity of notation, we will suppress the upper and lower limits throughout the paper.

Intuitively, the marginal distribution obtains a weighted average of  $x_o$ , where the weight is determined by the CACT prior distribution,  $\pi_{\gamma_o}(\psi)$ . The value of the parameter,  $\gamma_o$ , influences the marginal distribution of cohorts.

A closed form of  $f_{X_o}(x_o; \gamma_o)$  is typically unavailable, but we assume that obtaining instances from the distribution of physiological cohorts,  $x_o$ , is possible. Therefore, marginalization is implemented using sampling schemes, such as Monte Carlo methods.

The virtual cohort,  $y_o$  (Fig. 1 ③) is obtained by applying each of the signals in the physiological cohort to the device model:

**DEFINITION 3 (VIRTUAL ENDPOINT COHORT).** Given a physiological cohort  $x_o$ , the virtual endpoint cohort or virtual cohort,  $y_o$ , is a multivariate random variable  $Y_o$  of dimension  $N_o$ , such that,

$$Y_o \sim f_{Y_o}(y_o | x_o, \theta_o; \gamma_o) \quad (6)$$

where  $f_{Y_o}$  is the distribution of  $Y_o$  indexed by  $\theta_o$  and  $\gamma_o$  is the parameter value for the CACT prior distribution.

Here,  $\theta_o$  is value of the parameter of interest that is obtained with the virtual cohort. Note, the endpoints comprising the virtual cohort are simulations of the outcomes in a patient cohort, such as the number of inappropriate therapy for a device, and not the physiological signals. As before, the marginal distribution  $f_{Y_o}(y_o | \theta_o; \gamma_o)$  is defined as:

$$f_{Y_o}(y_o | \theta_o; \gamma_o) = E_{X_o}(y_o, x_o | \theta_o; \gamma_o) = \int f_{y_o}(y_o | x_o, \theta_o; \gamma_o) dF(X_o) \quad (7)$$

From the virtual cohort, we wish to obtain estimates of  $\theta$ . Assuming (1a) and (1b), for a single virtual cohort  $y_o$ , we define the likelihood function with respect to  $\theta$  as  $L(\theta | y_o; \gamma_o) = f_{Y_o}(y_o | \theta; \gamma_o)$ .

Following a typical Bayesian procedure, we place a minimally-informative prior on  $\theta$ ,  $\pi_o(\theta)$ , such as the uniform prior, and define the posterior distribution of  $\theta$ ,  $\pi_o(\theta | y_o)$  for a single virtual cohort



$y_o$  such that:

$$\pi_o(\theta | y_o; \gamma_o) \propto L(\theta | y_o; \gamma_o) \pi_o(\theta) \quad (8)$$

In special cases when  $\pi_o(\theta)$  and  $f_{Y_o}$  are conjugates, then we have the closed-form of the posterior:

$$\pi_o(\theta | y_o; \gamma_o) = f_{Y_o}(y_o | \theta; \gamma_o) \pi_o(\theta) \quad (9)$$

As before, we marginalize out the uncertainty from  $y_o$  to obtain the *pre-clinical prior distribution of  $\theta$*  (Fig. 1 ④):

DEFINITION 4 (PRE-CLINICAL SIMULATION PRIOR DISTRIBUTION).

For the parameter of interest  $\theta$  and (4) with parameter set to  $\gamma_o$ , the pre-clinical simulation prior distribution or pre-clinical prior distribution for  $\theta$  is defined as,

$$\pi_o(\theta; \gamma_o) = \int \pi_o(\theta | y_o; \gamma_o) dy_o \quad (10)$$

Similar to before, this can be thought of as a weighted average of  $\theta$ , where the weights are determined according to the distribution of  $y_o$ . Note, (10) is an informed prior and is different from the CACT prior.

Finally, the pre-clinical prior distribution is used to determine the pre-clinical simulation outcome. Consider evaluating the assertion  $\theta \in \phi_o$ , i.e., the parameter of interest falls in the region  $\phi_o$ . For instance, when evaluating two medical devices, this could mean: the difference of inappropriate therapy rate of Device A and Device B is non-negative (A is worse than B), i.e.,  $\theta \geq 0$ . Here  $\phi_o = [0, \infty)$ .

We define a function which encodes the results of evaluating the assertion  $\theta \in \phi_o$  based on the posterior distribution of  $\theta$  as the outcome of pre-clinical simulation (Fig. 1 ⑤):

DEFINITION 5 (PRE-CLINICAL SIMULATION CACT OUTCOME).

$$H(\phi_o; \pi_o(\theta; \gamma_o), \alpha) = \begin{cases} 1 & \text{if } P(\pi_o(\theta; \gamma_o) \in \phi_o) \geq 1 - \alpha \\ 0 & \text{o.w.} \end{cases} \quad (11)$$

Here, an assertion of  $\theta \in \phi_o$  means that at least  $(1 - \alpha)\%$  of the posterior distribution of  $\theta$  is supported on  $\phi_o$ . The relation to the assumption is denoted by  $\gamma_o$ . For the previous example, with  $\alpha = 0.05$ ,  $H(\phi_o; \pi_o(\theta; \gamma_o), \alpha) = 1$  means that the probability that Device A has a higher inappropriate therapy rate than Device B is at least 95%.

In addition to the outcome itself, (10) can be used to estimate various statistics, such as the mean and variance of  $\theta$  which is used to derive parameters relevant to the design of the clinical trial, such as the desired sample size  $N$ .

**CACT interim simulation and post-trial simulation** Next, we describe how to combine the results from pre-clinical simulation with endpoints from real patients available during trial conduct and at the conclusion of the trial. In order to combine the virtual cohort  $y_o$  with the real cohort  $y$ , the power-prior framework [16] is applied to virtual cohorts to define the pre-clinical simulation power-prior distribution for a single cohort  $y_o$ :

DEFINITION 6 (PRE-CLINICAL SIMULATION POWER-PRIOR DISTRIBUTION FOR  $y_o$ ). For a single virtual  $y_o$  of size  $N_o$  and real patient outcomes  $y$ , the pre-clinical simulation power-prior distribution for  $y_o$  is defined as,

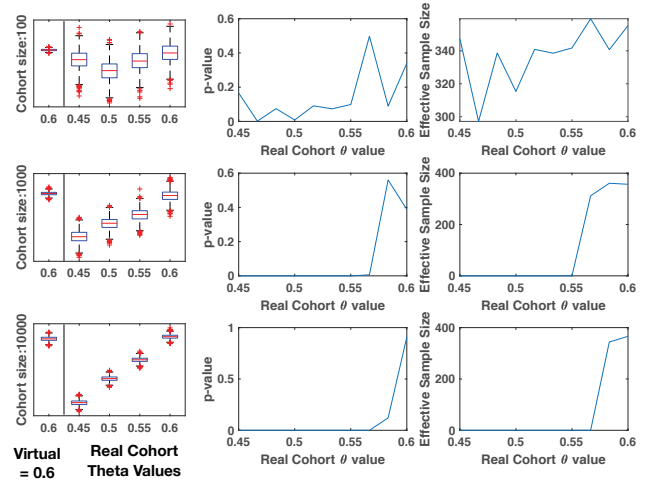
$$\pi_o(\theta | y_o, n_o; \gamma_o) \propto L(\theta | y_o; \gamma_o)^{\frac{n_o}{N_o}} \pi_o(\theta) \quad (12)$$

where  $n_o$ , the effective virtual cohort sample size is determined through a discount function which evaluates the similarity of  $y$  and  $y_o$ .

An example of a discount function and a method for determining the parameters of the function is proposed in [14]. Fig. 2, illustrates

how the discount function monitors the similarity between the virtual cohort and the real cohort by outputting a *p-value*. In the example, the real cohort follows a Binomial distribution with proportion parameter  $\theta$  ranging from 0.45 to 0.6. When the virtual cohort is sampled from a Binomial distribution  $Bin(N, p = 0.6)$  the *p-value* output from the increases as the real  $\theta$  value approaches 0.6 and the effective sample size,  $n_o$  increases accordingly. Also, as observed in Fig. 2, as the real cohort size is increased from 100 to 10000 the influence of the virtual cohort is only apparent when the two cohorts are more similar.

The details of applying the procedure to this work is described in Section 4. A discussion about choosing the appropriate parameters for the discount function is provided in Sec. 5.1.



**Figure 2: Monitoring of CACT prior using discount function. As the virtual cohort becomes similar to the real cohort, the *p-value* increases and the corresponding effective sample size  $n_o$  increases. With a larger real cohort, the effective sample size only increases with a strong degree of similarity.**

As in (7), the uncertainty in the virtual cohort  $y_o$  is marginalized out (Fig. 1 ⑥):

DEFINITION 7 (PRE-CLINICAL SIMULATION POWER-PRIOR DISTRIBUTION).

$$\pi_o(\theta | n_o; \gamma_o) = \int L(\theta | y_o; \gamma_o)^{\frac{n_o}{N_o}} \pi_o(\theta) dF(y_o) \quad (13)$$

(13) can be interpreted as a weighted average over the different virtual cohorts  $y_o$ , where the weights are determined according to the distribution of  $y_o$ .

For the real patient outcomes  $y$ , we define the likelihood  $L(\theta | y) = f_Y(y | \theta)$ , according to assumption (1b).  $L(\theta | y)$  and (13) are combined to form the post-trial posterior distribution (Fig. 1 ⑦):

DEFINITION 8 (POST-TRIAL SIMULATION POSTERIOR DISTRIBUTION).

$$\pi(\theta | y, n_o; \gamma_o) \propto L(\theta | y) \pi_o(\theta | n_o; \gamma_o) \quad (14)$$

If  $f_Y(y | \theta)$  and  $\pi_o(\theta; \gamma_o)$  are conjugates then,

$$\pi(\theta | y, n_o; \gamma_o) = f_Y(y | \theta) \pi_o(\theta | n_o; \gamma_o) \quad (15)$$

The effective sample size,  $n_o$ , can be interpreted as discounting the influence of the virtual cohort  $y_o$  on the overall posterior and the outcome if  $y$  and  $y_o$  differ greatly.

Finally, based on the post-trial posterior distribution  $\pi(\theta | y, n_o; \gamma_o)$ , the post-trial outcome of the CACT is defined (Fig. 1 ⑧):

DEFINITION 9 (POST-TRIAL SIMULATION CACT OUTCOME).

$$H(\phi_0; \pi(\theta; \gamma_o), \alpha) = \begin{cases} 1 & \text{if } P(\pi(\theta; \gamma_o) \in \phi_0) \geq 1 - \alpha. \\ 0 & \text{o.w.} \end{cases} \quad (16)$$

In Sec. 4, we apply the framework to the Rhythm ID Goes Head-to-head Trial (RIGHT). This trial sought to compare the performance of the discrimination algorithm in ICDs from two different manufacturers[4].

To quantify the influence of the CACT prior on the CACT outcome, we next define the concept of  $\delta$ -robustness and present a method for estimating  $\delta$ -robustness.

### 3 ESTIMATION OF THE $\delta$ -ROBUSTNESS VALUE CACT OUTCOME

#### 3.1 The $\delta$ -Robustness of CACT outcomes

Ideally, we would like a CACT outcome that is 'robust' to 'large' perturbations in the CACT prior distribution. This would lead to an increased confidence in the CACT outcomes. We formalize the the concepts of *large* and *robust* by defining  $\epsilon$ -perturbations and  $\delta$ -robustness.

The sensitivity of the CACT outcome to a base prior distribution can be determined by 'perturbing' the prior distribution by an  $\epsilon$  amount and evaluating the effect on the outcome.

For parametric distributions,  $\pi_\gamma$ , this is accomplished by perturbing the parameter  $\gamma$ :

DEFINITION 10 ( $\epsilon$ -PERTURBATION OF  $\pi_{\gamma_o}$ ). Let  $\pi_{\gamma_o}$  be the base distribution with base parameter  $\gamma_o$  and  $\pi_{\gamma(\epsilon)}$  be the perturbed distribution, with parameter set to  $\gamma(\epsilon)$ . Given a distance measure between two probability distributions  $D(P, Q)$ , an  $\epsilon$ -perturbation of  $\pi_{\gamma_o}$  is a set of parameters  $\gamma(\epsilon)$  such that,

$$D(\pi_{\gamma_o}, \pi_{\gamma(\epsilon)}) = \epsilon \quad (17)$$

From this, we define the robustness of a CACT outcome:

DEFINITION 11 ( $\delta$ -ROBUSTNESS OF A CACT OUTCOME). Given a base prior distribution  $\pi_{\gamma_o}(\cdot)$  and the CACT outcome for a parameter of interest  $\theta$ ,  $H(\cdot, \pi(\theta; \gamma_o))$ , the CACT outcome has a robustness of  $\delta$  if the following condition holds:

$$\begin{aligned} \forall \gamma \in \{\gamma(\epsilon) : \epsilon \in [0, \delta]\}, \\ H(\cdot, \pi(\theta; \gamma_o)) = H(\cdot, \pi(\theta; \gamma)) \end{aligned} \quad (18)$$

We now formally define the estimation problem as follows:

PROBLEM (ESTIMATION OF THE  $\delta$ -ROBUSTNESS VALUE OF A CACT OUTCOME). Given a base prior distribution  $\pi_{\gamma_o}(\cdot)$  and function for the CACT outcome  $H(\cdot, \pi)$ , determine,

$$\delta = \operatorname{argmin}_\epsilon H(\cdot, \pi(\theta; \gamma_o)) \neq H(\cdot, \pi(\theta; \gamma(\epsilon))) \quad (19)$$

Here, the range of  $\epsilon$  depends on the distance measure used. The resulting  $\delta$  is defined as the  $\delta$ -robustness value.

Intuitively, starting from a base prior distribution, the  $\delta$ -robustness value is the minimal perturbation of magnitude  $\delta$  required to detect a change in the CACT outcome. Fig. 3 depicts how the CACT outcome with a base prior distribution is initially 1. As the magnitude of perturbation ( $\epsilon$ ) increases, the outcome remains constant until  $\delta_o$  is reached. At  $\delta_o$ , the outcome changes from the initial positive outcome to 0. In this example, the  $\delta$ -robustness for the CACT outcome is  $\delta_o$ . Our procedure for  $\delta$ -robustness estimation is presented in the next section.

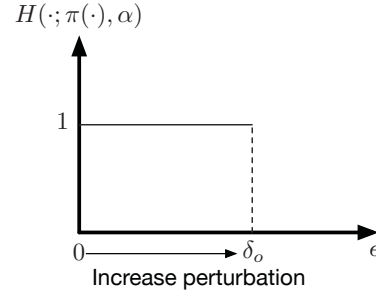


Figure 3:  $\delta$ -robustness of outcome  $H(\phi; \pi(\cdot), \alpha)$ . Perturbations  $\epsilon$  of  $\pi_{\gamma_o}$  are increased until a magnitude of  $\delta_o$ , at which the outcome differs from the initial outcome.

#### 3.2 Procedure for estimation of $\delta$ -robustness

Depending on the choice of distance measure, as well as the form of the base prior distribution, the  $\delta$ -robustness value for a CACT outcome will vary. In this work, we use the Hellinger distance [20] between two discrete distributions, which is defined as follows:

DEFINITION 12 (HELLINGER DISTANCE BETWEEN TWO DISCRETE DISTRIBUTIONS). For two discrete distributions  $\pi_{\gamma_o} = (\gamma_{o1} \dots \gamma_{ok})$  and  $\pi_\gamma = (\gamma_1 \dots \gamma_k)$ , the Hellinger distance is defined as,

$$D_H(\pi_\gamma, \pi_{\gamma_o}) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{\gamma_i} - \sqrt{\gamma_{oi}})^2} \quad (20)$$

In this work, we estimate the  $\delta$ -robustness with respect to the base distribution (e.g. CACT prior) using the Hellinger distance. One advantage to using the Hellinger distance include is that it is a symmetric measure of distance, unlike the Kullback-Liebler divergence. Other advantages are described in [22]. The functional form of the distance also enables convenient numerical computation of distributions which are at  $\epsilon$  distance (see Appendix A for details).

##### Procedure

- (1) Evaluate the CACT outcome with the base distribution  $\pi_{\gamma_o}$
- (2) Starting from a small  $\epsilon$  and gradually increasing the perturbation over the range of  $\epsilon \in (0, 1]$ , for each  $\epsilon$ , sample multiple perturbed distributions  $\pi_{\gamma(\epsilon)}$  and evaluate the CACT outcome.
- (3) Continue until the CACT outcome differs from the initial outcome.

There exist situations where the assumption of the CACT does not affect the outcome, trivially. In this case, the robustness would be infinite, which is ambiguous in meaning. For this reason, we define the weighted  $\delta$ -robustness:

DEFINITION 13 (WEIGHTED  $\delta$ -ROBUSTNESS).

$$\text{Weighted } \delta\text{-robustness} = \begin{cases} \delta W(\delta), & \text{if } W(\delta) > 0 \\ \text{N.A.}, & \text{otherwise} \end{cases} \quad (21)$$

The weight  $W(\delta)$  can be defined such if the weight is zero, then that means the assumption has no impact on the outcome.

In [22], the  $\delta$ -local circular sensitivity  $S_{\gamma_o}^c(\delta)$  is defined as:

$$S_{\gamma_o}^c(\delta) = \left\{ \frac{D(\pi_\gamma(\theta | y), \pi_{\gamma_o}(\theta | y))}{\delta}, \text{ for } \gamma \in G_{\gamma_o}(\delta) \right\} \quad (22)$$

A sensitivity value of 0 indeed indicates that the base prior distribution has no influence over the posterior distribution.

In this paper, we used the weighting function:

$$W(\delta) = E[S_{\gamma_o}^c(\delta)] \quad (23)$$

Therefore, the robustness value is weighted by the average of sensitivity at a perturbation of  $\delta$ . The weighted  $\delta$ -robustness is zero only if the average weight is zero, which can only happen if perturbations in the assumption have no effect on the outcome of the CACT. This resolves the ambiguity for cases of infinite robustness.

#### 4 CASE STUDY: CACT FOR RIGHT (CACT-RIGHT)

In this section, we apply the CACT framework to the Rhythm ID Goes Head-to-head Trial (RIGHT) [4] or Computer-Aided Clinical Trial for RIGHT (CACT-RIGHT).

RIGHT was a large trial with 2187 patients lasting from 2005-2010 and sought to compare the VT/SVT arrhythmia discrimination algorithms used by two ICD models, one from Vitality 2 (V2) and the other from Medtronic (MDT), with regards to the time-to-first inappropriate therapy. At the conclusion of the CT, the effect direction and size for the performance was *opposite* of what was hypothesized, with V2 ICDs having a 34% increase in the risk of inappropriate therapy compared to MDT ICDs.

For CACT-RIGHT, we generate a virtual cohort and apply the signals to measure the inappropriate therapy rate for each device.

**DEFINITION 14 (PARAMETER OF INTEREST FOR CACT-RIGHT).** For each device  $d$ , the inappropriate therapy rate for a cohort of size  $N_o$  is  $\varphi_d$ , where,  $d \in \{0: \text{Vitality II}, 1: \text{Medtronic Devices}\}$ . The parameter of interest for CACT-RIGHT is the difference in inappropriate therapy rate is defined as,

$$\theta = \varphi_0 - \varphi_1 \quad (24)$$

In this paper, we focus on the difference in inappropriate therapy rate, unlike the original time-to-inappropriate therapy. This difference was addressed in [17], where under certain assumptions, the time-to-inappropriate therapy could be approximated and the CACT outcome remained consistent with RIGHT.

##### 4.1 Pre-clinical simulation for CACT-RIGHT

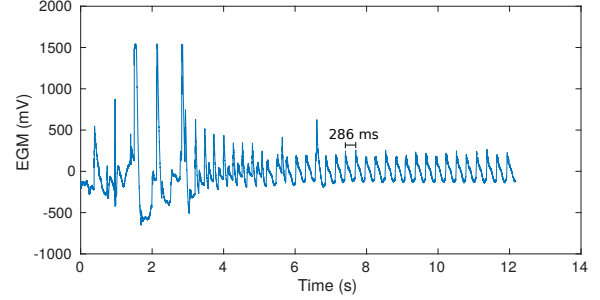
**Physiological cohort generation.** The physiological model consists of a timing model and a morphology model. The details of the physiological model and the signals that are generated can be found in [17][18][19]. Here, we model the physiological cohort  $x_o$  of size  $N_o$ . The cohort consists of  $K = 8$  different types of electrogram (EGM) episodes listed in Table 1. For each type  $i$ , the  $j$ th signal of type  $i$  is I.I.D. such that,

$$X_{oi}^{(j)} \sim f_{X_{oi}}(x_{oi} | \eta) \quad (25)$$

where  $\eta$  is the vector of parameters of the heart model to generate condition  $i$ . The model is then simulated to obtain a time-series of the physiological signal. Fig. 4 shows such a signal with a Ventricular Fibrillation (VF) of cycle length 286(ms).

**CACT-RIGHT prior distribution** A single cohort  $x_o$  of size  $N_o$  is generated as a composition of the different types according to  $\psi = (\psi_1, \psi_2, \dots, \psi_K)$ , where  $\psi_i$  is the number of episodes of type  $i$  in the cohort.  $\psi$  is a multivariate random variable that is governed by a multinomial distribution defined as,

$$\Psi \sim \pi(\psi | \gamma_o) = \text{Mult}(N_o, \gamma_o) \quad (26)$$



**Figure 4: Example of a generated physiological signal of type VF with VCL = 286(ms). The signals in the cohort are indexed by the type of condition and the average VCL for a condition.**

where  $\gamma_o = (\gamma_{o1}, \gamma_{o2}, \dots, \gamma_{oK})$  is a discrete probability distribution representing the proportions of each type in the physiological cohort. Setting  $\gamma_{oi} = 1/k$ , the CACT-RIGHT prior distribution  $\pi(\psi | \gamma_o)$  assumes a uniform composition of the  $K$  types of arrhythmia in the cohort.

Together with the CACT-RIGHT prior distribution, the distribution of the cohort  $f_{X_o}(x_o | \psi; \gamma_o)$  becomes:

$$f_{X_o}(x_o | \psi; \gamma_o) = \prod_{i=1}^K \prod_{j=1}^{\psi_i} f_{X_{oi}}(x_{oi} | \eta) \pi_{\gamma_o}(\psi) \quad (27)$$

The marginal distribution of physiological cohorts is (5).

**Device model and virtual cohort generation.** The device model for the two different devices is implemented according to the description of the algorithms in [6][7]. Each signal in the physiological cohort  $x_o$  is applied to the device model  $d$  to generate a cohort of device outputs,  $y_{od}$ . We assume that the device model is a deterministic function for a particular instance of the cohort  $x_o$ . This allows for the output  $y_{od}$  for a cohort instance to be modeled as a binomial random variable, where  $y_{od}$  is the number of inappropriate therapies in a cohort of size  $N_o$ :

$$Y_{od} \sim f_{Y_{od}}(y_{od} | x_o, \varphi_i; \gamma_o) = \text{Bin}(N_o, \varphi_d) = f_{Y_{od}}(y_{od} | \varphi_d; \gamma_o) \quad (28)$$

where  $d \in \{0, 1\}$  (refer to Def. 14).

We define the likelihood function  $L(\varphi_d | y_{od}) = f_{Y_{od}}(y_{od} | \varphi_i; \gamma_o)$  and assuming a non-informative Beta prior,  $\pi_o(\varphi_d) = \text{Beta}(1, 1)$ , by conjugacy, the posterior distribution for the inappropriate therapy rate is a Beta distribution such that:

$$f_{\varphi_d}(\varphi_d | y_{od}; \gamma_o) = L(\varphi_d | y_{od}) \pi_o(\varphi_d) = \text{Beta}(1 + y_{od}, N_o - y_{od} + 1) \quad (29)$$

Sampling over the many cohorts of  $y_{od}$ , we obtain the marginal distribution inappropriate therapy rate for device  $d$ :

$$f_{\varphi_d}(\varphi_d; \gamma_o) = \int f_{\varphi_d}(\varphi_d | y_{od}; \gamma_o) dF(y_{od}) \quad (30)$$

(30) is a weighted mixture of Beta distributions, for which the weighting is by the distribution of  $y_{od}$ .

The conditional distribution of the pre-clinical prior distribution of  $\theta$  for the CACT-RIGHT can be defined as,

$$\pi_o(\theta | \varphi_1, \varphi_2; \gamma_o) \quad (31)$$

There is no closed form for this distribution, hence a sampling scheme using ordinary Monte-carlo methods is employed. The samples obtained from the posterior distribution form an empirical distribution, as illustrated in Figure 5, from which we can then evaluate the pre-clinical simulation CACT outcome.

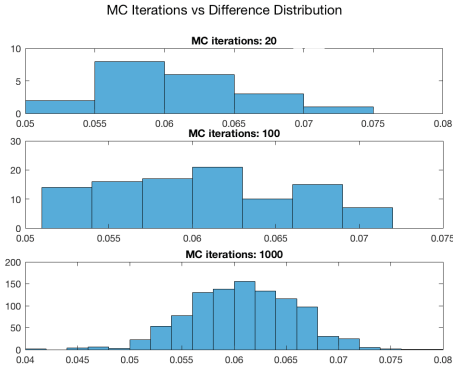


Figure 5: Example of marginalization using Monte Carlo methods to obtain samples from pre-clinical simulation prior.

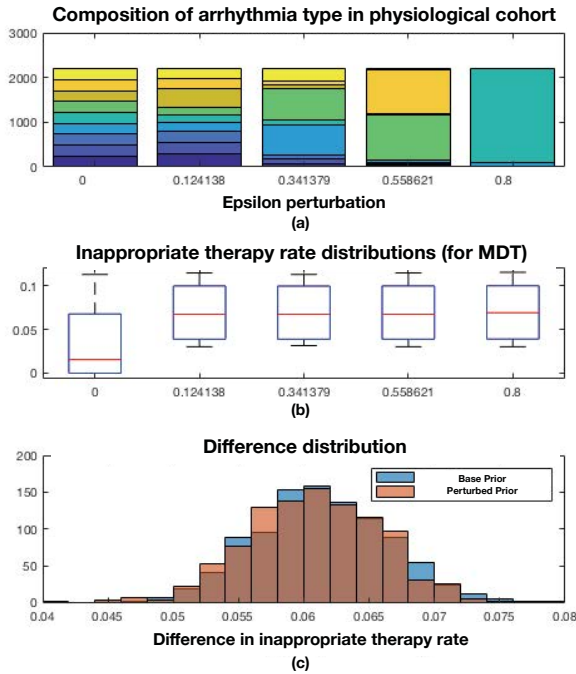


Figure 6: Pre-clinical simulation robustness evaluation. As  $\epsilon$ -perturbations are increased in magnitude, the composition of the cohort from the base uniform distribution is distorted. (b) As a result of perturbation, the pre-clinical distribution shifts. (c) Pre-clinical prior distribution (blue) and perturbed distribution of  $\epsilon = 0.8$ (red). The pre-clinical prior has a large robustness and does not shift. Color in online version.

**CACT-RIGHT pre-clinical simulation outcome.** The pre-clinical outcome of CACT-RIGHT is defined with regards to an assertion  $\theta \in \phi_0 = \{\theta : \theta < 0\}$ , according to the trial. The assertion tests if the inappropriate therapy rate of V2 ICDs is less than MDT ICDs. See (24).

**$\delta$ -Robustness estimation of CACT pre-clinical simulation outcome.** The  $\delta$ -robustness of the pre-clinical simulation outcome with respect to (26) is estimated according to the procedure outlined in Sec. 3.2.

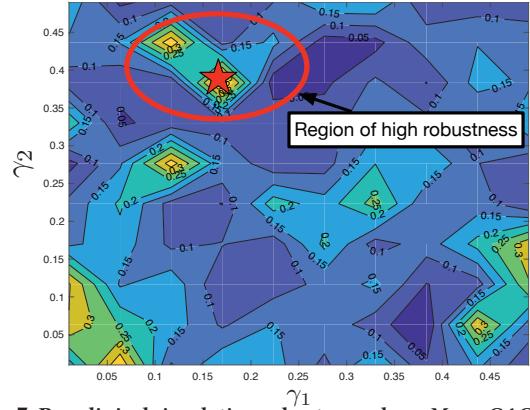


Figure 7: Pre-clinical simulation robustness plane. Many CACT priors are evaluated as the base distribution in order to form the robustness plane. Less emphasis may be placed on recruiting patients corresponding to regions of high robustness (star).

## 4.2 Pre-clinical simulation results

Fig. 6 shows the results of pre-clinical simulation. During pre-clinical simulation, estimates of the inappropriate therapy rate for Vitality II and Medtronic was  $9.99 \pm 0.04\%$  and  $3.88 \pm 0.06\%$ , respectively. Despite the discrepancy in magnitude, the CACT-RIGHT successfully predicted the effect direction as in RIGHT.

Fig. 6(b) illustrates how perturbing the parameters of the CACT prior affects the distribution of simulated endpoints (inappropriate therapy). From the initial base distribution ( $\epsilon = 0$ ), as the magnitude of the  $\epsilon$ -perturbation increases, the difference in inappropriate therapy is more pronounced in the same direction. This indicates the CACT-RIGHT outcome is robust with respect to the assumption of a uniform distribution of conditions.

The resulting pre-clinical prior distribution and the distribution after perturbation is shown in Fig. 6(c). With respect to the CACT-RIGHT prior (26), the  $\delta$ -robustness value of the outcome was 0.778. The maximum robustness in this case is 1, therefore we can conclude that the outcome is relatively large, adding to the confidence of the pre-clinical simulation outcome.

### $\delta$ -Robustness plane of CACT pre-clinical simulation outcome.

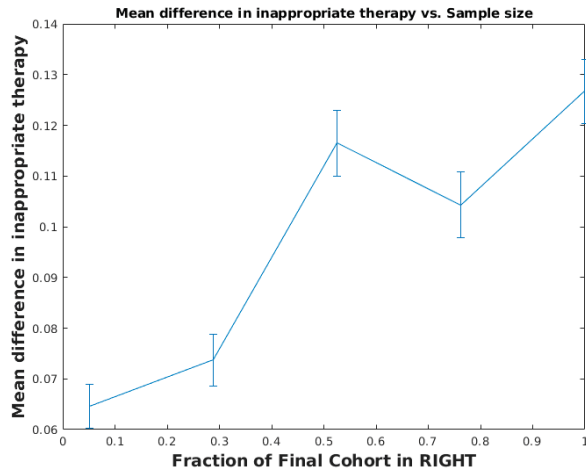
For CACT-RIGHT,  $\gamma_0$  represents the average proportion of the different types of arrhythmia in the physiological cohort and the robustness value for other  $\gamma_0$  can be evaluated. For illustration, we evaluated  $\gamma_0 \in \mathbb{R}^k$ , such that  $\gamma_{0i}$ , for  $i = 1, 2$  was uniformly distributed in the interval  $(0, 0.5)$ , and the remaining probability mass was distributed evenly, such that  $\gamma_{0j} = (1 - (\gamma_{01} + \gamma_{02}))/k, \forall j = 3, \dots, k$ . The robustness value was computed for each instance of  $\gamma_0$  and plotted to form the *pre-clinical simulation robustness plane*. Fig. 7 shows the weighted  $\delta$ -robustness plane.

A region on the robustness plane can be interpreted as a subset of the population that exhibits similar proportions of arrhythmias. The red star in the figure indicates a region of higher robustness. Although the robustness value is related to the distribution space, a point with a higher robustness value means there is a relatively larger region in the parameter space of  $\gamma_0$  where the simulation outcome will not change. From a trial planning perspective, this implies that less effort could be put into recruiting patients that correspond to this region, though not excluded entirely, as there is more confidence in the simulation outcome.



n episodes (% of total events)		
Adjudicated Rhythm	Vitality II	Medtronic
Ventricular tachycardia	23(1.1)	90(4.6)
Ventricular fibrillation	705(34.9)	994(51.0)
Sinus tachycardia	59(2.9)	220(11.3)
Atrial fibrillation	431(21.3)	101(5.2)
Atrial flutter	66(3.3)	19(1.0)
Atrial tachycardia	20(1.0)	100 (5.1)
Other SVT	178(8.8)	325(16.7)
Sinus rhythm with PVC	18(0.9)	1(0.1)

**Table 1: Summary of RIGHT[13] results. Inappropriate therapy per condition for each of the devices.**



**Figure 8: Change in mean difference distribution as more patients enroll. Initially, a small difference in inappropriate therapy rate becomes more pronounced as the cohort size reaches the size of the cohort at the conclusion of RIGHT.**

### 4.3 Interim trial simulation and final results

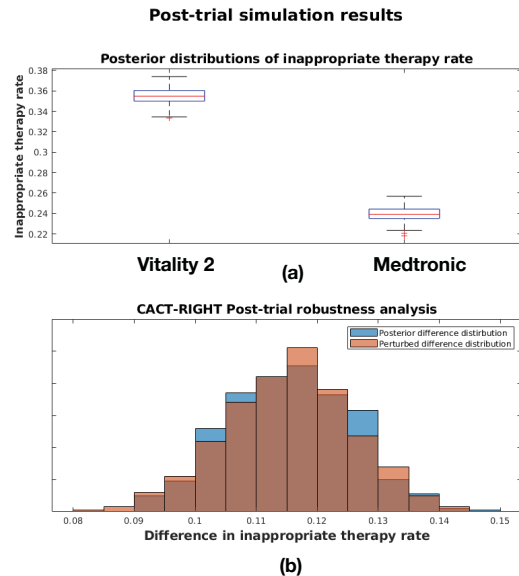
**Simulation of real patient cohort endpoints** Recruitment of the real cohort was emulated by utilizing the results reported at the conclusion of RIGHT [13], summarized in Table 1. By assuming the inappropriate therapy rate will remain constant throughout the trial, we sample from a binomial distribution  $Bin(N(t), \varphi_d)$ , where  $N(t)$  is the number of inappropriate therapies at a time during the trial and  $\varphi_d$  is the final inappropriate therapy rate reported for device  $d$ .

The discount function based on the Weibull cumulative distribution  $F(p | \kappa, \lambda)$  is defined:

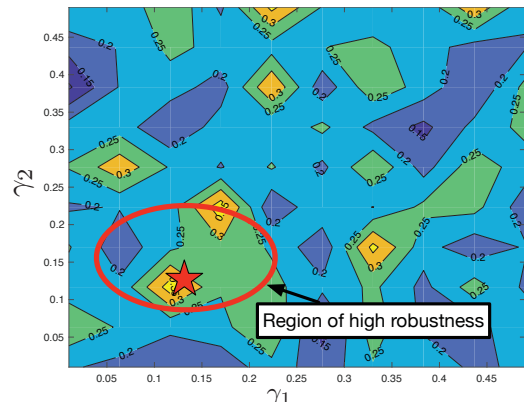
$$n_0 = n_{max} * F(p | \kappa, \lambda) \tag{32}$$

where  $F(p | \kappa, \lambda) = 1 - e^{-(\frac{p}{\lambda})^\kappa}$ ,  $p$  is the ‘p-value,’ and  $n_{max}$  is the maximum effective sample size for the virtual cohort (refer to Def. 6). A discussion about selecting  $\lambda$  and  $\kappa$  is in Sec. 5.1.

**Interim and final results of CACT-RIGHT.** Fig. 8 shows the average difference in appropriate therapy rate where the number of episodes increases up to the final number of episodes observed in RIGHT. The figure shows how initially, the mean difference between V2 ICDs and MDT ICDs is less pronounced when the sample size is small, but becomes more pronounced as the cohort size reaches the final cohort size. The final difference  $9.8 \pm 0.2$  % is in line with the actual difference in inappropriate therapy rate  $10.1$  % reported at the conclusion of RIGHT.



**Figure 9: Results of post-trial simulation and robustness evaluation for CACT-RIGHT. (a) Estimated inappropriate therapy rate for each device. The difference is in line with actual RIGHT (b) Post-trial robustness analysis. Similar to Fig. 6, the CACT outcome demonstrates high robustness. Color in online version.**



**Figure 10: Post-trial analysis robustness plane. Subsequent trial may focus on areas of low robustness.**

### 4.4 Post-trial simulation robustness analysis

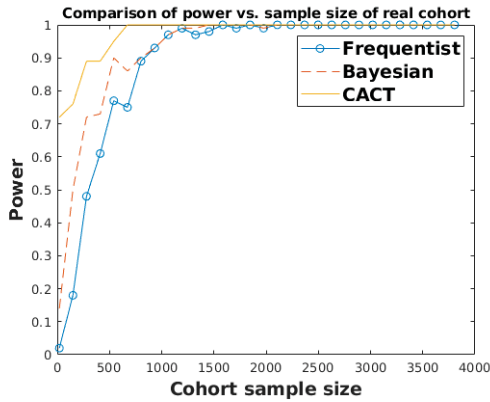
At the conclusion of the trial, information about the CACT prior (i.e. the incidence of the types of arrhythmia) would be available, allowing for a post-trial robustness evaluation to assess how the outcomes may change if a different cohort was observed. We set the parameters for the CACT prior to the values according to Table. 1. The robustness evaluation, Fig. 9(c), shows that the posterior outcome distribution is robust to changes in the CACT prior.

Fig. 10 shows the robustness plane for the post-trial simulation, as derived in Sec. 4.2. The results could be used in two ways:

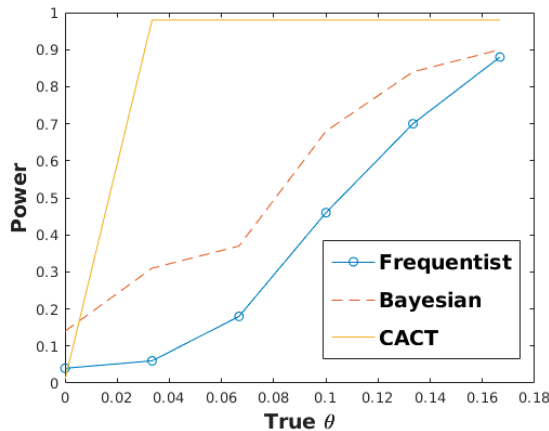
First, when planning a follow-up to the current trial, recruitment efforts and resource allocation can be focused on regions of low robustness.

Second, the robustness plane could be utilized for personalization of treatments. For example, for new patients corresponding to regions of high robustness, the outcome of the CACT could be used to determine whether a patient should receive the device from one manufacturer or another.





**Figure 11:** Comparison of CACT, Bayesian approach with minimal prior and frequentist approach in terms of power vs. sample size of RIGHT. CACT requires a smaller cohort compared to the other methods for the same power. Color in online version.



**Figure 12:** Comparison of CACT, Bayesian approach with minimal prior and frequentist approach in terms of power vs true  $\theta$ . In the positive  $\theta$  direction, the CACT increases in power more rapidly than other methods.

### 4.5 CACT vs other standard approaches

In order to assess the benefits of a CACT, three methods were compared: A frequentist approach comparing only proportions, a standard Bayesian approach using a minimally-informative prior, and the CACT. Cohorts of endpoints were generated repeatedly according to the inappropriate therapy rate reported for RIGHT.

While varying the size of the cohort, the power was estimated by determining the proportion of times each method correctly detected the relative difference in performance between the two devices. Fig. 11 shows how as the sample size increases, the power of the CACT increases much more rapidly compared to the other approaches. This implies that a CACT could achieve the same power as standard approaches using a smaller cohort, which could lead to a significant saving in costs due to the reduction in cohort size.

Fig. 12 shows the results of varying the underlying 'true' difference in inappropriate therapy rate with fixed cohort size of 400. As before, the power curve of the CACT increases rapidly compared to the other methods, further demonstrating the advantages of the CACT.

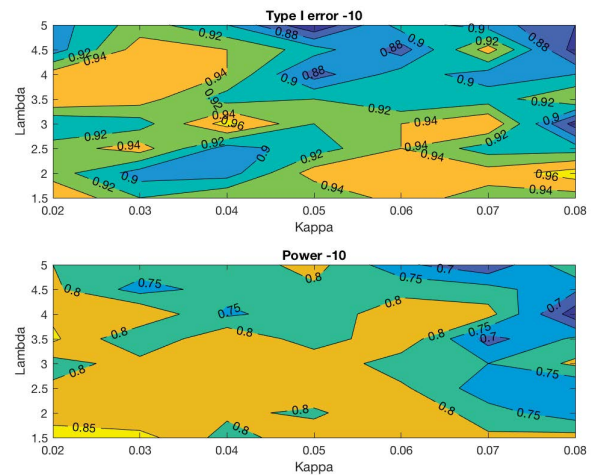
## 5 DISCUSSION

### 5.1 Parameter selection for simulation

In CACT-RIGHT, the parameters for the Weibull discount need to be selected (see Sec. 4.3, (32)). The detailed procedure for evaluating the Type I error and power can be found in [15]. For CACT-RIGHT, a range of 'true' values for the parameter of interest, the difference in inappropriate therapy, is assumed based on information from clinical trials. For each of the true values, a cohort of endpoints was generated according to the that value. The values for  $\kappa$  and  $\lambda$  in (32) were set to a value within the range of  $0.02 < \kappa < 1$  and  $0.5 < \lambda < 1$ , similar to [15]. For each value of  $\kappa$  and  $\lambda$ , numerous 'mock' trials were run and the type I errors and type II errors were tabulated. Note, here, we consider a type I error as when the true value was within the range of the null hypothesis, but the CACT concluded the opposite. Type II errors are defined similarly.

Fig. 13 shows an example of the Type I error and power over the range of  $\kappa$  and  $\lambda$  when a difference of 10% in the inappropriate therapy rate in favor of Vitality II is assumed. For CACT-RIGHT  $\kappa = 0.05$  and  $\lambda = 3$  was chosen for a power and type I error of 80% and 0.05%, respectively.

A similar procedure is used to determine the maximum effective sample size of the virtual cohort,  $n_{0,max}$ .



**Figure 13:** Results of parameter search for the Weibull discount function.

### 5.2 Related work

Clinical trial simulation and modeling has been used extensively in determining components of clinical trials [2]. These include testing for statistical model selection, determining sample size, and also predicting the outcome of clinical trials [12]. Additionally, work has been presented regarding the incorporation of prior information into the design of a prospective trial in the form of historical trials [16][23]. Some notable examples of utilizing physiological models as virtual endpoints include [14][8]. This work builds upon results in Bayesian sensitivity analysis and a review of which can be found in [5][22].

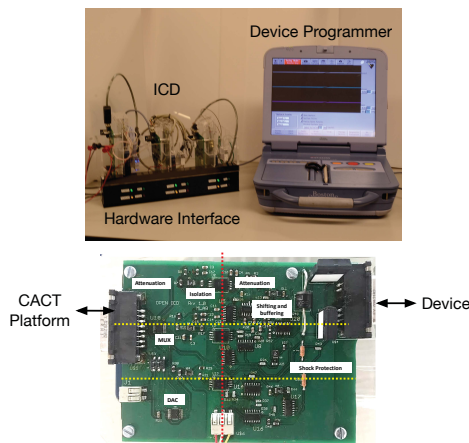
### 5.3 Limitations and extensions

The current method for evaluating robustness is limited in terms of the scalability. For more complicated models and CACT prior distributions, searching over the entire space of  $\epsilon$ -perturbations may not be feasible. Moreover, there are limitations in evaluating the

robustness with regard to only one assumption at a time. However, approaches to analyzing multiple parameters at once exist and may be applicable to  $\delta$ -robustness evaluation[3]. Finally, the procedure for estimation of  $\delta$ -robustness results in an upper bound for the actual  $\delta$ -robustness value, as a sampling scheme is used for the computation. This could be improved using more sophisticated sampling methods.

### 5.4 Hardware interface for CACT

In order to facilitate automated execution of a CACT with the actual physical device, a hardware interface was developed as shown in Fig. 14. The interface allows inputs physiological signals to the device and records the response automatically. Scenarios combining several types of rhythms can be programmed and applied to the device.



**Figure 14: Hardware interface for conducting CACT for ICDs. Episodes from physiological cohort can be applied to the physical device. The interface can also be used for device model validation**

## 6 CONCLUSION

We presented a method to quantify the uncertainty in the outcome of Computer-Aided Clinical Trials (CACTs) due to simulation assumptions in the form of a CACT prior. To this end, we formulated a CACT within a Bayesian statistical framework and defined a notion of perturbation in the assumptions, which we mapped to a formal definition of robustness ( $\delta$ -robustness) of the trial outcome with respect to the CACT prior. We also presented a procedure to estimate this quantity. Through a comparative study of an actual clinical trial of two implantable cardiac devices, we demonstrated how this notion can be used in the planning and design of a clinical trial. We compared our CACT based approach to two other standard methods, demonstrating the benefits. The framework provides a quantitative and explicit pathway to integrate modeling and simulation as regulatory-grade evidence in evaluating the safety and efficacy of medical devices at reduced cost and effort. Future work will include, advancing the methodology to account for multiple priors and applying the CACT to other closed-loop devices, such as the target controlled infusion systems and fluid resuscitation systems.

## ACKNOWLEDGMENTS

The authors would like to thank Tarek Haddad for the insightful discussions about the work and Alena Rodionova for comments on earlier versions of the manuscript. This work was funded in

part by NSF:1446664 CPS: Frontier: Collaborative Research: Compositional, Approximate, and Quantitative Reasoning for Medical Cyber-Physical Systems

## REFERENCES

- [1] H. Abbas, Z. Jiang, K. J. Jang, M. Beccani, J. Liangy, and R. Mangharam. 2016. High-level modeling for computer-aided clinical trials of medical devices. In *Proc. IEEE Int. High Level Design Validation and Test Workshop (HLDVT)*. 85–92. <https://doi.org/10.1109/HLDVT.2016.7748260>
- [2] Ismail Abbas. 2016. Modeling and Simulation in Clinical Trials Modeling and Simulation in Clinical Trials. 6, April (2016), 2016–2018. <https://doi.org/10.22360/SpringSim.2016.MSM.002>
- [3] Sensitivity Analysis, in Bayesian, Networks: From, Single to Multiple, and Parameters. 2004. *UAI 2004 CHAN & DARWICHE*. Technical Report is.
- [4] Ronald D Berger et al. 2006. The Rhythm ID Going Head to Head Trial (RIGHT): Design of a Randomized Trial Comparing Competitive Rhythm Discrimination Algorithms in Implantable Cardioverter Defibrillators. *Journal of Cardiovascular Electrophysiology* 17, 7 (2006), 749–753. <https://doi.org/10.1111/j.1540-8167.2006.00463.x>
- [5] Emanuele Borgonovo and Elmar Plischke. 2016. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research* 248, 3 (2016), 869–887. <https://doi.org/10.1016/j.ejor.2015.06.032>
- [6] Boston Scientific Corporation. 2007. The Compass - Technical Guide to Boston Scientific Cardiac Rhythm Management Products. *Device Documentation* (2007).
- [7] C. D. Swerdlow and others . 2002. Discrimination of Ventricular Tachycardia from Supraventricular Tachycardia by a Downloaded Wavelet Transform Morphology Algorithm: A Paradigm for Development of Implantable Cardioverter Defibrillator Detection Algorithms. *J. Cardiovascular Electrophysiology* 13 (2002).
- [8] C. Toffanin, M. Messori, F. Di Palma, G. De Nicolao, C. Cobelli, L. Magni. 2014. Artificial Pancreas: Model Predictive Control Design from Clinical Experience. *J Diabetes Sci Technol* 7 (2014), 1470–1483.
- [9] Gregory Campbell. 2017. Bayesian methods in clinical trials with applications to medical devices. *Communications for Statistical Applications and Methods* 24, 6 (2017), 561–581.
- [10] C. Cobelli, E. Renard, and B. Kovatchev. 2011. Artificial pancreas: past, present, future. *Diabetes* (2011).
- [11] Food, Drug Administration, et al. 2011. Guidance for industry and FDA staff: Guidance for the use of bayesian statistics in medical device clinical trials. (2011).
- [12] A Giovagnoli and M Zagoraio. 2012. Simulation of clinical trials: a review with emphasis on the design issues. *Statistica* 1 (2012), 63–80.
- [13] Michael R. Gold, Saleem Ahmad, Kevin Browne, Kellie Chase Berg, Lisa Thackeray, and Ronald D. Berger. 2012. Prospective comparison of discrimination algorithms to prevent inappropriate ICD therapy: Primary results of the Rhythm ID Going Head to Head Trial . *Heart Rhythm* 9, 3 (2012), 370 – 377. <https://doi.org/10.1016/j.hrthm.2011.10.004>
- [14] Tarek Haddad, Adam Himes, and Michael Campbell. 2014. Fracture prediction of cardiac lead medical devices using Bayesian networks. *Reliability engineering & system safety* 123 (2014), 145–157.
- [15] Tarek Haddad, Adam Himes, Laura Thompson, Telba Irony, and Rajesh Nair. 2017. Incorporation of stochastic engineering models as prior information in Bayesian medical device trials. *Journal of Biopharmaceutical Statistics* 00, 00 (2017), 1–15. <https://doi.org/10.1080/10543406.2017.1300907>
- [16] Joseph G Ibrahim, Ming-Hui Chen, Yeongjin Gwon, and Fang Chen. 2015. The power prior: theory and applications. *Statistics in medicine* 34, 28 (2015), 3724–3749.
- [17] Kuk Jin Jang, James Weimer, Houssam Abbas, Zhihao Jiang, Jackson Liang, Sanjay Dixit, and Rahul Mangharam. 2018. Computer Aided Clinical Trials for Implantable Cardiac Devices. *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2018).
- [18] Zhihao Jiang, Houssam Abbas, Kuk Jin Jang, Marco Beccani, Jackson Liang, Sanjay Dixit, and Rahul Mangharam. 2016. In-silico pre-clinical trials for implantable cardioverter defibrillators. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the IEEE*, 169–172.
- [19] Zhihao Jiang, Miroslav Pajic, and Rahul Mangharam. 2012. Cyber-Physical Modeling of Implantable Cardiac Medical Devices. *Proc. IEEE* 100, 1 (Jan. 2012), 122 –137.
- [20] L. Le Cam. 1986. *Asymptotic Methods in Statistical Decision Theory*.
- [21] W.H Maisel, W. Sweeney, and Stevenson. 2001. Recalls and safety alerts involving pacemaekrs and implantable cardioverer-defibrillator generators. *JAMA* (2001).
- [22] Magorzata Roos, Thiago G. Martins, Leonhard Held, and Hvard Rue. 2015. Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis* 10, 2 (2015), 321–349. <https://doi.org/10.1214/14-BA909> arXiv:arXiv:1312.4797v1
- [23] Kert Viele, Scott Berry, Beat Neuenschwander, Billy Amzal, Fang Chen, Nathan Enas, Brian Hobbs, Joseph G Ibrahim, Nelson Kinnersley, Stacy Lindborg, et al. 2014. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical statistics* 13, 1 (2014), 41–54.

## A NUMERICAL COMPUTATION OF $\epsilon$ PERTURBATIONS FOR DISCRETE DISTRIBUTIONS

We solve the following feasibility problem to find the parameters  $\alpha \in \mathbb{R}^n$  of a distribution that is  $\epsilon$  Hellinger distance away from a given distribution with parameters  $\beta \in \mathbb{R}^n$ . Let  $\gamma \in \mathbb{R}^n$  be such that  $\gamma_i = \sqrt{\alpha_i} \forall i = 1, \dots, n$ . The (non-convex) feasibility problem then becomes:

minimize  $\gamma$ , subject to,

$$\mathbf{1} \succcurlyeq \gamma \succcurlyeq \mathbf{0} \quad (33a)$$

$$\gamma^T \gamma = 1 \quad (33b)$$

$$\gamma^T \gamma - \begin{bmatrix} 2\sqrt{\beta_1} & 2\sqrt{\beta_2} & \dots & 2\sqrt{\beta_n} \end{bmatrix} \gamma + \mathbf{1}^T \beta - 2\epsilon^2 = 0 \quad (33c)$$

Here,  $\mathbf{1} \in \mathbb{R}^n$  (and  $\mathbf{0}$ ) is a vector with all elements being 1 (and 0). Eqs. 33a and 33b ensure that  $\alpha_i = \gamma_i^2 \forall i$  form the parameters of a probability distribution, i.e. are bounded between 0 and 1, and sum up to 1 (respectively). Eq. 33c enforces the condition that the square of the Hellinger distance between the distributions given by  $\alpha$  and  $\beta$  equals  $\epsilon^2$ . This follows directly from Def. 12.