Title: Estimating the Probability of Cellular Arrhythmias with Simplified Statistical Models that Account for Experimentally Observed Uncertainty

Qingchu Jin[1], Joseph L. Greenstein[1], Raimond L. Winslow[1*]

[1]Department of Biomedical Engineering, Johns Hopkins University, MD 21218, USA

*To whom correspondence may be addressed.

Email: rwinslow@jhu.edu

Abstract

Early after-depolarizations (EADs) are action potential (AP) repolarization abnormalities that can trigger lethal arrhythmias in, for example, Long QT Syndrome and heart failure. Simulations using biophysically-detailed cardiac myocyte models can reveal how model parameters influence the probability of these cellular arrhythmias, however such analyses often pose a huge computational burden. Here, we develop a simplified approach in which logistic regression models (LRMs) are used to define a mapping between the parameters of complex cell models and the probability of EADs. Specifically, we develop an LRM for predicting the probability of EADs (P(EAD)) as a function of slow-activating delayed rectifier current (I_{Ks}) parameters, and for identifying those parameters with greatest influence on P(EAD). This LRM, which requires negligible computational resources, is also used to demonstrate how uncertainties in experimentally measured values of I_{Ks} model parameters influence P(EAD). We refer to this as arrhythmia sensitivity analysis. In the investigation of five different I_{Ks} parameters associated with Long QT syndrome 1 (LQTS1) mutations, the predicted P(EAD) when rank ordered for 6 LQTS1 mutations matches the trend in risk from patients with the same mutations as measured by clinical cardiac event rates. We also demonstrate the degree to which parameter uncertainties map to uncertainty of P(EAD), with I_{Ks} conductance having the greatest impact. These results demonstrate the potential for arrhythmia risk prediction using model-based approaches for estimation of P(EAD).

Keywords

Cardiac electrophysiology, myocyte computational modeling, model simplification, early afterdepolarization, Long QT syndrome, uncertainty analysis

Introduction

Early after-depolarizations (EADs) are a form of cellular arrhythmia that can trigger cardiac arrhythmias. EADs are abnormal repolarization events that occur during the plateau or rapid repolarization phase of action potentials (APs). EADs can trigger large scale cardiac arrhythmias in the setting of long QT syndrome (1-3). Computational modeling of EADs has provided insights into the mechanisms by which they can trigger arrhythmias in the heart (4-6).

We seek to understand how variations in the underlying biophysical properties or physiological state of the myocyte influence the probability of occurrence of cellular arrhythmias. We refer to this as arrhythmia sensitivity analysis. To do this, we use a three-dimensional model of the ventricular myocyte developed by Walker et al (7) in which the fundamental events governing intracellular calcium (Ca^{2+}) dynamics are modeled stochastically. This model has previously been shown to reproduce realistic Ca^{2+} waves and delayed after-depolarizations (DADs) driven by stochastic gating of L-type Ca^{2+} channels (LCCs) and Ca^{2+} release channels (ryanodine receptors, RyRs). Stochastic mechanisms also play a role in EAD initiation. For example, Fowler et al. showed that stochastic Ca^{2+} sparks and local Ca^{2+} waves can generate sufficient Na^{+}/Ca^{2+} exchanger current to trigger EADs (8). Tanskanen et al. showed that a phosphorylation-induced change in LCC stochastic gating contributes to EAD initiation (9). Here we explore an approach designed for estimating the probability of EADs that is more computationally efficient than our previous methods for estimating the probability of DAD induced ectopic beats (7). To do this we leverage techniques first developed by Sobie et al (10, 11) to formulate a logistic regression model (LRM) that maps myocyte model inputs (MMIs), which refers collectively to myocyte model parameters as well as state variable initial conditions, to probabilities of cellular arrhythmias. This approach allows for evaluation of the relationship between these event probabilities and their associated MMIs.

Action potential duration (APD) prolongation and QT-interval prolongation have been widely used as surrogate indicators for increased arrhythmogenic potential (12-15). However, the relationship between these surrogate features and the likelihood of arrhythmia is indirect. For example, ranolazine leads to QT prolongation (surrogate proarrhythmic indicator) in multiple clinical trials (16, 17) but has anti-arrhythmic effects (18). Rather than relying upon indirect indicators of arrhythmic potential, in this work we develop model-based methods for directly estimating the probability of occurrence of the actual cellular arrhythmic events of interest.

Long QT Syndrome type 1 (LQTS1) arises from mutations in the alpha subunit KCNQ1 of I_{Ks} (19). Studies show that mutant I_{Ks} in LQTS1 contributes to generation of EADs, which may then lead to Torsades de Pointes at the level of the whole heart (20, 21). Jons et. al. described electrophysiological changes associated with different LQTS1 mutations by identifying the altered values of five I_{Ks}-related parameters for each mutation (6). In the work reported here, these five slow potassium current (I_{Ks})-related biophysical parameters were chosen as MMIs to parameterize the LRM for P(EAD).

Use of an LRM to map MMIs to the probability of cellular arrhythmias also enables analyses of how experimental uncertainty in the measurement of MMIs influences estimates of these probabilities. One mainstream method for experimental uncertainty analysis is Monte Carlo (22), which has been used extensively in system biology and myocyte modeling (23-25). However, such methods require many repeated model simulations, an approach which becomes intractable when using complex models of the myocyte such as that of Walker et. al. (7). LRMs provide a computationally efficient approach for performing uncertainty analyses. We studied the impact of uncertainty in LQTS1 related I_{Ks} parameters on P(EAD).

The LRM mapping five MMIs (I_{Ks}-related parameters) to P(EAD) is developed. The trend in predicted P(EAD), when rank ordered for 6 well-characterized LQTS1 mutations, matches the trend in risk from patients with the same mutations as measured by clinical 30-year cardiac event rates. We also demonstrate the degree to which uncertainty in each parameter maps to uncertainty of P(EAD), with I_{Ks} conductance having the greatest impact. Results demonstrate the potential for arrhythmia risk prediction using model-based approaches for estimation of P(EAD).

Results

Modeling the Role of I_{Ks} on Probability of EADs. An LRM was formulated to capture the quantitative relationship between P(EAD) and 5 slow K^+ current (I_{Ks})-related MMIs (activation time constant scaling factor τ_+_sf; deactivation time constant scaling factor τ_-_sf; shift of the half maximal activation voltage $\Delta V_{1/2}$; scaling factor for voltage-dependent activation curve slope k_sf; and maximal conductance scaling factor G_{Ks}_sf). These 5 MMIs were chosen because they have been experimentally measured for a number of different LQTS1 mutations (6). Fig. 1A shows examples of APs without (blue) and with (red) EADs. Because of the stochastic nature of the myocyte model, each realization may generate an AP with or without EADs. For each realization, an EAD is considered to have occurred if the membrane voltage at 500 ms following the time of stimulus current is greater than -40 mV (Fig. 1A). We defined a biophysically meaningful range (region of interest, ROI) of MMIs within which the LRM is built (Table S3). Fig. 1B demonstrates the relationship between MMIs within the ROI and P(EAD), where $\underline{B}_f^T\underline{P}_f$ (the argument of the logistic regression function) is the weighted ($\underline{B}_f$) summation of LRM features ($\underline{P}_f$) (see Materials and Methods Eq. 2). This nonlinear logistic relationship allows for identification of three domains in the ROI of MMIs: transition domain (TD), left plateau domain (LPD), and right plateau domain (RPD). The TD is a region in which P(EAD) is a sensitive function of $\underline{B}_f^T\underline{P}_f$, corresponding to the steep part of the curve in Fig. 1B. The LPD and RPD are both regions where P(EAD) is relatively insensitive to MMI perturbation, with P(EAD) = ~0 in the LPD and P(EAD) = ~1 in the RPD. To build the LRM, we sampled 200 MMI sets within the ROI and ran 100 realizations for each MMI set from which we obtain P(EAD). Since the TD is relatively narrow and P(EAD) is a very steep function within it, i.e., strongly dependent on MMI values, we have developed a two-stage sampling strategy to ensure adequate sampling of the TD (see Materials and Methods and Supplement for details). These 200 MMI sets and their corresponding P(EAD)s were used to train the LRM. The LRM enables a direct calculation of P(EAD) for any MMI set without simulation of the myocyte model (MM). Figure 1B shows the fidelity with which the LRM model (blue line) reproduces P(EAD) estimated from MM simulations (red points) for the training data set (200 MMI sets) as a function of $\underline{B}_f^T\underline{P}_f$. Fig. 1C shows the predicted P(EAD) from LRM versus actual P(EAD) computed from MM simulations using the training data set. The LRM performs well in predicting P(EAD) (R^2 = 0.971 for the 96 MMI sets inside the TD, mean prediction error of 0.022 ± 0.036 as per Materials and Methods Eq. 3), despite the highly complex nonlinear properties of the MM. To test generalizability of the LRM, Figs. 1D-E compare the LRM predicted P(EAD)s to actual MM-generated P(EAD)s using test data consisting of 100 independent MMI sets. The LRM performs well on the test set (R^2 = 0.933 for the 27 MMI sets within the TD, mean prediction error = 0.019 ± 0.043). Detailed modeling methods and performance metrics are described in Materials and Methods.

The LRM for P(EAD) was formulated with a total of 16 features, 5 being the linear features τ_+_sf, τ_-_sf, $\Delta V_{1/2}$, k_sf, and G_{Ks}_sf, and 11 being quadratic features. Table 1 shows the LRM weights for each feature ranked by feature importance. In order to ensure interpretability of feature importance from the LRM weights, all MMIs were scaled into a range from 0 to 1. Feature G_{Ks}_sf is the most important, and this is followed in rank-order by τ_+_sf*G_{Ks}_sf, k_sf, and $\Delta V_{1/2}$*G_{Ks}_sf. The second and fourth most important features are both quadratic features, which improve LRM accuracy when compared to an LRM formulated using only the linear features as inputs (Fig. S5B, R^2 = 0.859 within TD, mean prediction error = 0.028 ± 0.063 for 100 test MMI sets).

These results presented here demonstrate that LRMs can accurately predict P(EAD), the probabilities of physiologically relevant cellular arrhythmia events. For the 100 test MMI sets used in each case, thousands of CPU hours are required to obtain event probabilities from the MM whereas evaluation of LRMs requires negligible computing time on the order of milliseconds. The seemingly intractable task of estimating these probabilities in an efficient manner is made possible by this LRM approach.

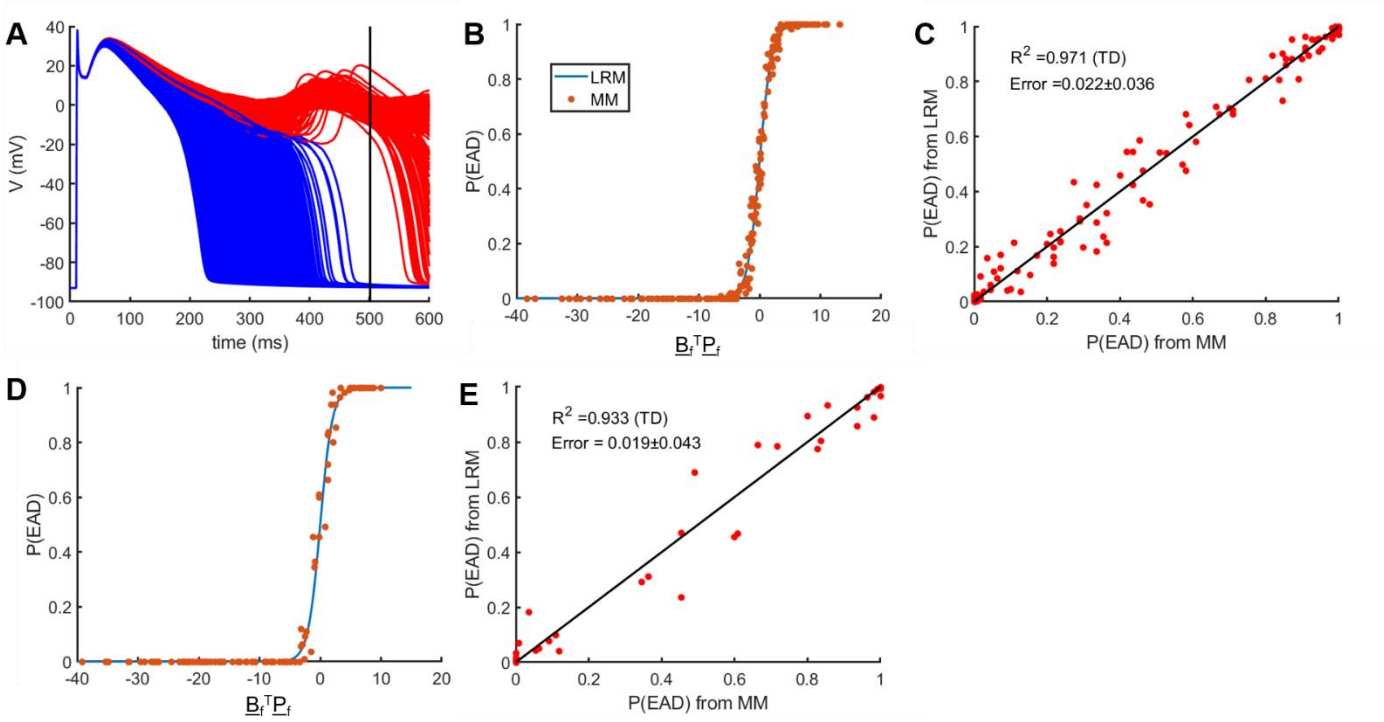

Fig. 1 EAD study. (A) Example APs from myocyte model simulations. Representative APs without EADs (blue) and APs with EADs (red) are shown. (B) Comparison between LRM-predicted P(EAD) (blue) and MM-generated actual P(EAD) (red) as a function of $\underline{B}_f^T\underline{P}_f$ for the training set. (C) Direct comparison between LRM-predicted P(EAD) (y-axis) and MM-generated actual P(EAD) (x-axis) for the training set. (D) and (E) show the same analyses described in (B) and (C), respectively, for the independent test set. R^2 is calculated only from MMI sets within the TD, LRM: logistic regression model, MM: myocyte model

Table 1. EAD features ranked by relative importance.

Ranking	Feature	Weight ± SE from LRM
1	G_{Ks}_sf	-63.89 ± 1.86
2	τ_+_sf * G_{Ks}_sf	34.35 ± 2.07
3	k_sf	14.57 ± 0.65
4	$\Delta V_{1/2}$ * G_{Ks}_sf	12.83 ± 1.61
5	τ_+_sf	11.90 ± 0.59
6	τ_+_sf ^{2}	-10.27 ± 0.73
7	τ_-_sf * G_{Ks}_sf	8.54 ± 1.11
8	k_sf ^{2}	-7.67 ± 0.41
9	k_sf * $\Delta V_{1/2}$	-6.31 ± 0.43
10	$\Delta V_{1/2}$	-6.05 ± 0.51
11	τ_-_sf ^{2}	-5.29 ± 0.36
12	τ_+_sf * $\Delta V_{1/2}$	-4.66 ± 0.69
13	G_{Ks}_sf * k_sf	-4.00 ±1.11
14	τ_-_sf	3.09 ± 0.62
15	k_sf * τ_-_sf	2.07 ± 0.37
16	τ_-_sf * $\Delta V_{1/2}$	-1.26 ± 0.37

Effect of LQTS1 Mutations on EAD Probability. LQTS1 arises from mutations in KCNQ1, the alpha subunit of the I_{Ks} channel (19). The symptoms of LQTS1 include syncope, cardiac arrest, and sudden death (26). Jons et al. experimentally characterized the electrophysiological properties of I_{Ks} expressed in Xenopus oocytes for LQTS1 mutations and reported the 30-year cardiac event rate for each mutation (6). The 30-year cardiac event rate is

defined as the percentage of people within a population of carriers of the mutant gene, who experienced syncope, cardiac arrest, or sudden death in the first 30 years of life. For each mutation, 5 I_{Ks}-related functional parameters were measured experimentally (6). Measurements were obtained from no less than 16 oocytes for each mutation, and experimental uncertainty (EU) of each parameter was estimated using standard error (SE). Detailed experimental and clinical data are summarized in Table S1. In this study, we evaluate the: 1) effect of experimental uncertainty of the estimated I_{Ks}-related functional parameters on P(EAD); and 2) use the LRM formulation for P(EAD) to predict the severity of LQTS1 mutations.

An LRM maps a given MMI set to a specific P(EAD). However, there is experimental uncertainty (EU) in the measured values of each MMI set. Jons et al. quantify this uncertainty by providing a mean (μ) and standard deviation (σ) for each of the five parameters measured experimentally (Table S1). To investigate the effect of this EU on P(EAD) for each mutation, MMI sets are constructed by choosing each of the five MMIs as independent normally distributed RVs with μ and σ as measured by Jons et al. For each random choice of MMI set, the LRM is used to compute P(EAD). In this way, P(EAD) itself is a random variable with a distribution. We estimate this distribution using Monte Carlo methods by choosing 10^6 normally distributed MMI sets for each mutation (see details in Materials and Methods). Distributions estimated in this way reveal the impact of EUs on estimates of the P(EAD).

Fig. 2A-F show the P(EAD) distribution for the six well-characterized mutations G314S, L266P, T312I, G168R, W305S, and Y315C. The P(EAD) distribution for G314S is narrow with a single peak at P(EAD) near one. This indicates that despite the measured experimental uncertainties, we can be quite certain that this mutation leads to EADs. The remaining distributions are bimodal with peaks near P(EAD) equal to zero and one. In these cases, MMI sets belong to two subgroups, one that almost always leads to EADs and another that seldom does. These bimodal distributions of P(EAD) can be approximated by a Bernoulli distribution with parameter p, where p = P(P(EAD) >0.5) for any particular MMI set and 1-p = P(P(EAD) <0.5). The parameter p, which can be approximated as the mean probability of EAD, is shown by the red vertical line in panels A-F. It increases gradually in the following mutation order: Y315C, W305S, G168R, T312I, L266P, and G314S.

To explain the observed P(EAD) distributions, a comparison of probability density function (PDF) of $\underline{B}_f^T\underline{P}_f$ (red) for each mutation and the model characteristic (MC) curve (blue) of $\underline{B}_f^T\underline{P}_f$ is shown in Fig. 2G. Recall that $\underline{B}_f^T\underline{P}_f$ is a function of the MMIs which appears in the LRM so that PDFs of $\underline{B}_f^T\underline{P}_f$, for each mutation, can be derived from the Monte Carlo method with μ's and σ's measured by Jons et al (6). The MC curve of $\underline{B}_f^T\underline{P}_f$ is the $\underline{B}_f^T\underline{P}_f$ - P(EAD) relationship curve determined by the LRM. For mutation G314S, the PDF (squares) coincides with the RPD and part of the TD of the MC curve. This results in a P(EAD) distribution that is unimodal and close to 1. For the remaining 5 mutations, PDFs all coincide with and span all domains (i.e. LPD, TD, and RPD) of the MC curve, thus resulting in a P(EAD) distribution that is bimodal.

In Fig. 2H, the mean P(EAD)s for each mutation are compared with the 30-year cardiac event rate reported by Jons et al (6). The correlation of predicted P(EAD) and cardiac event rate demonstrate that the LRM-predicted P(EAD) may be a useful indicator of the mutation-specific risk of clinical arrhythmias in patients. To validate this use of the LRM in uncertainty analyses, we performed Monte Carlo simulations of the MM with 200 randomly generated sample MMI sets for the L266P mutation (200 realizations were run for each MMI set). Fig. 2I shows that the distribution of P(EAD) generated from the MM faithfully matches that calculated from the LRM. The total computation time required for all MM simulations was ~17,000 CPU hours whereas the LRM model required only less than 1 CPU second.

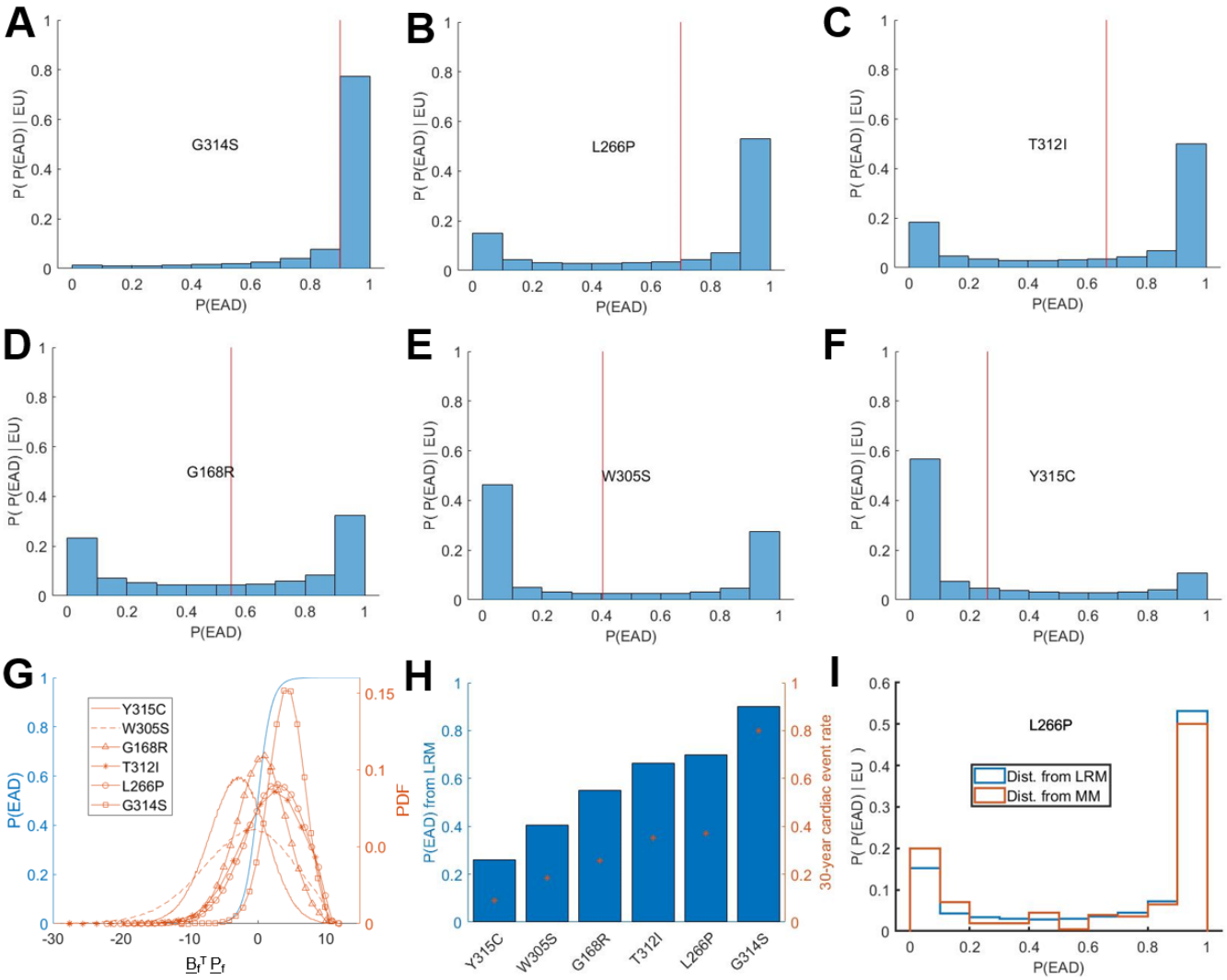

Fig. 2. LQTS1 study. LRM-based estimates of P(EAD) distribution for mutations: (A) G314S, (B) L266P, (C) T312I, (D) G168R, (E) W305S, and (F) Y315C. Red vertical lines in A-F represent the mean P(EAD) of each mutation. P(P(EAD)|EU) on the y-axis, refers to the P(EAD) distribution given mutation-specific experimental uncertainties (EUs) by the Monte Carlo method. (G) PDFs of $\underline{B}_f^T\underline{P}_f$ (red) are shown for each mutation (Y315C: solid line, W305S: dashed line, G168R: triangle markers, T312I: star markers, L266P: circle markers, and G314S: square markers). $\underline{B}_f^T\underline{P}_f$ is the weighted summation of 16 I_{Ks} features (5 MMIs + 11 MMI-derived quadratic features) for each mutation. The MC curve (blue) relates $\underline{B}_f^T\underline{P}_f$ to P(EAD). (H) Mean P(EAD) of each mutation versus the corresponding 30-year cardiac event rate (6). (I) Comparison of P(EAD) distributions for L266P computed using the LRM (blue) and MM (red).

To further delineate the effects of EU in the measured properties of I_{Ks} for each mutation, we evaluated the contribution of each individual MMI to the uncertainty of P(EAD) separately for all mutations. To do this, for each mutation, only a single MMI (i.e. a single member of the MMI set) is allowed to vary while all others are assigned a constant value equal to their experimentally measured mean value, and this process is repeated for each MMI.

In Fig. 3A, PDFs (red curves) are shown for each mutation arising from variations in τ_+_sf based on its EU with all other MMIs held constant. Figs. 3B-E show the same analyses for τ_-_sf, $\Delta V_{1/2}$, k_sf, and G_{Ks}_sf.

In Fig. 3B and D, the widths of the PDFs are quite narrow, indicating that variability in τ_-_sf and k_sf have relatively small impact on P(EAD). In Fig. 3A and C, the widths of the PDFs are wider indicating that uncertainty in τ_+_sf and $\Delta V_{1/2}$ have a greater influence on predicted P(EAD). The uncertainty in P(EAD), however, is most heavily influenced by the EU of G_{Ks}_sf as shown in Fig. 3E. Note that the PDFs in Fig. 2G, where all MMIs vary simultaneously, are nearly identical to those of Fig. 3E, indicating that G_{Ks} uncertainty is the dominant player determining uncertainty in P(EAD) for all mutations.

PDFs for each mutation are quantitatively compared in Table 2, based on their μ and σ. For each mutation, regardless of which MMI is allowed to vary, μ is always close the μ_{all} (μ of the PDF where all 5 MMIs vary based on their EUs). The contribution of MMIs to P(EAD) uncertainty is ranked by computing the average ratio between σ where only one MMI is allowed to vary and σ_{all} where all 5 MMIs are allowed to vary across all mutations. G_{Ks}_sf ($\sigma_{GKs_sf}/\sigma_{all}$ = 0.95 ± 0.03) contributes most, then followed by τ_+_sf ($\sigma_{\tau+_sf}/\sigma_{all}$ = 0.23 ± 0.05), $\Delta V_{1/2}$ ($\sigma_{\Delta V1/2}/\sigma_{all}$ = 0.17 ± 0.06), k_sf ($\sigma_{k_sf}/\sigma_{all}$ = 0.06 ± 0.05), and τ_-_sf ($\sigma_{\tau-_sf}/\sigma_{all}$ = 0.03 ± 0.01). The finding that $\sigma_{GKs_sf}/\sigma_{all}$ = ~1 is consistent with Fig. 3E indicating that G_{Ks} uncertainty is the dominant player determining uncertainty in P(EAD) for all mutations.

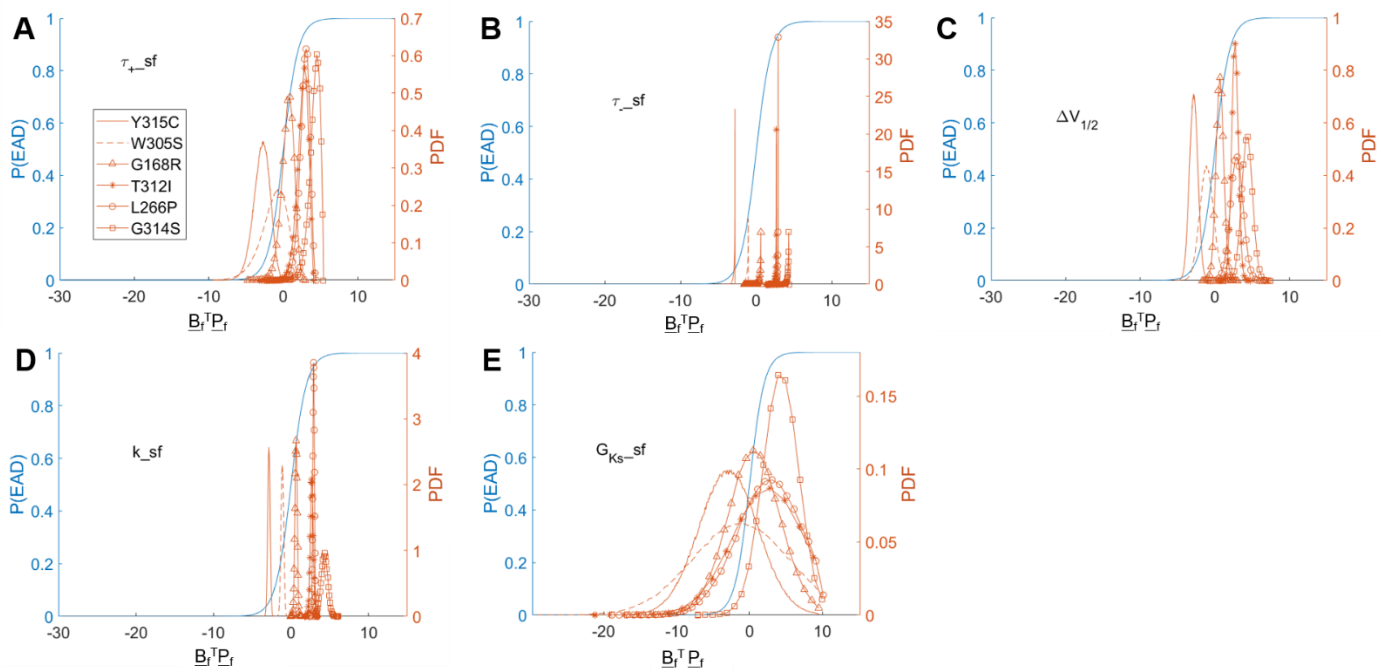

Fig. 3 Uncertainty analysis for individual features underlying LQTS1 mutations. PDFs of $\underline{B}_f^T\underline{P}_f$ (red) are shown for each mutation (Y315C: solid line, W305S: dashed line, G168R: triangle markers, T312I: star markers, L266P: circle markers, and G314S: square markers). $\underline{B}_f^T\underline{P}_f$ is the weighted summation of 16 I_{Ks} features (5 MMIs + 11 MMI-derived quadratic features) for each mutation. The MC curve (blue) relates $\underline{B}_f^T\underline{P}_f$ to P(EAD). Single feature uncertainties: (A) τ_+_sf, (B) τ_-_sf, (C) $\Delta V_{1/2}$, (D) k_sf, or (E) G_{Ks}_sf are sampled from an experimentally based distribution while other MMIs are assumed fixed at their mean.

Table 2. Mean (μ) and standard deviation (σ) of PDFs of $\underline{B}_f^T\underline{P}_f$. Only the indicated MMI is allowed to vary.

	τ_+_sf		τ_-_sf		$\Delta V_{1/2}$	
	$\mu_{\tau+_sf}$	$\sigma_{\tau+_sf}$	$\mu_{\tau-_sf}$	$\sigma_{\tau-_sf}$	$\mu_{\Delta V1/2}$	$\sigma_{\Delta V1/2}$
Y315C	-2.921	1.104	-2.903	0.138	-2.851	0.564
W305S	-1.354	1.744	-1.203	0.220	-1.119	0.923
G168R	0.585	0.834	0.564	0.133	0.657	0.516
T312I	2.583	0.741	2.633	0.075	2.679	0.438
L266P	2.844	0.664	2.863	0.078	2.918	0.847
G314S	4.195	0.712	4.262	0.100	4.305	0.714
	k_sf		G_{Ks}_sf		all 5 MMIs	
	μ_{k_sf}	σ_{k_sf}	μ_{GKs_sf}	σ_{GKs_sf}	μ_{all}	σ_{all}
Y315C	-2.853	0.158	-2.863	4.023	-2.987	4.221
W305S	-1.121	0.175	-1.791	6.058	-2.106	6.449
G168R	0.656	0.151	0.613	3.519	0.444	3.669
T312I	2.676	0.190	1.984	4.338	1.840	4.450

L266P	2.916	0.103	2.380	4.057	2.246	4.224
G314S	4.301	0.415	4.245	2.343	4.090	2.604

Discussion

Regression models have been used previously by Sobie et al. to estimate action potential duration, peak and resting membrane potential, and Ca^{2+} transient amplitude from deterministic myocyte models, as well as the probability of the occurrence of Ca^{2+} sparks from a stochastic Ca^{2+} release site model (10, 11). Based on their method, we developed a new pipeline for establishing simplified models (LRMs) for estimating the probability of cellular cardiac arrhythmia events simulated by a complex stochastic myocyte model. Fig. 1 show good prediction performance of the LRM in predicting the probability of EADs. The method developed in this work is universal in the sense that it can be considered a general approach to simplifying highly complex nonlinear stochastic models in a wide range of applications. To confirm this point, we did another study on the probability of ectopic beats (see Supplement) in which an LRM is developed that maps 4 MMIs (Diastolic cytosolic Ca^{2+} concentration ($[Ca^{2+}]_i$), sarcoplasmic reticulum (SR) Ca^{2+} load ($[Ca^{2+}]_{SR}$), conductance of the inward rectifier current I_{K1} scaling factor(G_{K1}_sf), and ryanodine receptor (RyR) opening rate scaling factor (k_{RyR^+}_sf)) to the probability of delayed afterdepolarization-induced ectopic beats (P(EB)). Fig. S2 show good performance (test set: $R^2 = 0.993$ for the 22 MMI sets within the TD, mean prediction error = 0.006 ± 0.013) of the LRM in predicting the probability of ectopic beats.

Our approach differs from that of Lee at al. (11) at two points in the modeling pipeline. First, we developed a two-iteration MMI set generation strategy. The second iteration ensures that the transition domain (TD) is adequately sampled. Second, we derived LRM quadratic features from the MMIs. The rationale for this approach is that since the MM is highly nonlinear, we expect improved performance by allowing for a nonlinear relationship between MMIs and P(EAD) in the LRM. To demonstrate the performance improvement of these two steps, we built LRMs from only the first iteration MMI sets and using only MMIs as features (i.e. without derived quadratic features). For prediction of P(EAD), this approach yields an average error of 0.030 ± 0.059 (Fig. S5A), which is worse than that (0.019 ± 0.043) of the full pipeline (and S7D). Fig. S6A and Fig. S6D shows the same result for prediction of P(EB).

The LRM approach has the benefit of computational efficiency. The average computational time required for each MMI set in the EAD study is ~0.4 (realizations per hour) * 100 (realizations) = 40 CPU hours. The computational time required to run the test set (100 MMI sets) in the EAD study is therefore 4000 CPU hours. In contrast, the computational time for LRMs is negligible (~5 CPU ms). To build the LRM, however, 200 training MMI sets must be simulated with the MM, which takes ~8000 CPU hours. For the uncertainty analysis, in Fig. 3I, the L226P P(EAD) distribution using MM (200 MMI sets and 200 realizations for each set) costs ~16,000 CPU hours (C++) and the LRM costs ~0.5 CPU second (MATLAB R2018b), which again shows the dramatic improvement in computational efficiency of the LRM.

LRMs enable us to quantitatively explore the relationship between MMIs and event probabilities. In the EAD study, G_{Ks} is the most impactful parameter of I_{Ks} with regard to P(EAD) because in the ranking of feature importance, it is associated with the first, second, and fourth features in Table 1. In Fig. 3, we also showed that, for 6 LQTS1 mutations, G_{Ks} experimental uncertainty is the dominant contributor to P(EAD) uncertainty. Fig. S7 shows that it is negatively correlated with P(EAD). This result follows from the fact that I_{Ks} is a repolarizing current. With the decrease of conductance of I_{Ks}, the action potential becomes more prone to depolarizing events, thus leading to higher likelihood of EAD occurrence (27). In the ectopic beat study, Fig. S3A-D shows the relationship between each MMI and P(EB) from the LRM. These findings agree with published experiments and simulations (see details in Supplement).

Action potential duration (APD) and QT prolongation have been widely used as surrogates, both in basic research and in clinical settings, as indicators of disease (12, 13) and drug-related arrhythmogenic effects (14, 15). However, there are examples which have shown the dissociation of these two surrogates with arrhythmia or arrhythmogenic

effects. For example, alfuzosin prolongs APD and QT interval but is not linked to occurrence of Torsades de Pointes, a polymorphic ventricular tachycardia (28). Ranolazine prolongs QT interval but has been shown to be antiarrhythmic (18). Unlike these indirect surrogates, EAD is a fundamental cellular arrhythmia event. The stochastic myocyte model, in principle, can be used to estimate P(EAD) but doing so is unrealistic because of its high computational burden. Such estimates become possible with our LRM. We successfully utilized P(EAD) to correctly rank the severity of 6 LQTS1 mutations out of 17 mutations (based on experiments of Jons et al. (6)) based on the severity determined by clinical outcome (Fig. 3A). Predictions failed to match clinical results for the remaining 11 mutations, potentially due to the following reasons: 1) The stochastic myocyte model is a canine midmyocardial ventricular myocyte model, which is quantitatively distinct from human myocytes in terms of some essential ion channel properties such as the Ca^{2+}-independent transient outward K^+ current (I_{to1}) (29). 2) Tissue-level networking effects are ignored. Xie et. al. used computational models to show that gap junction conductance is essential for EAD and DAD propagation inside the heart (30). 3) The mathematical characterization for LQTS1 mutations may be insufficient. In our study, I_{Ks} activation and deactivation processes are characterized by a single time constant (see Eq. 1) whereas Eldstrom et al. used two activation time constants and three inactivation time constants to characterize four LQTS1 mutations (31). 4) The patient pool for some mutations is relatively small such that the derived ranking of clinical outcomes may be sensitive to sample bias. For seven of these mutations, the size of the patient cohort was less than 16 (6).

In Fig. 2B-F, a bimodal shape of the P(EAD) distribution is observed when analyzing the effect of 5 LQTS1 mutations (Y315C, W305S, G168R, T312I, and L266P) on P(EAD), which is explained by Fig. 2G. To understand the occurrence of the bimodal shape, a more exhaustive study on the P(EB) distribution shape, which can be generalized to that of the P(EAD) distribution, was performed (Fig. S4). By manually varying the $[Ca^{2+}]_{SR}$ uncertainty, three patterns of P(EB) distributions were observed: unimodal, approximately uniform, and bimodal. This demonstrates that the P(EB) distribution pattern depends on the width of the PDF of the MMI (or PDFs of $\underline{B}_f^T\underline{P}_f$ for multiple MMIs) as well as the relative position of the mean MMI (or $\underline{B}_f^T\underline{P}_f$) with respect to the TD of its MC curve. Further details are provided in the Supplement.

Given the bimodal shape of the P(EAD) distribution for the 5 mutations studied here, small perturbations of the 5 I_{Ks}-related parameters would result in model cells that either exhibit normal APs exclusively, or APs with EADs exclusively. This is consistent with the fact that EADs are a bifurcation phenomenon in the stochastic myocyte model. Other studies also show that EADs arise as a result of a bifurcation (32, 33). In Fig. S4C, results showed that when $[Ca^{2+}]_{SR}$ is larger than 30μM, the P(EB) distribution becomes bimodal. Given the fact that reported EUs (standard errors) of $[Ca^{2+}]_{SR}$ vary widely from 10μM to 290μM (34-36), the phenomenon of EBs is also a bifurcation behavior (37).

In Fig. 3 and Table 2, we evaluated the contribution of each individual MMI to the uncertainty of P(EAD) for six LQTS1 mutations. An average value of $\sigma_{GKs_sf}/\sigma_{all}$ = ~1 (Table 2) indicates that G_{Ks}_sf uncertainty alone provides enough information to approximately reconstruct the P(EAD) distribution. For each mutation, the PDF of $\underline{B}_f^T\underline{P}_f$ in Fig. 3E arising from variability in G_{Ks}_sf alone nearly matches the PDF in Fig. 2G, confirming this idea. This result indicates that for future LQTS1 studies, I_{Ks} expression variability is more impactful than variations in its kinetic or voltage-dependent properties with respect to the uncertianty of P(EAD). A similar conclusion has been made in LQTS type 2 (LQTS2). Anderson et al. showed that protein trafficking (which ultimately impacts cell-surface expression), rather than gating or permeation (intrinsic functional properties of the channel), is the dominant mechanism for loss of Kv11.1 channel function in LQTS2 (38).

In summary, we developed a pipeline for building logistic regression models for the probability of cellular arrhythmias. These simplified models faithfully reproduce predictions of biophysically-detailed stochastic myocyte models with far less computational burden. As far as we know, this is the first simplified model enabling quantitative investigation of determinants of the probability of cellular arrhythmias (EAD and EB). In the LQTS1 study, we showed that ranking of P(EAD) is qualitatively consistent with the ranking of 6 LQTS1 mutation severities according to their 30-year cardiac event rate. We then performed the study of LQTS1 experimental parameter uncertainty

analysis. We showed that the EU of G_{Ks} is the dominant contributor to P(EAD) uncertainty.

Materials and Methods

Modified Walker Ventricular Myocyte Model: This study uses a modified version of the Walker et al. stochastic ventricular myocyte model (7). The Walker et al. model is a biophysical-detailed three-dimensional spatial stochastic model of a cardiac ventricular myocyte containing 25,000 Ca^{2+} release units (CRUs) (each representing a dyad) along with their sub-membrane compartments. A subset of CRUs (5400) was used for all simulations to accelerate the computation and all CRU Ca^{2+} fluxes were scaled to the total number of CRUs. This model has been shown to reproduce realistic Ca^{2+} waves as well as DADs. Chu et. al. (39) recently developed a model of the Na^{+}-Ca^{2+} exchanger (NCX) that includes allosteric regulation of NCX function by Ca^{2+}. This NCX model was incorporated into the Walker et al. model. Following Chu et. al. (39), NCX was localized 15% in dyad, 30% in the submembrane compartment, and 55% in the cytosol. In addition to these changes, NCX conductance was increased 20%, the time constant of Ca^{2+} flux from submembrane to cytosol was reduced by 20%, I_{Kr} conductance increased by 20%, SERCA pump rate decreased by 25%, and number of I_{to2} channels reduced from 8 to 5 in each submembrane compartment to assure a more realistic canine APD (~290 ms) and Ca^{2+} transient peak (~800 nM), similar to that of the normal canine ventricular myocyte (40). β-adrenergic stimulation is simulated to facilitate generation of cellular arrhythmias. This was done using the β-adrenergic stimulation protocol described in Walker et. al. (7).

Long QT syndrome type 1 experimental data and I_{Ks} model.

LQTS1 clinical and experimental mutation data are obtained from Jons et. al. 2011 (6). The 30-year cardiac event rate is extracted from their Fig. 2. The 30-year cardiac event rate represents the percentage of mutation carriers that had their first cardiac event in their first 30 years of life. The focused cardiac events comprise syncope, aborted cardiac arrest, and sudden cardiac death. Clinical and experimental data from their study is summarized in Table S1. Five I_{Ks}-related parameters for each mutation are extracted from their Tables 2 and S1. In order to incorporate these five I_{Ks}-related parameters into our myocyte model, we also incorporated their I_{Ks} model into our myocyte as model described by the following equations

$$
\begin{aligned}
&I_{Ks} = G_{Ks} \bullet G_{Ks_sf} \bullet \chi_{Ks} \bullet (V - E_{Ks}) \\
&\frac{d\chi_{Ks}(t)}{dt} = \frac{\chi_{Ks}(\infty) - \chi_{Ks}(t)}{\tau_{xKs}} \\
&\chi_{Ks}(\infty) = \frac{1}{1 + e^{-\frac{V - 9.4 + \Delta V_{1/2}}{11.8 \bullet k_sf}}} \\
&a_1 = 0.0000719 \frac{V}{1 + e^{-0.74V}} \bullet \frac{1}{\tau_{+}_sf} \\
&a_2 = 0.000786 \frac{V + 40}{e^{0.103(V+40)} - 1} \bullet \frac{1}{\tau_{-}_sf} \\
&\tau_{xKs} = \frac{1}{a_1 + a_2}
\end{aligned}
\qquad (1)
$$

, where τ_{+}_sf, τ_{-}_sf, $\Delta V_{1/2}$, k_sf, and G_{Ks}_sf are mutation-specific terms.

We selected 6 well-characterized mutations (Y315C, W305S, G168R, T312I, L266P, and G314S) for study covering low clinical risk to high clinical risk based on 30-year cardiac event rate.

Simulation Protocol for early afterdepolarization modeling: In the EAD studies, initial states of model state variables were the same in each simulation. These values were obtained in diastole after 10 beats at 1 Hz pacing during the β-adrenergic stimulation protocol. Each model simulation had a duration of 600ms. At 10ms, a 1-ms 10000 pA inward current was injected to trigger an AP.

Statistical Modeling. Fig. 4 illustrates the general workflow for building a logistic regression model (LRM) from an underlying biophysically-detailed myocyte model. A more detailed workflow (Fig. S1) and its mathematical development is provided in the supplement. The steps of the workflow in Fig. 4 are performed twice to build the

LRM. In the first iteration, relevant myocyte model inputs (MMIs), which refers collectively to myocyte model parameters as well as state variable initial conditions, and their ranges of variation (regions of interest, or ROI) are identified. ROI is established to keep MMIs within biophysically relevant ranges. Multiple randomly selected MMI sets are generated by uniform sampling within these ROIs. For each sampled MMI set, multiple simulations (realizations) are performed using the Modified Walker model. A binary indicator random variable is defined using the output of each simulation. This random variable indicates whether the cellular arrhythmia event of interest did or did not occur. For each MMI set, an event (occurrence of an EAD) probability is estimated based on its frequency of occurrence over all realizations. All MMI sets and their corresponding event probabilities are then used to derive a logistic regression model relating the probability of the event to a weighted sum of MMIs:

$$P(\text{event})=\frac{1}{1+\exp(-\underline{B}^{T}\underline{P})} \qquad (2)$$

, where $\underline{P} = \underline{P_o}$ is a vector of features, $\underline{B} = \underline{B_o}$ is a vector of weights. The subscript "*o*" stands for "initial", representing the first iteration of the pipeline. P(event) represents the estimated event occurrence probability.

After this first iteration, the logistic equation (1.2) is then used as an additional constraint to estimate the transition domain (TD), which is a subregion within the ROI for the MMI set. The second iteration then samples only within the TD, and steps 4 – 5 of the workflow are repeated. MMI sets and corresponding event probabilities obtained from both iterations are combined, and the final LRM is built again using equation 1.2, where $\underline{P} = \underline{P_f}$ is the final feature vector and $\underline{B} = \underline{B_f}$ is the final weights vector. The subscript "*f*" stands for "final", representing the second iteration of the pipeline. We refer to this model as the LRM. To ensure the generalizability of the LRM, an additional 100 test MMI sets are generated within the ROI to be used for evaluation of the performance of the LRM. 100 MMI sets are generated in both the first and second iterations. For each MMI set, 110 realizations (for EAD study) are simulated to estimate the probability of the event of interest. All logistic regression model fits were performed with MATLAB R2018b *fitglm* function.

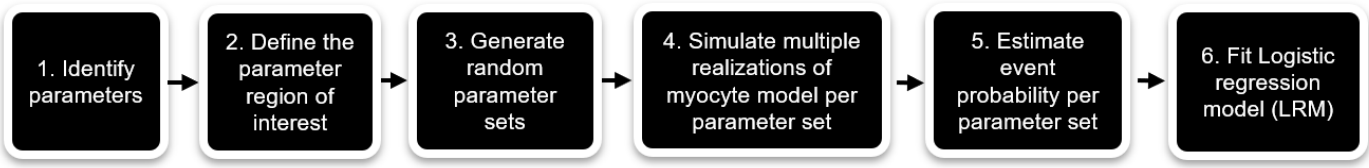

Fig. 4. General workflow pipeline for logistic model formulation

MMI Identification and MMI ROI: When building the LRM for predicting probability of occurrence of EADs in the setting of LQTS1, 5 MMIs (I_{Ks}-related parameters) were selected to be varied (Table S3). These MMIs were varied over an experimentally determined range of values (± 2 standard errors) based on the work of Jons et. al.(6). Those 5 MMIs are activation time constant scaling factor (τ_+_sf), deactivation time constant scaling factor (τ_-_sf), shift in the half maximal activation voltage ($\Delta V_{1/2}$), scaling factor of voltage-dependent activation curve slope (k_sf), and maximal conductance scaling factor (G_{Ks}_sf).

The ROI can be separated into two mutually exclusive subspaces in terms of the arrhythmia event probability: plateau domain (PD) and transition domain (TD). The PD is the space where MMI sets simulated in the MM yield P(event) strictly equal to 0 or 1. The TD is the space where MMI sets simulated in the MM yield P(event) strictly > 0 and < 1. We can specify the Left PD (LPD) as corresponding to P(event) = 0 and the Right PD (RPD) as corresponding to P(event) = 1 since these are also mutually exclusive. Because of the limited number of realizations performed with the MM for each MMI set, P(event) takes on a discrete set of values such that we can strictly follow this definition.

LRM Validation Metrics. Linear regression was performed between MM-generated actual P(event)s and LRM predicted P(event)s. The R-squared (R^2) from the linear regression was calculated. Absolute error between those

was also calculated as

$$Error = \frac{\sum_{i=1}^{n} | P_i(event)_{MM} - P_i(event)_{LRM} |}{n} \quad (3)$$

,where n is the number of MMI sets, $P_i(event)_{MM}$ is event probability of ith MMI set obtained from the MM, and $P_i(event)_{LRM}$ is event probability of ith MMI set obtained from the LRM.

Uncertainty Analyses: The LRM derived in this study is a function that maps MMIs to P(event). If experimental measurements of these MMIs are variable, then this variation can be modeled by assuming these measurements are drawn from an underlying probability distribution. In this case, the event probability produced by the LRM is also a random variable (RV). We wish to characterize this uncertainty by computing the distribution of the estimated event probability. To do this, when the mean and variance of MMI estimates are available from experimental data, we assume MMIs are RVs drawn from a normal distribution with that mean and variance. To ensure the values are limited to be within a biophysically meaningful range for each MMI, distributions were truncated to exclude values outside this range. In cases where the argument of the LRM ($\underline{B}_f^T\underline{P}_f$) involves a weighted sum of individual MMIs not including quadratic features, this distribution can be calculated analytically, and is the logit-normal distribution (Eq. S5) (41). In cases where the feature vector $\underline{P}_f$ includes elements containing quadratic features, P(event) has to be solved numerically. In this latter case, we randomly generate (10^6) sample vectors for a specific MMI set based on the underlying MMI uncertainty distributions and use the LRM to predict P(event) for all sample vectors. Following this approach, the P(event) distribution can be estimated.

References

1. P. F. Cranefield, R. S. Aronson, Torsade de pointes and other pause-induced ventricular tachycardias: the short-long-short sequence and early afterdepolarizations. *Pacing Clin Electrophysiol* **11**, 670-678 (1988).
2. Z. Qu, J. N. Weiss, Mechanisms of ventricular arrhythmias: from molecular fluctuations to electrical turbulence. *Annu Rev Physiol* **77**, 29-55 (2015).
3. M. R. Rosen, Cellular electrophysiology of digitalis toxicity. *J Am Coll Cardiol* **5**, 22A-34A (1985).
4. R. L. Winslow *et al.*, Integrative modeling of the cardiac ventricular myocyte. *Wiley Interdiscip Rev Syst Biol Med* **3**, 392-413 (2011).
5. M. Fink, P. J. Noble, D. Noble, Ca(2)(+)-induced delayed afterdepolarizations are triggered by dyadic subspace Ca2(2)(+) affirming that increasing SERCA reduces aftercontractions. *Am J Physiol Heart Circ Physiol* **301**, H921-935 (2011).
6. C. Jons *et al.*, Use of mutant-specific ion channel characteristics for risk stratification of long QT syndrome patients. *Sci Transl Med* **3**, 76ra28 (2011).
7. M. A. Walker, V. Gurev, J. J. Rice, J. L. Greenstein, R. L. Winslow, Estimating the probabilities of rare arrhythmic events in multiscale computational models of cardiac cells and tissue. *PLoS Comput Biol* **13**, e1005783 (2017).
8. E. D. Fowler *et al.*, Arrhythmogenic late Ca2+ sparks in failing heart cells and their control by action potential configuration. *Proceedings of the National Academy of Sciences* **117**, 2687-2692 (2020).
9. A. J. Tanskanen, J. L. Greenstein, B. O'Rourke, R. L. Winslow, The role of stochastic and modal gating of cardiac L-type Ca2+ channels on early after-depolarizations. *Biophysical Journal* **88**, 85-95 (2005).
10. E. A. Sobie, Parameter sensitivity analysis in electrophysiological models using multivariable regression. *Biophys J* **96**, 1264-1274 (2009).
11. Y. S. Lee, O. Z. Liu, H. S. Hwang, B. C. Knollmann, E. A. Sobie, Parameter sensitivity analysis of stochastic models provides insights into cardiac calcium sparks. *Biophys J* **104**, 1142-1150 (2013).
12. P. J. Schwartz, L. Crotti, R. Insolia, Long-QT syndrome: from genetics to management. *Circulation: Arrhythmia and Electrophysiology* **5**, 868-877 (2012).
13. R. Hoefen *et al.*, In silico cardiac risk assessment in patients with long QT syndrome: type 1: clinical predictability of cardiac models. *Journal of the American College of Cardiology* **60**, 2182-2191 (2012).
14. I. H. T. Guideline, The non-clinical evaluation of the potential for delayed ventricular repolarization (Qt Interval

Prolongation) by human pharmaceuticals. *S7B (http://www. ich. org/products/guidelines/safety/article/safety-guidelines. html)* (2005).
15. K. Blinova *et al.*, International multisite study of human-induced pluripotent stem cell-derived cardiomyocytes for drug proarrhythmic potential assessment. *Cell reports* **24**, 3582-3592 (2018).
16. M. J. Koren, M. R. Crager, M. Sweeney, Long-term safety of a novel antianginal agent in patients with severe chronic stable angina: the Ranolazine Open Label Experience (ROLE). *Journal of the American College of Cardiology* **49**, 1027-1034 (2007).
17. B. R. Chaitman *et al.*, Effects of ranolazine with atenolol, amlodipine, or diltiazem on exercise tolerance and angina frequency in patients with severe chronic angina: a randomized controlled trial. *Jama* **291**, 309-316 (2004).
18. S. Tzeis, G. Andrikopoulos, Antiarrhythmic properties of ranolazine–from bench to bedside. *Expert opinion on investigational drugs* **21**, 1733-1741 (2012).
19. Q. Wang *et al.*, Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nat Genet* **12**, 17-23 (1996).
20. W. Shimizu, C. Antzelevitch, Cellular basis for the ECG features of the LQT1 form of the long-QT syndrome: effects of β-adrenergic agonists and antagonists and sodium channel blockers on transmural dispersion of repolarization and torsade de pointes. *Circulation* **98**, 2314-2322 (1998).
21. I. Mikuni, C. G. Torres, M. W. Bienengraeber, W.-M. Kwok, Partial restoration of the long QT syndrome associated KCNQ1 A341V mutant by the KCNE1 β-subunit. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1810**, 1285-1293 (2011).
22. JCGM (2008) Joint Committee for Guides in Metrology. (Bureau International des Poids et Mesures Sèvres, France).
23. S. Marino, I. B. Hogue, C. J. Ray, D. E. Kirschner, A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology* **254**, 178-196 (2008).
24. R. H. Johnstone *et al.*, Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models? *Journal of molecular and cellular cardiology* **96**, 49-62 (2016).
25. O. J. Britton *et al.*, Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology. *Proceedings of the National Academy of Sciences* **110**, E2098-E2105 (2013).
26. P. J. Schwartz *et al.*, Genotype-phenotype correlation in the long-QT syndrome: gene-specific triggers for life-threatening arrhythmias. *Circulation* **103**, 89-95 (2001).
27. D. Bers, *Excitation-contraction coupling and cardiac contractile force* (Springer Science & Business Media, 2001), vol. 237.
28. A. E. Lacerda *et al.*, Alfuzosin delays cardiac repolarization by a novel mechanism. *Journal of Pharmacology and Experimental Therapeutics* **324**, 427-433 (2008).
29. F. G. Akar *et al.*, Phenotypic differences in transient outward K+ current of human and canine ventricular myocytes: insights into molecular composition of ventricular Ito. *Am J Physiol Heart Circ Physiol* **286**, H602-609 (2004).
30. Y. Xie, D. Sato, A. Garfinkel, Z. Qu, J. N. Weiss, So little source, so much sink: requirements for afterdepolarizations to propagate in tissue. *Biophysical journal* **99**, 1408-1415 (2010).
31. J. Eldstrom, Z. Wang, D. Werry, N. Wong, D. Fedida, Microscopic mechanisms for long QT syndrome type 1 revealed by single-channel analysis of IKs with S3 domain mutations in KCNQ1. *Heart Rhythm* **12**, 386-394 (2015).
32. D. X. Tran *et al.*, Bifurcation and chaos in a model of cardiac early afterdepolarizations. *Phys Rev Lett* **102**, 258103 (2009).
33. P. Kugler, Early Afterdepolarizations with Growing Amplitudes via Delayed Subcritical Hopf Bifurcations and Unstable Manifolds of Saddle Foci in Cardiac Action Potential Dynamics. *PLoS One* **11**, e0151178 (2016).
34. E. Bovo, S. R. Mazurek, L. A. Blatter, A. V. Zima, Regulation of sarcoplasmic reticulum Ca2+ leak by cytosolic Ca2+ in rabbit ventricular myocytes. *The Journal of physiology* **589**, 6039-6050 (2011).
35. T. R. Shannon, K. S. Ginsburg, D. M. Bers, Quantitative assessment of the SR Ca2+ leak-load relationship. *Circulation research* **91**, 594-600 (2002).
36. A. V. Zima, E. Bovo, D. M. Bers, L. A. Blatter, Ca2+ spark-dependent and-independent sarcoplasmic reticulum Ca2+ leak in normal and failing rabbit ventricular myocytes. *The Journal of physiology* **588**, 4743-4757 (2010).
37. T. R. Chay, Y. S. Lee, Studies on re-entrant arrhythmias and ectopic beats in excitable tissues by bifurcation analyses.

J Theor Biol **155**, 137-171 (1992).

38. C. L. Anderson *et al.*, Most LQT2 mutations reduce Kv11.1 (hERG) current by a class 2 (trafficking-deficient) mechanism. *Circulation* **113**, 365-373 (2006).
39. L. Chu, J. L. Greenstein, R. L. Winslow, Modeling Na(+)-Ca(2+) exchange in the heart: Allosteric activation, spatial localization, sparks and excitation-contraction coupling. *J Mol Cell Cardiol* **99**, 174-187 (2016).
40. B. O'Rourke *et al.*, Mechanisms of altered excitation-contraction coupling in canine tachycardia-induced heart failure, I: experimental studies. *Circ Res* **84**, 562-570 (1999).
41. J. Atchison, S. M. Shen, Logistic-normal distributions: Some properties and uses. *Biometrika* **67**, 261-272 (1980).