

Model-Based Clinical Trials for Medical Devices

Houssam Abbas¹, Zhihao Jiang¹, Kuk Jin Jang¹, Marco Beccani¹
Jackson Liang², Sanjay Dixit³, and Rahul Mangharam¹

¹Dept. of Electrical and Systems Eng.

University of Pennsylvania

{habbas, zhihaoj, jangkj, beccani, rahulm}@seas.upenn.edu

²Cardiovascular Division, ³Cardiac Electrophysiology

Hospital of the University of Pennsylvania

{jackson.liang, sanjay.dixit}@uphs.upenn.edu

ABSTRACT

Regulatory authorities require that the safety and efficacy of a new high-risk medical device be proven in a Clinical Trial (CT), in which the effects of the device on a group of patients are compared to the effects of the current standard of care. Phase III trials can run for several years, cost millions of dollars, and pose an inherent risk to the patients by exposing them to an unproven device. In this paper, we demonstrate how to use closed-loop models of a set of devices and a large patient cohort to improve the planning and execution of a CT so as to increase the chances of a successful trial. We illustrate our approach by applying it to a real CT that compares the algorithms within implantable cardioverter defibrillators (ICDs) for the treatment of potentially fatal cardiac arrhythmias. The CT failed to prove its premise that the Boston Scientific algorithm resulted in fewer inappropriate shocks than the Medtronic algorithm. We begin by modeling the heart, processing 100's of real patients' signals, mapping the timing and morphology components of the signals to the heart model. This is followed by generating a population of 10,000+ synthetic heart models and implementing the device algorithms for both Boston Scientific and Medtronic devices. We perform conformance testing to validate our device models against real ICDs. Now, using the closed-loop of the device models and synthetic patient population we conduct multiple trials for the comparative study of inappropriate therapy across a range of heart conditions. The results of our model-based clinical trials (MBCT) indicate that Boston Scientific's algorithm resulted in a higher risk of inappropriate therapy than Medtronic's. We further demonstrated that the result continues to hold if we vary the characteristics of the synthetic population and device parameters - thus indicating that the CT was unlikely to prove the desired effect. While MBCTs do not seek to replace a CT, they provide statistically significant insight into the factors which affect the outcome and provide early insight prior to a CT at a fraction of the cost and duration and without the ethical issues. This effort is a early step towards using computer modeling as regulatory-grade evidence for medical device certification.

1. INTRODUCTION

The lives of millions of patients around the world depend on medical devices. In the domain of cardiac devices, for example, 10,000 people in the U.S. receive an Implantable Cardioverter Defibrillator (a heart rhythm adjustment device) every month [1]. Clinical trials have presented evi-

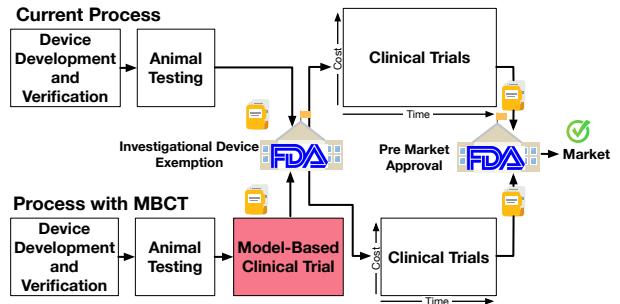


Figure 1: Bringing a device to market. Clinical trials are the last step before a new device's market approval. Model-based clinical trials will provide insight during planning and execution of clinical trials, leading to reduction in costs and increasing the chance of a successful trial.

dence that patients implanted with ICDs have a mortality rate reduced by up to 31% [2]. Unfortunately, ICDs suffer from a high rate of *inappropriate therapy*, which takes the form of unnecessary electric shocks or pulse sequences delivered to the heart. Inappropriate therapy increases patient stress, reduces their quality of life, and is linked to increased morbidity [3]. Depending on the particular ICD and its settings, the rates of inappropriate therapy range from 46% to 62% of all delivered therapy episodes [4]. The closed loop formed by the organ (heart) and device (ICD) is an example of a life-critical Cyber-Physical System (CPS): the device's software is the cyber component, and the physiological phenomenon (e.g., cardiac rate and electrical activity) is the physical component.

After the verification and testing effort is completed, regulatory agencies like the F.D.A. require that the safety and efficacy of new devices be demonstrated in a *Clinical Trial (CT)* (Fig. 1). In a trial, a group of patients that are treated with the new device (this is the 'intervention group') are compared to a group of patients who are treated with the current standard of care (e.g., a different device currently on the market; this is the 'control group'). The objective is to see whether the different devices result in significantly different effects on the patients. Clinical trials are major endeavors, involving physicians, patients, statisticians, clinical centers, companies and regulators, sometimes in several countries. Late-phase trials can run for several years, and cost millions of dollars. For example, a 2002 trial for stents lasted 2 years, enrolled 800 patients and cost \$10 to \$12 million and lasted 24 months [5]. Trials also pose an inherent risk to the patients in the intervention group by exposing them to an unproven device. Thus it is crucial that they be well planned, and rigorously executed.

In reality, any trial runs the risk of errors during its planning and execution stages, which can invalidate the results of the trial. In this paper, we pose and propose an answer to the following question: *how can modeling of CPS assist in the planning and execution of a clinical trial, so as to increase the chances of a successful trial?*

Specifically, in this paper, we demonstrate how computer models can be used for early, affordable and reproducible testing of a clinical trial’s premises and assumptions. Model-based empirical validation of the premises reduces the risk of conducting a trial that fails to demonstrate the desired effect (typically, an improvement of new intervention over the control). We work with the Rhythm ID Going Head to Head Trial (RIGHT) [4], which lasted five years and sought to compare two ICDs for the treatment of potentially fatal tachycardias (abnormally fast heart rhythms). Our contributions are as follows:

- We defined a heart model capable of producing several types of tachycardias, in terms of both timing characteristics and morphology of the electrical signals (the *electrograms (EGMs)*) (Section 4). This model allows us to generate thousands of arrhythmia exemplars which serve as our virtual trial patients.
- For modeling purposes, we annotated segments of over 100 EGM records of real patients to extract the tachycardias that they suffered from at time of recording. These electrograms are essential to simulate our models.
- Using the openly available literature, we implemented the tachycardia detection algorithms of two major ICD manufacturers: Boston Scientific and Medtronic. (See Section 5).
- We developed an experimental setup to validate our implementation of the Boston Scientific algorithm against a real ICD (Section 5.3).

The connected heart model and ICD form the CPS under study.

- With the above elements in place, we generated a synthetic cohort, consisting of 11,000+ heart models displaying a wide range of tachycardias. These models are then connected to the ICD implementations and arrhythmia detection results are recorded. This allows us to estimate their relative efficacy on a condition-by-condition basis, which can be used for estimating the effect size (Section 6.)

We call our approach Model-Based Clinical Trial (MBCT), or MBCT. An MBCT is a trial whose subjects are computer models of the (heart, device) closed-loop system. By generating large cohorts, we can answer several questions contributing to the planning and execution of a clinical trial.

Modeling as regulatory grade evidence.

Most medical models today are aimed at either better understanding the phenomenon under study [6] or at device debugging and verification [7]. There is only one case in which a computer model has been used to intervene in the regulatory process of medical devices, namely the T1 Diabetes Model (T1DM) of UVA/PADOVA [8]. T1DM models glucose kinetics in hypoglycemia, and has been accepted by the FDA as a substitute for animal trials. The T1DM has a fixed virtual cohort with 300 patients. Its objective is to test

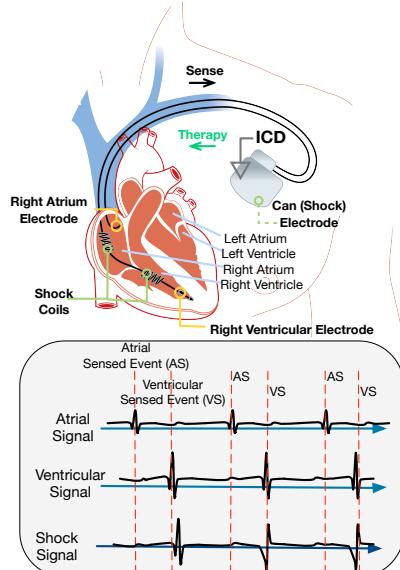


Figure 2: ICD connected to the heart. The atrial, ventricular, and shock electrogram signals are measured by the device, which uses them to diagnose the current state of the heart and determine whether therapy is required.

the efficacy of new glucose control algorithms by simulating them on the virtual cohort. While our models (and indeed, most models) can be used in this way, our objective here is to target specific clinical trials steps and improve how they are conducted. This dictates the experimental setup and the cohort generation considerations.

The Avicenna consortium [9] lays out a vision for ‘In-Silico Clinical Trials’ similar to our approach. However, the emphasis in Avicenna is on individualized patient models, as they propose to customize the model to each patient enrolled in a trial. In the present work, we propose a usage of MBCT *prior* to recruitment. Thus our models need not be fitted to a given patient’s data, which might be impossible, invasive, or burdensome for the conduct of the trial.

2. CLINICAL TRIALS AND RIGHT

At the clinical trial stage (Fig. 1), the objective is no longer to find bugs in the device: it is, rather, to evaluate the safety and efficacy of the validated device on humans. Randomized Controlled Trials (RCTs) are the gold standard for evaluating the safety and efficacy of a new medical device [10]. They constitute the only time prior to market where the effects of the device on humans are actually observed, and are legally mandated for new high-risk medical devices like implantable cardioverter defibrillators (ICDs). The planning and execution of an RCT requires carefully navigating a number of technical, logistical and ethical issues to obtain reliable and statistically significant results.

Because of the very high cost of RCTs in terms of money, time, and the risk of harm they present to enrolled patients, our focus in this paper is on the use of CPS models, formalized in an MBCT, *to validate the assumptions made by the investigators and thus increase the chances of success of an RCT*. We illustrate our approach by applying it to the Rhythm ID Going Head-to-Head Trial (RIGHT) [4], which we present next.

2.1 The RIGHT trial

We first provide a brief background to better understand

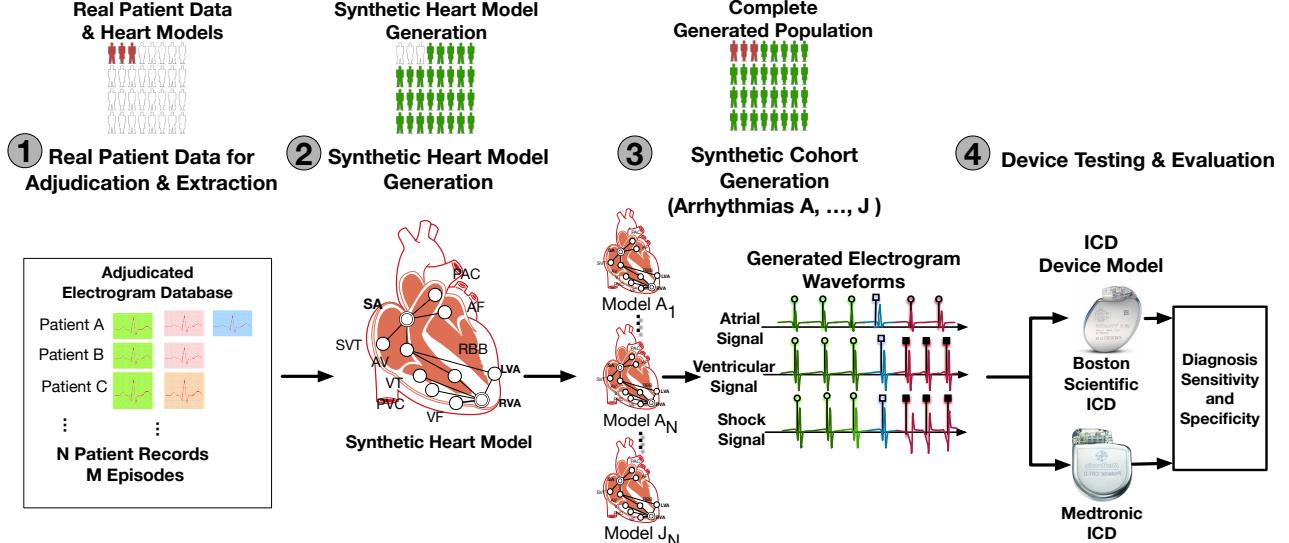


Figure 3: Overview of an MBCT. 1) EGM recordings of real patients are adjudicated to create a table of EGM morphologies of various tachycardias. 2) Subsets of these morphologies are combined with a timing model to create a synthetic heart model. 3) Through variation of the parameters of the model, an entire synthetic cohort is generated and simulated to produce synthetic EGM signals. 4) Various device evaluation experiments can be executed with this synthetic cohort.

RIGHT (see Fig.2). Tachycardias (abnormally elevated heart rates) can be divided into Ventricular Tachycardias (VTs), which originate in the heart's ventricles, and SupraVentricular Tachycardias (SVTs), which originate above the ventricles. A sustained VT is fatal within seconds of onset, while an SVT is typically non-fatal. The therapy applied by the ICD often takes the form of a high-energy electric shock. The shock can be pro-arrhythmic, and was even linked to increased morbidity [3]. Therefore, one of the biggest challenges for ICDs is to guarantee shock delivery during fatal VTs, and simultaneously reduce inappropriate shocks for non-fatal SVTs [11].

RIGHT is a trial that sought to compare the VT/SVT discrimination abilities of two competing detection algorithms from two ICD manufacturers [4]: the detection algorithm found in Boston Scientific's Vitality II ICDs (which uses the Rhythm ID discriminators [12]), and the PR Logic + Wavelet detection algorithm found in a number of Medtronic's ICDs. *Inappropriate therapy* was defined as therapy applied to an arrhythmia other than VT or Ventricular Fibrillation (VF) (VF is a type of VT). RIGHT enrolled 1962 patients and ran for approximately five years. It was fully sponsored by Boston Scientific.

One of the trial's assumptions was that Boston Scientific's algorithm would reduce the risk of inappropriate therapy by 25% over Medtronic's algorithm [13]. The outcome of the trial [4], however, was that patients implanted with Boston Scientific ICDs had a **34% risk increase** of inappropriate therapy as compared to patients implanted with Medtronic ICDs. This result is the opposite of the effect hypothesized by the trial investigators. In this paper, we design an MBCT to verify early and cheaply whether the hypothesized effect holds by comparing the two ICDs on a large *synthetic* cohort.

Organization: In the following sections we describe the building blocks of the MBCT: modeling the heart, processing 100's of real patients' data, mapping the timing and morphology components of the signal to a heart model we developed, generating a population of 10,000+ synthetic heart

models, implementing the device algorithms and conducting multiple trials for the comparative rate of inappropriate therapy, condition-level rates and evaluating the effect of device parameters on discrimination rates.

3. Model-Based Clinical Trial

We demonstrate how a heart model we developed is used to generate a synthetic cohort and test whether the Boston Scientific detection algorithm is indeed better than Medtronic's as was assumed by the RIGHT investigators prior to trial start. A negative answer to this question would cause the trial investigators to revise their assumptions before the costs of a trial had been sunk. Because an MBCT is designed and conducted in support of a given RCT, the details necessarily depend on the RCT we consider. In Fig. 3 we give an overview of the process of the MBCT we conducted in support of RIGHT.

(1) Modeling starts by adjudication of 380 individual episodes from a database of 123 EGM records of real patients. Each such episode provides EGM morphologies that are annotated with the tachycardia that produced them (see Fig. 6 for examples), resulting in the identification of 19 different rhythms.

(2) An automata-based timing model is used to simulate the timing characteristics of various tachycardias. Combined with the annotated EGM morphologies, we can now generate parametrized probabilistic heart models that simulate different tachycardias, and variations on each tachycardia.

(3) A cohort of > 11,000 models is generated by varying the parameters of the heart model. The parameter ranges depend on the tachycardia being simulated.

(4) Every member of the cohort is then simulated to produce EGM signals that are fed to both ICD algorithms. The rates of inappropriate therapy from the two algorithms are analyzed. We repeated this analysis for various distributions of arrhythmias in our cohort. The conclusion was that the Medtronic algorithm had fewer inappropriate therapies applied than Boston Scientific's.

3.1 Advantages of an MBCT

The use of computer models in MBCT gives us significantly more latitude in procedures since we don't have the ethical constraints of a clinical trial. In particular, in an RCT, a patient will typically be on either intervention (new device) or control (currently accepted device or other standard treatment). It is usually not possible or ethical to retrieve the first device from the patient just for the purposes of the trial, and implant them with the other device. In an MBCT on the other hand, the same model can be subjected to both intervention and control, so that perfect comparability between the two groups is assured.

A significant advantage of MBCTs is that we can generate very large cohorts, thus lending statistical strength to the results. We control the variability in the cohort, so it is possible to test the effects of the new device on a particular sub-group, e.g., people with one dominant type of arrhythmia. Moreover, we can test the outcome under varying device parameter values, something which is not feasible in a clinical trial.

3.2 What an MBCT is not

An MBCT does not seek to replace an RCT: the latter provides data on the safety and efficacy of the new device *in the clinical setting*. That is, under conditions that closely resemble the conditions under which the device will be used in real life, in clinics and hospitals around the world. Rather, an MBCT is designed to improve key steps in the planning and execution of an RCT, and to confirm early on assumptions about the effectiveness of the new device.

The difference between MBCT and current Model-Based Design (MBD) is one of goals and emphasis. The goal of an MBCT is to help better plan and conduct a subsequent RCT. It looks at things like statistical significance and performance in a target population. The goal of MBD is to design and debug the device. It looks at things like corner cases and implementation errors. The device (or software) that is evaluated in an MBCT has already been through the verification stage and is considered market-ready (pending trial results and regulatory approval).

4. HEART MODELING

4.1 Basic cardiac electrophysiology

The heart has two upper chambers called the *atria* and two lower chambers called the *ventricles* (see Fig. 2). The synchronized contractions of atria and ventricles assure an adequate supply of oxygenated blood to the rest of the body. This contraction is driven by electrical activity in the heart. A normal pattern of electrical activity is referred to as normal sinus rhythm (NSR). Disturbances of NSR are referred to as *arrhythmias*, and can result in insufficient blood supply and even death of a patient. *VT* is an example of an arrhythmia originating in the ventricles, in which the ventricles beat at a very high rate. If the VT is sustained, or degenerates into VF, it is fatal within seconds. An abnormally fast heart rate that originates in the atria is referred to as a *SVT*. This is a sub-optimal but not fatal condition, and many arrhythmias fall under this heading.

implantable cardioverter defibrillator (ICD) treat VTs and VFs by observing the electrical activity through three channels, as shown in Fig. 2. The measured signals are known as *electrograms*, or EGMs. VTs and SVTs can share similar characteristics and might even occur simultaneously, so an SVT is often mis-diagnosed as a VT. This is problematic

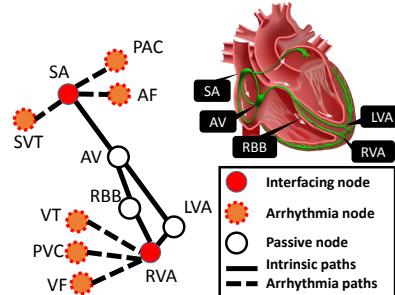


Figure 4: Timing model of the heart

because VT therapy consists of low and high energy electric shocks of 30-40 Joules (~800V) delivered directly to the heart, which causes severe stress to the patient, and might even be linked to increased morbidity [3]¹. Therefore, one of the biggest challenges for ICDs is to discriminate between fatal VTs that require a shock, and on-fatal SVTs that should not be shocked [11].

An EGM signal can be characterized by the *timing of events* that produced it, and the *morphology of the signal itself*. An ‘event’ is roughly characterized as the source of the largest peak in the EGM (e.g. a ventricular depolarization), and event timing is a crucial element of an arrhythmia’s definition in clinical Electrophysiology. The ‘morphology’ refers to the shape of the EGM (see Fig. 6 for examples). Both aspects are used by the ICD to make its decision. Correspondingly, our model has two components: a timing model, and a morphology model.

4.2 Timing Model

Computer models of the heart have been developed to model different aspects of the cardiac function to suit different applications ([14]). In [7], the authors developed a heart model structure that can be used to simulate the timing for generation and conduction of electrical events of the heart under a variety of conditions. The model structure consists of a set of node automata, which model the *generation* and *blocking* of electrical events by heart tissue, and a set of path automata, which model the *conduction delay* of electrical activity between node automata. The node and path topology used in the MBCT is shown in Fig. 4. The hollow nodes are passive nodes representing key locations within the heart where electrical events may be blocked. These include the Atrioventricular node (AV), Right Bundle Branch (RBB) and Left Ventricle Apex (LVA). The filled nodes in red, Sinoatrial (SA) node and Right Ventricle Apex (RVA) node, represent the heart locations where ICD electrodes are placed to measure the EGMs. The timing of the activation events at these nodes determines the timing of corresponding EGMs. Different sources for tachyarrhythmias are represented by arrhythmia nodes (dashed filled nodes) which are capable of self-activating at prescribed rates. These include Premature Atrial Complexes (PACs) and Premature Ventricular Complexes (PVCs) which are sources of rhythm disturbances.

Every node and path automaton has timing parameters that determine, for example, the delay between events, and how long it takes to conduct an electrical event between two nodes. These timing parameters can be directly derived from clinical data [15], and the model structure is compatible with clinical Electrophysiology concepts. Thus we know the ranges for these parameters. In [7], the timing model’s

¹Physicians compare a shock to a “horse kicking you in the chest”

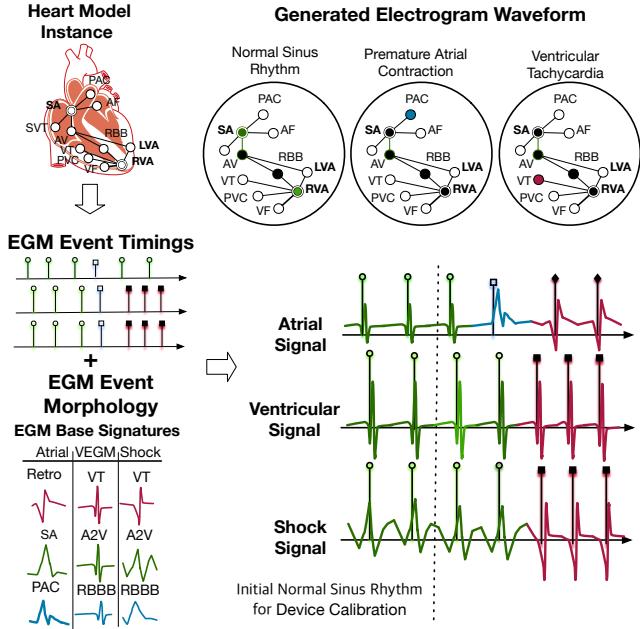


Figure 5: EGM waveform generation. From a given model instance and set of tachycardias, an EGM waveform is generated for the duration of an episode. The timing model determines event timings. When an event occurs, the EGM morphology for the event is output from the morphology model.

capability to simulate various normal and abnormal heart conditions was validated quantitatively and by cardiac electrophysiologists.

In this work we use the same heart model structure to ensure the correct timing of the EGM signals into the ICD. In order to account for inherent timing variability, during simulation the heart model randomly selects timing parameters within a pre-specified range, instead of choosing specific values. By choosing the range, we control the variability of the signals produced by a given model instance.

4.3 Morphology Model

The ICD uses the EGM morphology in two ways: first, the atrial and ventricular EGMs are used to *sense* when events occur via peak detection (Section 5.1). Second, the Shock channel EGM is used in the morphology comparison discriminators (Section 5.2, [16, 17]). It is known that sensing (the detection of events) can be responsible for up to 20% of inappropriate therapies [18]. Therefore, it is important that our model generate realistic and varied EGM waveforms for a proper evaluation of the detection algorithms.

The timing model provides the time stamps for electrical events to happen at the interfacing nodes (SA, RVA). From path conduction we also know the source of the signals. In the heart model structure shown in Fig. 4 there are 5 different sources for SA node activation and 5 different sources for RVA node activation. Based on the clinical observations that electrical events from the same source produce very similar EGM morphologies, we can generate EGM signals by overlaying EGM templates corresponding to different sources onto the timing event diagram. The procedure is shown in Fig. 5. We also introduce small variations on EGM templates. The variations are obtained by a wavelet decomposition of the signatures followed by a random scaling of the 25% smallest coefficients. We guarantee that this does

not change the signature of the EGM, by running one of the morphology comparison discriminators described in Section 5.2. This variation is parametrized, e.g. the percentage of modified coefficients, the range of the random scaling.

4.4 Patient Data Adjudication and EGM Template Extraction

In order to obtain realistic morphologies for our simulations we utilize the Ann Arbor Electrogram Libraries (AAEL), a database of over 500 EGM recordings made during clinical electrophysiology studies [19]. The AAEL is used by all major ICD manufacturers and is licensed by the US FDA. The AAEL provides descriptive annotations of records at a high level and needed additional detailed examination to precisely segment each record according to rhythm type. 123 records from 47 patients were manually examined and adjudicated into segments called *episodes* containing one specific rhythm, e.g. NSR or VF. The adjudication was performed by an experienced cardiologist. Fig. 6 (left) shows an example record (Record A185660) which has undergone this adjudication. From each episode, we developed an automated process which extracted EGMs from a given episode. The EGM are collected and organized by both patient record and by the type of rhythm which was annotated during the adjudication process. These extracted rhythm *signatures* provide the basis for the morphology information in the signal generated by our model. Fig. 6 (right) depicts an example of 10 signatures extracted from the record.

4.5 Cohort generation

Let $p = (p_1, \dots, p_n) \in \mathbb{R}^n$ be the vector of timing and morphological parameters of the heart model. Let $P_i \subset \mathbb{R}$ be the range of parameter p_i . We generate a *synthetic cohort* of N probabilistic model instances. To produce one of these instances, for each scalar parameter p_i , we randomly select a sub-interval I_i of its range: $I_i \subset P_i$. The sub-interval I_i is chosen so that it fits with the tachycardia that this model instance is meant to simulate. E.g., for modeling VT, the rest period of the VT node might be assigned the sub-interval $I_i = [260, 280]ms$, reflecting the firing rate in the ventricles. When a model instance is simulated, each parameter p_i 's value changes beat to beat by sampling it uniformly within its sub-interval I_i . Thus each generated model is probabilistic to reflect inherent rhythm variability.

5. IMPLEMENTING DEVICE ALGORITHMS

Due to the limited sensing capability of ICDs, device manufacturers developed different algorithm components to identify the electrical events and diagnose the heart conditions from the timing and morphology of EGM signals. In this paper we implemented the detection algorithm Rhythm ID of Boston Scientific [12, 11], and PRLogic+Wavelet of Medtronic [20, 17]. In available literature on the evaluations of device algorithms there are descriptions of the device algorithms which are not detailed enough for full implementation. To obtain the detailed implementation we also reviewed clinical execution traces from clinical literature like [20] that can be used to infer detailed executions of the algorithms. We set up a testing platform to validate our implementations against real ICDs using conformance testing.

5.1 Cardiac Signal Sensing

Sensing is the process by which cardiac signals measured through the leads of the ICD is converted to cardiac timing events. Appropriate sensing is essential for proper ICD de-

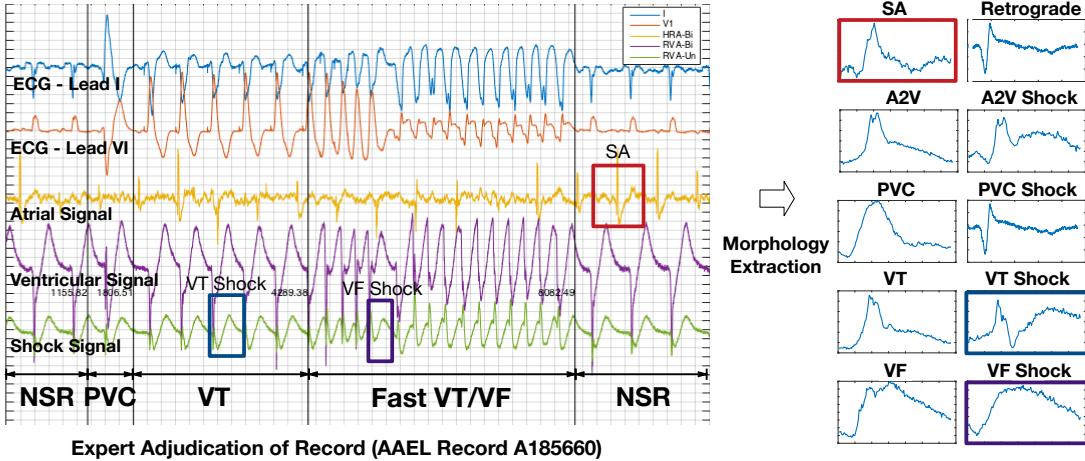


Figure 6: (Left) The EGM record is segmented into episodes with distinct rhythms in each. (Right) From each episode, individual EGMs morphologies are extracted and stored.

tecture algorithm operation which relies heavily on accurate event timing and morphology information provided by sensing. An *event* corresponds to a depolarization in the heart and manifests as a displacement from the baseline amplitude of the signal. In its simplest form, the sensing algorithm declares an event whenever the amplitude of the signal exceeds a given threshold. Once an event has been declared, the peak of the amplitude is measured and a *refractory period* begins, during which a consecutive event is ignored for a short period of time. This is to ensure that the same event is not counted repeatedly.

ICDs require a balance in sensitivity in order to operate in noisy, complex, environments where cardiac events can vary greatly in signal amplitude and frequency, such as during VF. Setting the threshold low achieves higher sensitivity to events of small amplitude, but increases chances for incorrectly sensing noise artifacts, such as T-waves, as events (oversensing). Conversely, setting the threshold too high allows sensing to be more robust to noise, but opens the potential for missing events, such as during VF when the peak amplitude can be low (undersensing).

In order to achieve such a balance, ICD sensing algorithms are enhanced by applying dynamic adjustment of the sensitivity threshold. Initially, the threshold is raised during the refractory period after an event and after the refractory period concludes, the threshold is decayed until set at a minimum threshold for maximum sensitivity. In our device models, we implemented the automatic gain control (AGC) algorithm of Boston Scientific and the auto-adjusting sensitivity (AAS) of Medtronic ICDs to incorporate dynamic threshold adjustment.

In addition to AGC, in our device model for Boston Scientific, we have implemented a specific *cross-chamber blanking* feature present in Boston Scientific ICDs. Cross-chamber blanking starts a blanking period in atrial sensing after a ventricular event has been detected. This reduces the chance of a ventricular event being interpreted as an atrial event.

The added complexity of these enhancements to the sensing algorithm adds to the difficulty of properly programming device settings of ICDs and requires calibration process at implantation. The calibration process is repeated in follow-up visits.

5.2 VT Detection Algorithm

The limited sensing resolution and noise in the EGM sig-

nals make it is impossible for the device to achieve 100% accuracy for SVT/VT discrimination. Device companies have developed different algorithmic components to distinguish SVT vs. VT, referred to as *discriminators*. These discriminators utilize the timing and morphological information in the EGM signals. No single discriminator is sufficient on its own to discriminate between SVT and VT, because these classes of arrhythmias can appear similar in a number of criteria. Therefore discriminators are organized in a decision tree, as shown in Fig. 7 for Boston Scientific's detection algorithm [12].

Initial Detection Duration: The algorithm starts with an initial detection phase, which examines the last 10 ventricular intervals and compares them with VT and VF thresholds. If 8/10 intervals are faster than the threshold the algorithm examines whether the fast rhythm is sustained for a certain *Duration*. Since VF is a lethal arrhythmia, if the rhythm remains faster than the VF threshold at the end of the *Duration*, the device delivers therapy without further diagnosis. At the end of the VT duration the algorithm runs through RhythmID, which further distinguish between VT/SVT. If the ventricular rate is greater than the atrial rate for the last 10 ventricular beats, which is an important indication of VT, RhythmID will determine the condition is VT. As shown in Fig. 7(a), an episode is identified at the end of the duration, and due to the faster rate in the ventricles, Rhythm ID decided the condition is VT and delivered therapy.

Rhythm ID: If the atrial rate is similar or faster than the ventricular rate, the Vector Timing and Correlation (VTC) component [16] compares the current EGM morphology with an EGM *template* saved during NSR. The VTC component is based on the assumption that the EGM morphology of the shock channel during VT is different from the morphology during SVT and NSR. If the correlation between the current morphology and the NSR morphology is above certain threshold, the current rhythm is more likely to be SVT than VT. As shown in Fig. 7(b), at the end of the duration, the atrial rate is faster than the ventricular rate. The morphology in the shock channel is similar to the NSR morphology (red dashed) so VTC is correlated and therapy is inhibited. As a comparison the morphology during VT is different from the NSR morphology (Marker 1 in Fig. 7).

Besides rate-based timing components, the Medtronic detection algorithm also continuously matches 19 different pat-

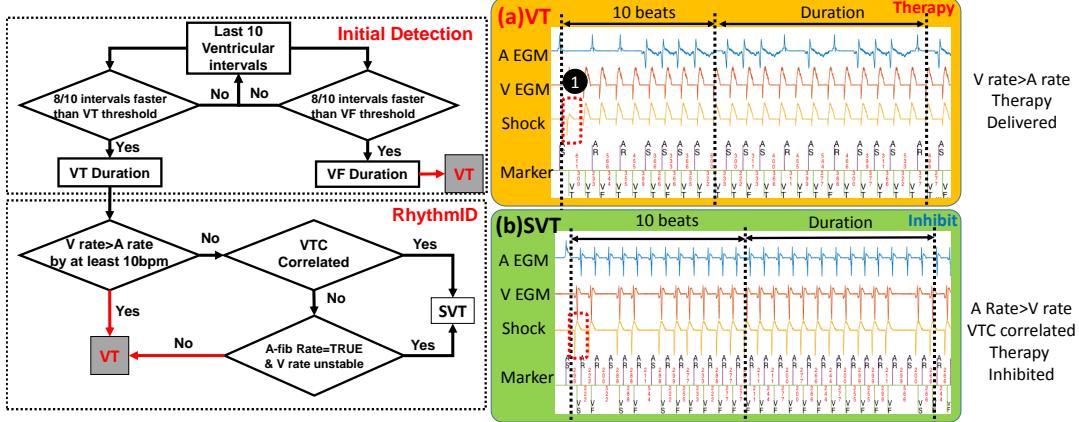


Figure 7: SVT/VT detection algorithm by Boston Scientific [12]

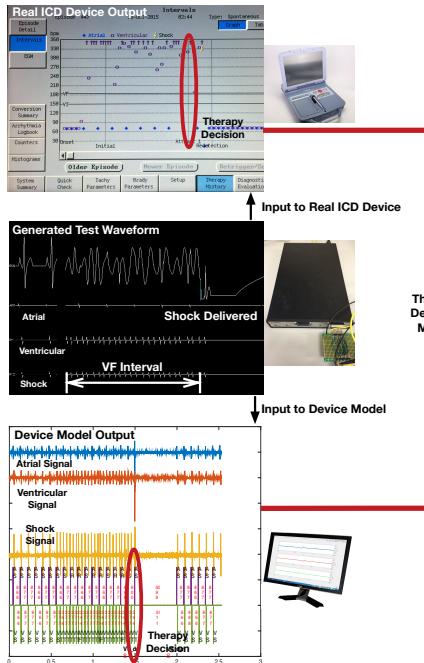


Figure 8: Example of validation output screenshots (Ventricular fibrillation) showing matching therapy decision for the ICD and our implementation.

terns for atrial and ventricular events. Different sequences of patterns are then categorized into different heart conditions [20]. Combined with a morphology component referred to as Wavelet [17], which is similar to the VTC algorithm of Boston Scientific, the Medtronic algorithm should perform better in terms of distinguishing SVT vs. VT. We implemented the Medtronic algorithm for the MBCT, but due to space limit the detailed description of PR-Logic will not be introduced here.

5.3 Validation

Conformance testing was used to validate the software implementations of the Vitality II device by Boston Scientific. The validation hardware setup is illustrated in Fig. 8. 14 different scenarios were specified and programmed into an EGM Waveform generator (CRM3 Simulator, Guidant,

USA) such that it would output a signal to the connected Vitality II device. The various scenarios traverse 7 out of 9 branches of the detection algorithm for Boston Scientific described in Sec. 5.2 and shown in Fig. 7. The response of the ICD interrogated using an ICD programmer (ZOOM Latitude, Boston Scientific). As the waveform was applied to the ICD, the waveform was simultaneously acquired using a National Instruments Data Acquisition (DAQ) board. The recorded waveform was then applied to the device model and response was compared. Fig. 8 shows an example of one such scenario, specifically VF. In this case, the software model matched to the decision of the actual ICD which also determined that therapy should be applied.

In all scenarios, the decision of model conformed to that of the ICD. The remaining two branches were not reachable due to the limited output capability of the programmer. The Medtronic software implementation can be validated using a similar process.

6. RESULTS

6.1 The rate of inappropriate therapy

The first objective of the MBCT is to estimate the rate of inappropriate therapy \bar{t} for each of the two algorithms for all arrhythmias combined, i.e., for the entire synthetic cohort. The rate of inappropriate therapy is defined as

$$\bar{t} = \frac{\text{Number of inappropriately applied therapies}}{\text{Number of applied therapies}}$$

From this we can confirm or invalidate the assumption that Boston Scientific's algorithm outperforms Medtronic's. We generated a synthetic cohort of 11,400 heart instances, equally distributed among the 19 arrhythmias. The number of instances was obtained from a Monte Carlo calculation.

Conclusion 1: Medtronic delivers less inappropriate therapy. The obtained rates of inappropriate therapy were 6.65% for Boston Scientific and 2.91% for Medtronic ($P < 0.0001$), assuming an equal number of patients from each arrhythmia in the synthetic cohort. The corresponding relative improvement of *Medtronic over Boston Scientific* is 56%. In other words, the MBCT reveals that Medtronic actually delivers *less* inappropriate therapy than Boston Scientific. This agrees with the findings of the RIGHT trial itself [4] from 2012, and is the opposite of what was hypothesized by the RIGHT investigators before the beginning of the trial [13] in 2006.

Conclusion 2: result holds across population characteristics. The above rates were obtained under the assumption that each arrhythmia is equally represented in the cohort. A significant feature of MBCT is that it allows us to study the endpoint of interest (here, rate of inappropriate therapy) on a variety of populations, which have the various arrhythmias in different proportions. This is clearly not feasible in a real clinical trial, which has to contend with the population present at the clinical centers where the trial is conducted. We may then ask: does Medtronic maintain a better rate of inappropriate therapy across different populations? To answer this question, we varied the distribution of the arrhythmias in the synthetic cohort, and re-computed the cohort-wide rates of inappropriate therapy. Fig. 9 shows the results for 10 random variations of the arrhythmia distribution. It can be seen that indeed, Medtronic’s algorithm maintains a better rate of inappropriate therapy across the board. Thus the results are robust to the characteristics of the population under study.

This illustrates very well the benefit that an MBCT can bring to the planning of an RCT: the fact that Boston Scientific’s algorithm could not be shown to be better than Medtronic’s (let alone 25% better as was hypothesized by the investigators) can cause the investigators to re-consider their assumptions, including the feasibility of the trial. In this case, the MBCT casts doubt on the assumed *direction* of the effect, i.e. whether intervention is better than control, or the other way around. This early check can mean the difference between an expensive trial that fails at showing the desired effect, and a trial that is appropriately sized to demonstrate the desired effect size (if it holds).

While we will not know the true distribution of the arrhythmias in the RCT until its completion, we can synthesize approximations of the population to get *early insight* at a small fraction of the RCT cost and duration and without the ethical issues. The role of the MBCT is not to replace or mimic the RCT as we can capture episode-level performance but not patient-level outcomes of the therapy.

6.2 Condition-level rates

Having a heart model allows us to estimate the *sensitivity* and *specificity* of the device, something which is not possible in a clinical trial because the device only records a limited number of episodes. These are defined as

$$\text{Sensitivity} = \frac{\text{Number of correctly classified VTs}}{\text{Number of true VTs}}$$

$$\text{Specificity} = \frac{\text{Number of correctly classified SVTs}}{\text{Number of true SVTs}}$$

In words, the sensitivity measures how well the device recognizes VTs. Specificity measures how well the device discriminates between VT and SVT. An ideal device would have 100% sensitivity and specificity. Unfortunately, these are typically competing goals: the more sensitive the device, the more likely it will mis-diagnose some SVTs as VTs, so its specificity will drop.

We calculated sensitivity and specificity in our MBCT, and report them in Table 1 on a per-arrhythmia basis. The conditions are drawn from RIGHT’s baseline characterization [4]. Specificity is reported for SVTs and sensitivity is reported for VTs. It can be seen from these results that in our synthetic cohort, Atrial flutter and other Supraventricular tachycardias are the main source of difficulty for Boston Scientific’s algorithm. In the case of Atrial flutter, the Boston Scientific algorithm provides therapy inappropri-

Table 1: Specificity for SVTs and sensitivity for VTs.

Arrhythmia	Boston Sci. ICD	Medtronic ICDs	P value
Specificity (%)			
Atrial Fibrillation	99.8	99.6	0.3167
Atrial flutter	58.3	79.33	<0.0001
Premature ventricular complexes	100	100	1
Nonsustained ventricular tachycardia	100	99.8	0.3171
Other Supraventricular tachycardia	96.3	99.7	<0.0001
Brady-Tachy	100	98.83	0.0079
Sensitivity (%)			
Ventricular fibrillation	100	100	1
Ventricular tachycardia	100	100	1

ately for 41.7% of the cases (i.e. it misdiagnoses the condition as a VT or VF condition when it is an atrial condition that does not require therapy).

Condition-level analysis pinpoints the specific pathways of the discrimination algorithm which must be addressed to reduce the device’s rate of inappropriate therapy. It is difficult to get such insight through an RCT as the patient population is fixed and the conditions are determined retroactively. Such analysis can be further used to investigate condition distributions across different patient population types (e.g. abnormal heart rhythms in children vs geographic region-specific or race-specific condition distributions).

6.3 Effect of Device Parameters on Discriminating Capability

ICDs have a number of parameters which can be tuned to accommodate specific patient conditions by the physicians. Currently there are very few clinical results on the effect of tuning parameters and their effect on sensitivity and specificity [2]. One of the main causes of VT/SVT misclassifications is inappropriate parameter settings [18]. In order for the physicians to set appropriate parameters, it is very important to understand how the change of one parameter can affect the discriminating capability of the device. It is costly to experiment this on real patients. With MBCT, one can use the same population across multiple devices with different parameter settings at virtually no cost.

In this section, we use MBCT to demonstrate the effects of changing two common parameters on SVT/VT discrimination specificity. The first parameter is the *duration* of arrhythmia before the ICD makes a therapy decision. For Boston Scientific ICD the value can be set to 1 to 30 seconds. In this experiment we explore the values {1,2,3,4,5,8,10}. The equivalent parameter for Medtronic ICD is the number of consecutive fast ventricular intervals which can be set from 8 to 20 beats. In this experiment we explore the values {8,10,12,16,18,24,30} which roughly correspond to the parameters of Boston Scientific ICD. Intuitively, with a longer duration the device can examine a longer history of the arrhythmia episode, which can prevent inappropriate therapy, and thus increase SVT/VT discrimination specificity. Setting the duration too long can also cause missed therapies, thus affecting sensitivity. These results are in agreement with the recently conducted ADVANCE-III RCT which showed that longer arrhythmia detection windows reduce shocks for Medtronic ICDs [21].

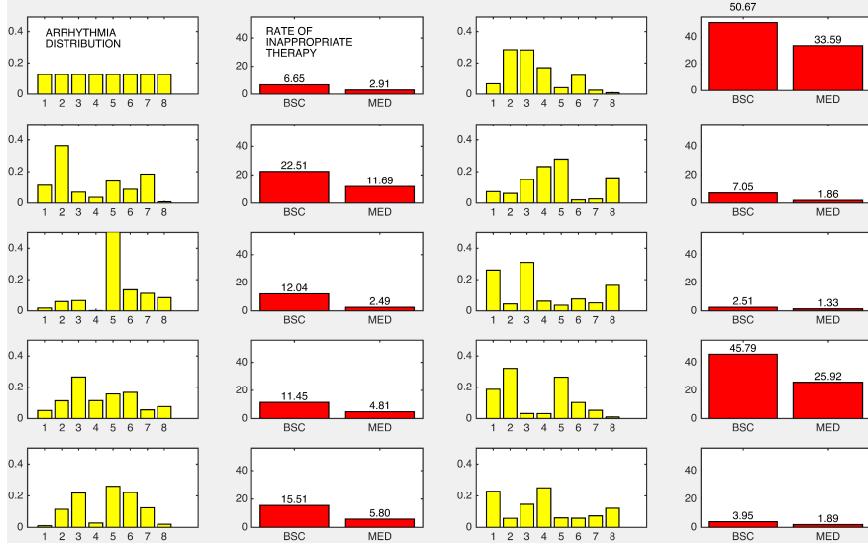


Figure 9: Rate of inappropriate therapy (2^{nd} and 4^{th} columns) for different arrhythmia distributions (1^{st} and 3^{rd} columns). The arrhythmias are (left to right on the x axis): Atrial fibrillation, Atrial flutter, Premature Ventricular Complexes, Nonsustained Ventricular Fibrillation, Supraventricular Tachycardia, Sinus Brady-Tachy, Ventricular Fibration, Ventricular Tachycardia [15]. The top left distribution is uniform, and the bottom right distribution is that of the baseline characterization in **RIGHT** [4].

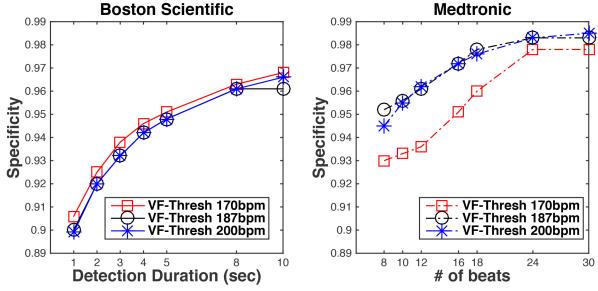


Figure 10: Effects of Duration and VF threshold parameters on Specificity

The second parameter we vary is the **VF threshold**. For both devices, if the ventricular rate is faster than the VF threshold for a period of time the devices will deliver therapy without going into the SVT/VT discrimination algorithm. So a higher VF threshold means that more signals are passing through the discrimination algorithm. The value can be set to 150 to 200BPM. In this experiment we explore the value $\{170, 184, 200\}$ for both devices. Intuitively the higher the threshold, the more episodes will be examined by the SVT/VT discrimination algorithm, which may increase specificity. However, VTs with rate less than the threshold may also be classified as SVT, causing missed therapies.

Conducting the Model-based Clinical Trial

For each of the 21 parameter combinations described above, we ran a MBCT with 11,400 EGM episodes on both device models. From the results we observe that for both devices the specificity increases monotonically with the length of the duration. When the duration is longer than 5, sensitivities also dropped below 100%, which is in line with the intuition.

However, Boston Scientific algorithm and Medtronic algorithm displayed opposite trends for VF threshold. For Medtronic algorithm, the specificity increases when the VF threshold increases from 170BPM to 184BPM - i.e. a higher threshold admits more signals through the discrimination algorithm which performs better across all rates. For Boston

Scientific algorithm the specificity dropped when the VF threshold increases from 170BPM to 184BPM - i.e. the discrimination algorithm is less effective at higher rates. One possible interpretation of the result is that the Boston Scientific algorithm is more prone to inappropriate therapies for SVTs with ventricular rate between 170BPM to 184BPM, which is a very useful insight for the physicians to consider during parameter settings.

7. DISCUSSION

The above experiment has illustrated a practical application of the MBCT approach in the use of computer modeling for the support of clinical trial planning and execution. We now present the medically relevant limitations of this particular experiment, then discuss the MBCT approach in general. First, we did not account for post-shock detection (a phase of detection that follows the delivery of therapy), which was part of the **RIGHT** results. The Onset discriminator was not implemented so the results exclude its effects. **RIGHT** included both dual-chamber and single-chamber devices, whereas we only implemented the algorithms for dual-chamber devices. For the VTC and Wavelet algorithms, the literature did not specify which samples are taken from the electrogram. We chose to sample the electrogram uniformly in time, and validated that this gives correct results.

Clinical trials study the effect of an intervention *in the patient*, and report patient-level results (e.g., “The event of interest was observed in X% of patients in Group 1”). Our results are at the condition level: they take the form “the event of interest was observed in X% of generated conditions”. To produce patient-level estimates requires an estimate of how conditions are distributed among patients. This low-level data is not publicly nor readily available. A trial’s investigators, however, should be able to obtain such data from previous trials.

It is important to stress that in general, one should not expect *absolute numbers* from an MBCT to match those from a clinical trial, nor should this be the goal of the MBCT. For example, in this work, it is unlikely that our MBCT will

yield rates of inappropriate therapy that are equal to the rates obtained by RIGHT itself. The reasons for this are many:

- The RIGHT in vivo cohort, and our synthetic cohort, are not comparable: indeed, a myriad of factors affect the outcome of a clinical trial, e.g., whether some patients take up smoking. These factors are not modeled.
- The adjudication of episodes in RIGHT (and other trials) is limited by the fact that only therapy episodes were recorded by the devices. The adjudication process is further limited by the lack of surface EKGs, which makes it hard to reliably distinguish certain atrial arrhythmias. Neither of these is a limitation in MBCT since we have the ground truth: we know exactly what arrhythmia is being simulated by the model. Furthermore, the AAEL signals have both device electrograms (EGMs) and the corresponding surface EKGs which allow for precise adjudication.
- Experts may disagree on how to adjudicate the more complex episodes, so our classification of episodes from the AAEL database and the classification of the RIGHT investigators have an irreducible discrepancy. Again, this will affect the statistics that they and we compute.

That said, we can expect that a good heart model will reveal *the trend* of the results, such as improvement of intervention over control or not, as shown in this paper. The MBCT conducted here clearly showed that the Medtronic algorithms outperforms the Boston Scientific's and resulted in a negative outcome of RIGHT across all population distributions and relevant heart conditions. In hindsight, the Boston Scientific-sponsored RIGHT would have needed reconsideration prior to running it to prevent a failed outcome.

8. CONCLUSION

We demonstrated how to design and conduct a Model-Based Clinical Trial in support of RIGHT, a prospective trial that compared two ICDs from two manufacturers. The results showed that under various distributions of the arrhythmias in the synthetic cohort, Medtronic's algorithm delivered less inappropriate therapy than Boston Scientific's, in agreement with the findings of the actual trial. By providing early and reproducible tests of the main assumptions of a clinical trial *before trial start*, MBCTs can increase the chances of trial success.

9. REFERENCES

- [1] Ask The ICD. <http://asktheicd.com>, 2015. Accessed on 10/11/2015.
- [2] A. J. Moss et al. Reduction in inappropriate therapy and mortality through icd programming. *New England Journal of Medicine*, 367(24):2275–2283, 2012.
- [3] M. Rosenqvist, T. Beyer, M. Block, K. Dulk, J. Minten, and F. Lindemans. Adverse Events with Transvenous Implantable Cardioverter-Defibrillators: A Prospective Multi-center Study. *Circulation*, 1998.
- [4] M. R. Gold et al. 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):370 – 377, 2012.
- [5] A. V. Kaplan et al. Medical device development: From prototype to regulatory approval. *Circulation*, 109(25):3068–3072, 2004.
- [6] K. Ten Tusscher, R. Hren, and A. V. Panfilov. Organization of ventricular fibrillation in the human heart. *Circulation Research*, 100(12):e87–e101, 2007.
- [7] Z. Jiang, M. Pajic, and R. Mangharam. Cyber-Physical Modeling of Implantable Cardiac Medical Devices. *Proc. of IEEE*, 100(1):122 –137, Jan. 2012.
- [8] C. Toffanin, M. Messori, F. Di Palma, G. De Nicolao, C. Cobelli, L. Magni. Artificial Pancreas: Model Predictive Control Design from Clinical Experience. *J Diabetes Sci Technol*, 7:1470–1483, 2014.
- [9] Avicenna Consortium. In-silico Clinical Trials: How Computer Simulations will Transform the Biomedical Industry. *Roadmap document*, 2015.
- [10] L. M. Friedman and C. D. Furberg and D. L. DeMets. *Fundamentals of clinical trials*. Springer, 2010.
- [11] K. Ellenbogen, G. N. Kay, C-P Lau, and B. L. Wilkoff. *Clinical Cardiac Pacing, Defibrillation, and Resynchronization Therapy*. Elsevier, 2011.
- [12] Boston Scientific Corporation. The Compass - Technical Guide to Boston Scientific Cardiac Rhythm Management Products. *Device Documentation*, 2007.
- [13] R. D Berger et al. 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):749–753, 2006.
- [14] N. A. Trayanova and P. M. Boyle. Advances in Modeling Ventricular Arrhythmias: from Mechanisms to the Clinic. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 6(2):209–224, 2014.
- [15] M.E. Josephson. *Clinical Cardiac Electrophysiology*. Lippincot Williams and Wilkins, 2008.
- [16] M. Gold et al. Advanced Rhythm Discrimination for Implantable Cardioverter Defibrillators Using Electrogram Vector Timing and Correlation. *J Cardiovasc Electrophysiol*, 2002.
- [17] C. D. Swerdlow et al. 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.
- [18] J. P. Daubert et al. Inappropriate Implantable Cardioverter-Defibrillator Shocks in MADIT II: Frequency, Mechanisms, Predictors, and Survival Impact . *Journal of the American College of Cardiology*, 51(14):1357 – 1365, 2008.
- [19] J. M Jenkins and R. E Jenkins. Arrhythmia database for algorithm testing: surface leads plus intracardiac leads for validation. *Journal of Electrocardiology*, 36, Supplement 1:157 – 161, 2003.
- [20] Igor Singer. *Interventional Electrophysiology*. Lippincott William & Wilkins, 2001.
- [21] M Gasparini et al. Effect of Long-detection Interval vs Standard-detection Interval for Implantable Cardioverter-Defibrillators on Antitachycardia Pacing and Shock Delivery: The ADVANCE III Randomized Clinical Trial. *JAMA*, 309(18):1903–1911, 2013.