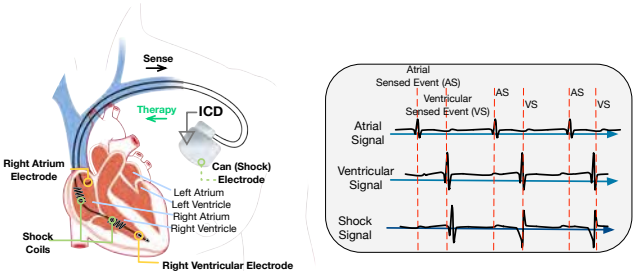
In-silico Pre-clinical Trials for Implantable Cardioverter Defibrillators

Zhihao Jiang[1] , Houssam Abbas[1] , Kuk Jin Jang[1] , Marco Beccani[1]
Jackson Liang[2] , Sanjay Dixit[3] and Rahul Mangharam[1]
[1]Dept. of Electrical and Systems Eng. University of Pennsylvania
{zhihaoj, habbas, jangkj, beccani, rahulm}@seas.upenn.edu
[2]Cardiovascular Division, [3]Cardiac 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 expose patients to an unproven device. In this paper, we demonstrate how to use a large group of synthetic patients based on computer modeling to improve the planning of a CT so as to increase the chances of a successful trial for implantable cardioverter defibrillators (ICDs). We developed a computer model of the electrical generation and propagation in the heart. This model was used to generate a large group of heart instances capable of producing episodes of 19 different arrhythmias. We also implemented two arrhythmia detection algorithms from the literature: Rhythm ID from Boston Scientific and PR Logic + Wavelet from Medtronic. Using this setup, we conducted multiple in-silico trials to compare the ability of the two algorithms to appropriately discriminate between potentially fatal Ventricular Tachy-arrhythmias (VT) and non-fatal Supra-Ventricular Tachy-arrhythmias (SVTs). The results of our in-silico trial indicate that Rhythm ID was less able to discriminate between SVT and VT and so may lead to more cases of inappropriate therapy. This corroborates the findings of the Rhythm ID Going Head to Head Trial (RIGHT), a clinical trial that compared the two algorithms in patients. We further demonstrated that the result continues to hold if we vary the distribution of arrhythmias in the synthetic population. We also used the same in-silico cohort to explore the sensitivity of the outcome to different parameter settings of the device algorithms, which is not feasible in a real clinical trial. In-silico trials can provide early insight into the factors which affect the outcome of a CT at a fraction of the cost and duration and without the ethical issues.**

I. Introduction

In the domain of cardiac devices, over 10,000 people in the U.S. receive an Implantable Cardioverter Defibrillator (ICD) every month [1]. An ICD is an electronic device that is implanted near and connected to the heart. It detects and treats chaotic, extremely fast, life-threatening heart rhythms, called fibrillations, by delivering a 30-40J (800V) shock to the heart, restoring normal rhythm of the heart . *After the device verification and testing effort is completed*, regulatory agencies such as the US FDA require that the safety and efficacy of new devices be demonstrated in a *Clinical Trial (CT)*. 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



% of Therapy Episodes Delivered Inappropriately	
Medtronic PRL + W	54.1%
Boston Scientific Rhythm ID	62.2%

Fig. 1. 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. [2]

physicians, patients, statisticians, clinical centers, companies and regulators, sometimes in several countries. For example, a 2002 trial for stents lasted 2 years, enrolled 800 patients and cost $10 to $12 million [3]. Trials might also expose patients in the intervention group 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. 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 ultimately fails to demonstrate the desired effect (typically, an improvement of the new intervention over the control).

We used the "Rhythm ID Going Head to Head Trial (RIGHT)" [2], which lasted five years (2005-2010) and sought to compare the VT/SVT discrimination algorithms used by two ICD models for correctly diagnosing potentially fatal tachycardias (abnormally fast heart rhythms). Upon completion, the evidence was *contrary* to the trail's hypothesis, hence the trail was considered a failure. ICDs suffer from a high rate of *inappropriate therapy*, which takes the form of unnecessary electric shocks or pacing sequences delivered to the heart during non-fatal heart rhythm. Inappropriate therapy increases patient stress and is linked to increased morbidity. Depending on the particular ICD and its settings, the rates of inappropriate therapy can reach 62% of all delivered therapy episodes [2]!

By modeling the heart's electrical activity and implementing the two VT/SVT discrimination algorithms, we find that our model-based results corroborate the findings of RIGHT

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

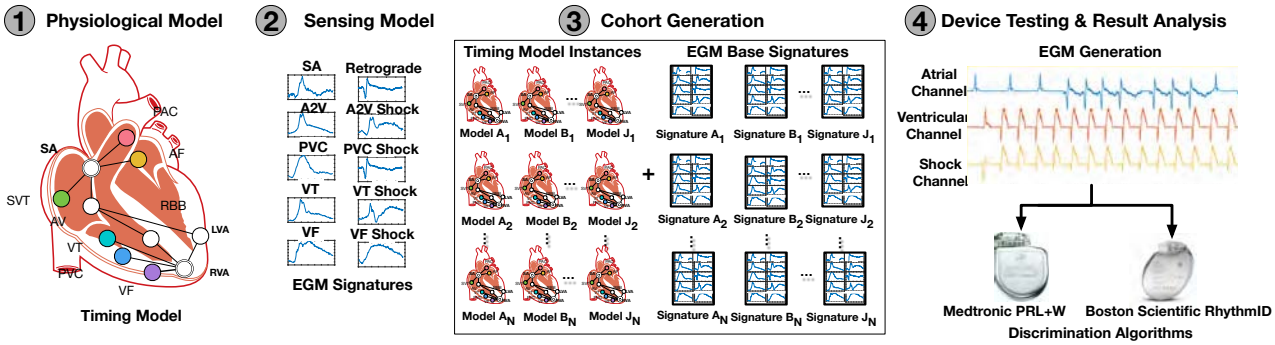


Fig. 2. Overview of an in-silico trial. 1) A model of the electrical conduction system is developed to model electrical activity of different heart conditions. 2) EGM morphologies corresponding to different signal sources were extracted from real patient data. 3) Through variation of the parameters of the model, a synthetic cohort, with over 11,000 heart models, is generated and simulated to produce synthetic EGM signals. 4) Device evaluation experiments are executed with this synthetic cohort using Medtronic PRL+W and Boston Scientific RhythmID discrimination algorithms (authors' implementation).

and provide stronger evidence to the sources of the trial's failure. Having these results early in the CT planning should allow the trial investigators to revise their assumptions and premises, and overall improve the chances of a successful trial.

II. Methods

Because an in-silico trial is designed and conducted in support of a given CT, the details necessarily depend on the CT we consider. Fig. 2 gives an overview of the in-silico trial we conducted in support of RIGHT.

We have access to a database of adjudicated arrhythmia episodes (electrograms (EGM), electrocardiograms (EKG) and signals sensed by an ICD) from real patients. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

(1) **Physiological Models - Timing:** Different heart conditions are described by the pattern and timing of electrical activity, which is studied in clinical Electrophysiology (EP) [4]. EP is the basis for ICDs thus it is the perfect level of abstraction needed for physiological modeling. We developed an automata-based EP heart model to simulate the electrical activities of the heart (see [5] for heart modeling details). The electrical conduction system of the heart is modeled with a set of nodes and the conduction paths connecting them (Fig. 2). Each node is an automaton modeling the timing of generation (T_{rest}) and blocking (T_{erp}) of electrical events, and each path connecting nodes is also an automaton modeling the conduction delay between nodes (T_{cond}). Tachycardia is modeled as nodes that can generate activation events with high frequency. The topology of nodes and paths represents the electrical connectivity between the atrial lead and the ventricular lead and different sources that can trigger depolarization events in the device.

(2) **Sensing Model - Morphology:** The ICD utilizes the timing and morphology of local electrical activities in EGM signals to diagnose the current heart condition. The activations of SA and RVA nodes generate EGM signals which are the inputs to the ICD. With different sources of activation, the morphology of the EGM signal may differ. We developed a sensor model which generates realistic EGM signals that capture possible sensing errors. Based on the clinical observation that electrical activities observed by the ICD lead have the same EGM morphology if they share the same origin, we create an *EGM signature* for each virtual patient which consists of EGM templates for 10 different signal sources. When the nodes (SA and RVA in Fig. 2 which interface with the ICD leads) are activated via any of the 10 different paths, the EGM for the activation will have the morphology corresponding to the respective EGM template. To capture inter-person variability in EGM morphology, EGM signatures were collected from adjudicated arrhythmia episodes from a database of EGM records of real patients (see Fig. 3 for examples). EGM signals are then synthesized by overlaying the EGM templates on the timing sequences of electrical events generated by the heart models (Fig. 2). For heart conditions with irregular EGM morphologies like in atrial and ventricular fibrillation, the EGM template is randomly alternating among several saved morphologies.

(3) **Cohort Generation:** The parameter ranges for the heart model are from two sources. For the cycle length of a given arrhythmia, we obtained a range from EP testing reports (Fig. 4). The ranges for the remaining parameters were obtained from open medical literature such as [4]. For each heart instance, ranges of parameters are uniformly sampled from the larger ranges in Fig. 4. For instance, the cycle length of a heart instance with SVT can be [290,300], which is a subset of the larger range [280,530]. In this manner, a synthetic cohort of $> 11,000$ models is generated, covering 19 common arrhythmic conditions.

(4) **Device Testing:** We implemented two VT/SVT discrimination algorithms: Rhythm ID from Boston Scientific [6], [7], and PR Logic + Wavelet (PRL+W) from Medtronic [8], [9], [7]. The ICD algorithms were implemented from open literature [6], [7], [9], [8] and the implementation was validated with real devices using conformance testing. Details can be found in the technical report [10]. Every

Condition	Number of Episodes	Cycle Length(msec)
SVT	14	280-530
VF	13	180-240
VT	119	200-540

Fig. 4. Arrhythmia timing parameters from EP testing reports for real patients

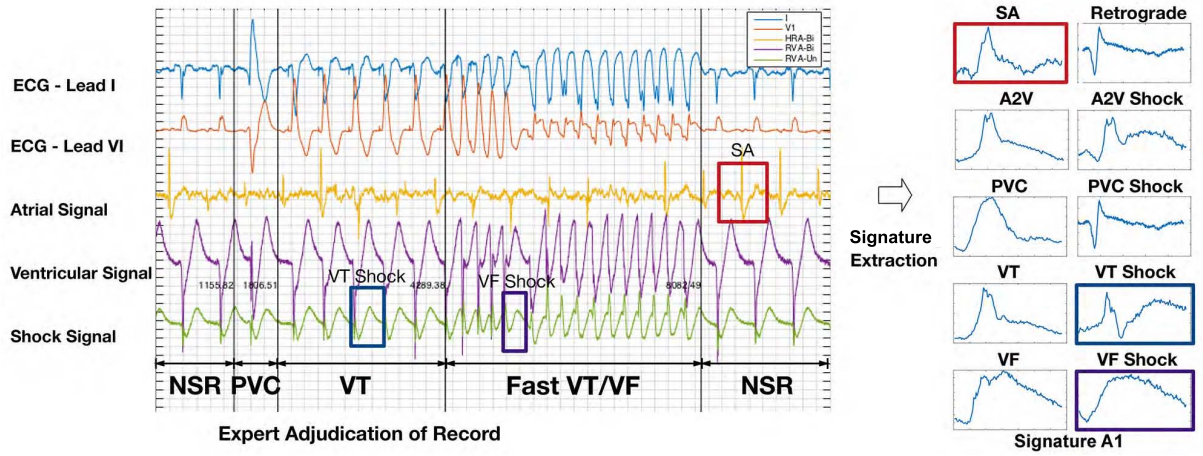


Fig. 3. EGM morphologies are identified and extracted from patient episodes. 10 EGM morphologies corresponding to different signal sources are extracted as EGM signature for each patient.

member of the synthetic cohort is then simulated to produce EGM signals that are fed to both detection algorithms, and their rates of inappropriate detection are analyzed. We repeated this analysis for various distributions of arrhythmias in our cohort.

III. Results and Discussion

A. The rate of inappropriate therapy

The first objective of the in-silico trial is to estimate the rate of inappropriate detection $\bar{t}$ for both algorithms across 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 Rhythm ID outperforms PRL+W. We generated a synthetic cohort of 11,400 heart instances, equally distributed among 19 arrhythmias. The number of instances was obtained from a Monte Carlo calculation and is over five times larger than the actual RIGHT trial.

Conclusion 1: PRL+W delivers less inappropriate therapy. The obtained rates of inappropriate detection were 6.65% for Rhythm ID and 2.91% for PRL+W ($P < 0.0001$), assuming an equal number of patients from each arrhythmia in the synthetic cohort. The corresponding relative improvement *of PRL+W over Rhythm ID* is 56%. In other words, the in-silico trial reveals that PRL+W algorithm differentiates between VT and SVT more often than Rhythm ID. Our findings are consistent with the observations of the RIGHT trial itself [2], and are purely model-based.

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 in-silico trials is that they allow us to study the endpoint of interest (here, rate of inappropriate detection) on a variety of populations, which have the various arrhythmias in different proportions. This may not be feasible in a real clinical trial, which has to contend with the population present at the clinical centers where the trial is conducted. We varied the distribution of the arrhythmias in the synthetic cohort, and re-computed the cohort-wide rates of inappropriate therapy. We conducted trials for 100 random variations of the arrhythmia distribution. Fig. 5 shows the results for the uniform distribution and a distribution that approximates that of RIGHT's cohort [2, Table 1]. It can be seen that indeed, PRL+W maintains a better rate of arrhythmia discrimination across the board.

In this case, the in-silico trial casts doubt on the assumed *direction* of the effect, i.e. whether intervention (Rhythm ID) is better than control (PRL+W), or the other way around. This early-stage 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.

B. Condition-level Analysis

A heart model allows us to better estimate the *sensitivity* and *specificity* of the diagnostic algorithms' performance, something which is not possible in a clinical trial because the device only records a limited number of episodes for which therapy was delivered. These are defined as

$$\text{Sensitivity} = \frac{\text{Nb of correctly detected sustained VTs/VFs}}{\text{Nb of sutained VTs/VFs}}$$

$$\text{Specificity} = \frac{\text{Nb of correctly detected SVTs/non-sus VTs}}{\text{Nb of SVTs/non-sustained VTs}}$$

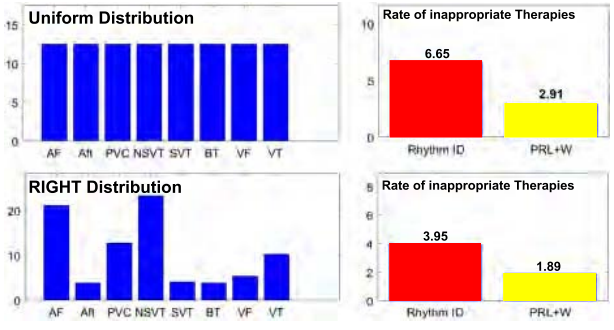


Fig. 5. Rate of inappropriate detection (2^{nd} column) for different arrhythmia distributions (1^{st} column). The upper-left distribution is uniform, and the lower-left distribution is that of the baseline characterization in RIGHT [2].

TABLE I
SPECIFICITY AND SENSITIVITY OF ICD VT/SVT DISCRIMINATION ALGORITHMS

Arrhythmia	Rhythm ID	PRL+W	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 (%)		**P value**
Ventricular fibrillation	100	100	1
Ventricular tachycardia	100	100	1

In words, the sensitivity measures how well the device recognizes sustained VTs. Specificity measures how well the algorithm discriminates between VT and SVT. An ideal algorithm would have 100% sensitivity and specificity.

We calculated sensitivity and specificity in our in-silico trial, and report them in Table I on a per-arrhythmia basis. The conditions are drawn from RIGHT's baseline characterization [2]. It can be seen from these results that in our synthetic cohort, Atrial flutter and other SVTs are the main source of inappropriate detection for Rhythm ID compared to PRL+W. In the case of Atrial flutter, Rhythm ID categorizes it inappropriately as Ventricular Tachycardia (VT) for 41.7% of the episodes.

Condition-level analysis pinpoints the specific decision 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 a CT 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 populations (e.g. abnormal heart rhythms in children vs geographic region-specific or race-specific condition distributions).

C. Effect of Device Parameters on Discriminating Capability

ICDs have a number of parameters which can be tuned by the physicians to accommodate specific patient conditions. Currently there are few clinical results on the effect of different parameter settings on sensitivity and specificity. One of the main causes of VT/SVT mis-classifications is inappropriate parameter setting [11]. 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 algorithm. With in-silico trials, one can subject the same synthetic population to different settings of the parameters at virtually no cost.

In this section, we use in-silico clinical trial 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. The parameter for PRL+W 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} . From the results (Fig. 6) we observe that the specificity increases monotonically with the length of the duration, which matches the intuition as the device can examine a longer history of the arrhythmia episode with longer duration, and also allows a greater chance for the arrhythmia to self-terminate. This can prevent inappropriate detections therefore prevent inappropriate therapies. However, setting the duration too long can delay, and in some cases withhold appropriate therapy, as sensitivities dropped below 100% when the number of consecutive beats is more than 18.

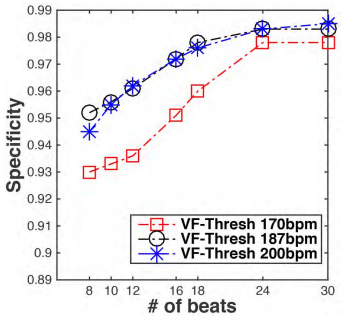


Fig. 6. Effect of Duration and VF threshold params on Specificity

The second parameter we varied is the ***VF threshold***. If the ventricular rate is faster than the VF threshold for a period of time the algorithm will confirm detection without going into the SVT/VT discrimination algorithm. In this experiment we explored the values {170,184,200} msec. As the parameter increases from 170BPM to 184BPM, more episodes will be examined by the SVT/VT discrimination algorithm, which may increase specificity (Fig. 6).

IV. LIMITATIONS AND DISCUSSION

In this paper, heart models were used to evaluate the capability of the VT/SVT discrimination algorithms to make appropriate therapy decisions. Therefore the heart models are not required to be able to respond to device therapies. Other applications of *in-silico* pre-clinical trials, i.e. evaluating the effectiveness of ICD therapies, require heart models with physiological complexity to interact with the ICD in closed-loop.

REFERENCES

[1] Ask The ICD. http://asktheicd.com, 2015. Accessed on 10/11/2015.

[2] 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.

[3] A. V. Kaplan et al. Medical device development: From prototype to regulatory approval. *Circulation*, 109(25):3068–3072, 2004.

[4] M.E. Josephson. *Clinical Cardiac Electrophysiology*. Lippincot Williams and Wilkins, 2008.

[5] Z. Jiang, M. Pajic, and R. Mangharam. Cyber-Physical Modeling of Implantable Cardiac Medical Devices. *Proc. of IEEE*, 100(1):122 – 137, Jan. 2012.

[6] Boston Scientific Corporation. The Compass - Technical Guide to Boston Scientific Cardiac Rhythm Management Products. *Device Documentation*, 2007.

[7] K. Ellenbogen et. al. *Clinical Cardiac Pacing, Defibrillation, and Resynchronization Therapy*. Elsevier, 2011.

[8] C. D. Swerdlow et al . Discrimination of Ventricular Tachycardia from Supraventricular Tachycardia by a Downloaded Wavelet Transform Morphology Algorithm. *J. Cardiovascular Electrophysiology*, 2002.

[9] Igor Singer. *Interventional Electrophysiology*. Lippincott William & Wilkins, 2001.

[10] H. Abbas, Z. Jiang, K. J. Jang, M. Beccani, J. Liang, S. Dixit, and R. Mangharam. Model-based Clinical Trials for Implantable Cardiac Defibrillators. *Technical Report (University of Pennsylvania):*, 2016.

[11] 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.