5-2014

Personalized anticoagulant management using reinforcement learning.

Michael Jacobs
University of Louisville

Personalized Anticoagulant Management Using Reinforcement Learning

By

Michael Jacobs
Bachelor of Science Bioengineering, University of Louisville, May 2012

A Thesis

Submitted to the Faculty of the
University of Louisville
J.B. Speed School of Engineering
As Partial Fulfillment of the Requirements
for the Professional Degree

MASTER OF ENGINEERING

Department of Bioengineering

May 2014

Personalized Anticoagulant Management Using Reinforcement Learning

Submitted by:_____
Michael Jacobs

A Thesis Approved On

_____

(DATE)

By the Following Reading and Examination Committee:

_____
Hermann Frieboes Ph.D., Thesis Director

_____
Robert Keynton Ph.D.

_____
Tamer Inanc Ph.D.

_____
Adam E. Gaweda Ph.D.

_____
Michael E. Brier Ph.D.

ACKNOWLEDGEMENTS

Thank you to all those who have helped me reach this point in my academic career. I would like to specifically thank Dr. Adam Gaweda and Dr. Michael Brier. Without their support and guidance over the last few years, I would never have been able to accomplish all that I have. I'd also like to thank my parents for their continued support throughout my life and academic career, and also for helping me to remain focused on the things that truly matter.

ABSTRACT

Introduction: There are many problems with current state-of-the-art protocols for maintenance dosing of the oral anticoagulant agent warfarin used in clinical practice. The two key challenges include lack of personalized dose adjustment and the high cost of monitoring the efficacy of the therapy in the form of International Normalized Ratio (INR) measurements. A new dosing algorithm based on the principles of Reinforcement Learning (RL), specifically Q-Learning with functional policy approximation, was created to personalize maintenance dosing of warfarin based on observed INR and to optimize the length of time between INR measurements. This new method will help improve patient's INR time in therapeutic range (TTR) as well as minimize cost associated with monitoring INR when compared to the current standard of care.

Procedure: Using the principles of Reinforcement Learning, an algorithm to control warfarin dosing was created. The algorithm uses 9 different controllers which correspond to 9 different levels of warfarin sensitivity. The algorithm switches between controllers until it selects the controller that most closely resembles the individual patient's response, and thus the optimal dose change ($\Delta\mathrm{Dose}$) and time between INR measurements **(**$\Delta\mathrm{Time}$) are personalized for each patient, based on INR observed in the patient. Three simulations were performed using data from 100 artificial patients,

generated based on data from real patients, each. The first simulation that was performed was an ideal case scenario (clean simulation where the coefficient of variance (CV) of noise added to the model output = 0) using only the warfarin RL algorithm to prove efficacy. The second simulation was performed using the current standard of care and a CV = 25% to simulate intra-patient variability. The third simulation was performed using the warfarin RL algorithm with a CV = 25%. 180 days were simulated for each patient in each simulation and the measurements that were used to benchmark the efficacy of the therapy were INR time in therapeutic range (TTR) and the number of INR measurements that were taken during simulation.

Results: The first simulation yielded a mean TTR = 92.1% with a standard deviation of 4.2%, and had a mean number of INR measurements = 7.94 measurements/patient. The second simulation yielded a mean TTR = 45.3% with a standard deviation of 16.4%, and had a mean number of INR measurements = 12.3 measurements/patient. The third simulation yielded a mean TTR = 51.8% with a standard deviation of 10.8%, and had a mean number of INR measurements = 8.05 measurements/patient. A p-value <.001 suggests that there is a statistically significant difference between the 2 algorithms.

Conclusion: Results from the simulations indicate that the warfarin RL algorithm performed better than the standard of care at keeping the patient's INR in therapeutic range and also reduced the number of INR measurements that were necessary. This algorithm could help improve patient safety by increasing the patient's INR TTR in the presence of intra-patient variability, and also help reduce the heavy cost associated with the therapy by minimizing the number of INR measurements that are necessary.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. Problem Statement

There are a variety of different disease states and conditions where the use of prophylaxis is recommended to reduce the risk of thromboembolism in patients. These disease states and conditions include, but are not limited to, atrial fibrillation, heart valve replacements, deep vein thrombosis (DVT), and myocardial infarction (MI) (Merli & Tzanis, 2009). Common prophylaxes that are used in clinical practice include mechanical methods, such as compression sleeves, and pharmaceutical methods, such as anticoagulants.

Mechanical methods used to reduce the risk of thromboembolism include a variety of different types of compression sleeves. Compression sleeves are used to apply pressure to areas of poor circulation, thereby reducing blood stasis (Larkin, Mitchell, & Petrie, 2012). There are many different types of compression sleeves used in clinical practice including uniform compression sleeves, graduated compression sleeves, and intermittent pneumatic sleeves. Uniform compression sleeves apply uniform pressure to the area that they are applied and are readily available to the entire population, whereas graduated compression sleeves vary the pressure they apply throughout the sleeve and are typically used in hospital settings (Larkin, Mitchell, & Petrie, 2012). Intermittent pneumatic sleeves use pressure cuffs to repeatedly inflate and deflate around the area they are applied. These sleeves can vary the amount of pressure they apply and can also be used to apply uniform or graduated pressure. While compression sleeves have proven to be effective in reducing thromboembolism in conditions such as DVT and surgery

(Morris & Woodcock, 2004; MacLellan & Fletcher, 2007; Larkin, Mitchell, & Petrie, 2012), there is no evidence to suggest that they could be effective in reducing thromboembolic events in conditions such as atrial fibrillation and MI.

Pharmaceutical methods for reducing the risk of thromboembolic event are typically anticoagulant drugs including injectable drugs such as heparin and oral anticoagulant drugs such as dabigatran, rivaroxaban, and warfarin. Heparin, specifically low molecular weight heparin, binds to antithrombin III which inactivates thrombin and factor Xa (Chuang, Swanson, Raja, & Olson, 2001). Dabigatran is a direct thrombin inhibitor (Miller, Grandi, Shimony, Filion, & Eisenberg, 2012). Rivaroxaban inhibits both free factor Xa and factor Xa (Miller, Grandi, Shimony, Filion, & Eisenberg, 2012). Warfarin is a vitamin K antagonist and works by inhibiting the synthesis of the Vitamin K dependent clotting factors II, VII, IX, and X (Porter, 2010). While all of these drugs have proven to be effective in clinical practice, there are drawbacks associated with each. Injectable heparin can cause patient discomfort because of the need for injections, and even when it is taken orally, it still has a higher monetary cost when compared to other oral anti-coagulants (Looi, et al., 2013). Due to the nature of these anticoagulant drugs, there is an increased risk of patient bleeding, and because of this, there is a need to take precautions while using anticoagulants for therapy. Warfarin has an easy reversibility of action when compared to other oral anticoagulants, such as dabigatran and rivaroxaban, and due to this fact, it remains the most widely used oral anticoagulant in clinical practice today (The International Warfarin Pharmacogenetics Consortium, 2009).

The standard for measuring the efficacy of warfarin therapy, first adopted in 1982 by the World Health Organization, is known as the International Normalized Ratio (INR)

(Wardrop & Keeling, 2008). Many of the indications for use of warfarin therapy specify a narrow therapeutic range of 2.0-3.0, and because of this there is a need for frequent INR measurements to minimize the risk of bleeding (when INR is too high), or thromboembolic event (when INR is too low) (Merli & Tzanis, 2009). The high cost associated with warfarin therapy comes from the need for frequent INR measurements. These costs include not only the health care expenses (such as laboratory tests, equipment, labor, etc.), but also indirect costs such as time lost from work, travel expenses, and many others (Chambers, Chadda, & Plumb, 2009; Harrington, Armstrong, Nolan, & Malone, 2013; Lafata, Martin, Kaatz, & Ward, 2000).

Because of the widely adopted use of warfarin oral anticoagulant therapy, there is a need for dosing algorithms to maintain the efficacy of the therapy while reducing the risk for bleeding or thromboembolic events. The current standard of care for warfarin oral anticoagulant therapy, as dictated by the American Society of Hematology, is an expert-system type algorithm that provides no dosing personalization and also does not explicitly optimize the monitoring frequency of the efficacy of the therapy (Cushman, Lim, & Zakai, 2011). While there are many other warfarin dosing algorithms that seek to improve the efficacy of warfarin therapy, including pharmacogenetic algorithms (Carlquist & Anderson, 2011) and computerized algorithms (Grzymala-Lubanski, Själander, Renlund, Svensson, & Själander, 2013; Dimberg, et al., 2012), these algorithms do a poor job of accounting for intra-patient variability (Kangelaris, Bent, Nussbaum, Garcia, & Tice, 2009) and do not explicitly optimize the monitoring frequency, which would reduce the overall cost of the therapy. Intra-patient variability during maintenance dosing can occur due to a variety of factors such as diet, disease

state, and drug interactions. (Ansell, et al., 2008; White, 2010). In fact, there are currently no warfarin dosing algorithms that optimize both the warfarin dose for an individual patient as well as the time between INR monitoring visits.

B. Current Warfarin Dosing Methods

There are three main types of algorithms that are used in clinical practice to manage warfarin oral anticoagulant therapy. These three main types of algorithms include dose titration, pharmacogenetic algorithms, and computerized algorithms. Warfarin dose titration is exemplified in the dosing protocol dictated by the American Society of Hematology, which is the current standard of care (Cushman, Lim, & Zakai, 2011). Dose titration algorithms slowly titrate a patient's warfarin dose until the patient's INR levels are within the therapeutic range, and the method of titrating a patient's dose until the desired effect is achieved is common practice in drug dosing even outside of the realm of anticoagulant therapy. This can be ineffective and slow to respond in the presence of intra-patient variability, resulting patient's INR values being outside of the therapeutic range (Wilson, Costantini, & Crowther, 2007).

Pharmacogenetic algorithms use pharmacogenetic information, specifically, variations in the genes CYP2C9 and VKORC1, to select a more accurate initial warfarin dose (Carlquist & Anderson, 2011). While pharmacogenetic algorithms have been proven in clinical practice to minimize the effect of inter-patient variability and select a more accurate initial warfarin dose (Carlquist & Anderson, 2011), they do nothing to account for intra-patient variability due to external factors, which is unhelpful during a patient's maintenance dosing period, and also are yet to gain wide acceptance among

physicians. The required pharmacogenetic testing to determine a patient's genetic variations are also costly and not part of standard clinical practice.

Computerized algorithms, exemplified by AuriculA (a Swedish national quality registry of patients treated with warfarin), use key patient characteristics, and information about the warfarin treatment and complications to make dose suggestions (Dimberg, et al., 2012). These algorithms operate according to 720 rules and patient history to make dose suggestions (Grzymala-Lubanski, Själander, Renlund, Svensson, & Själander, 2013). While these algorithms have been successful in clinical practice, they require massive databases of patient information, and also, in the presence of high intra-patient variability, still require manual (physician initiated) dose changes. Another issue with all of the algorithms in clinical practice are that there are no dosing algorithms that optimize INR measurement and dose change frequency.

C. Objective

The objective of this study was to develop a new method for dosing warfarin based on the control technique of Reinforcement Learning (RL) that will adapt to each patient based on feedback from the patient, and will also optimize the time between INR measurements. This new algorithm will help minimize the effect of intra-patient variability and reduce the number of INR measurements that are necessary. Because the current standard of care does a poor job of accounting for intra-patient variability and does nothing to optimize the time between INR measurements, the ultimate goal of this work is for the new warfarin RL algorithm to increase patient's time in the therapeutic

range (TTR) and reduce the overall cost of therapy by optimizing the INR monitoring frequency when compared to the current standard of care.

II. INSTRUMENTATION AND EQUIPMENT

The equipment that was used for the creation of the warfarin RL algorithm was a Lenovo ThinkPad Edge laptop with model number 0301-DBU. The laptop was manufactured by Lenovo (Singapore) Pte. Ltd., and it was made in China. All coding, calculations, and graphs were done using MATLAB 7.12.0 R2011a software created by The MathWorks Inc., located in Natick, MA.

III. PROCEDURE

A. Reinforcement Learning Overview

Reinforcement Learning (RL) is a control method used to control a system based on experience. Elements of Reinforcement Learning include: Environment, which is the system that is being affected, State, which is a measurement of the environment, Action, which is the control input into the system, Agent, which is the governing body that takes the action, and Reward, which identifies how well the agent is performing. This method utilizes the Markov Decision Process (MDP) to determine the optimal action to take, while in a given state, to achieve a desired state (Sutton & Barto, 1998). The rules governing what action to take, while in a given state, to achieve a desired state are known as a policy, and the goal of Reinforcement Learning is to determine the optimal policy.

One of the most popular RL methods is Q-Learning. Q- Learning is a type of Reinforcement Learning that seeks to maximize the action-value function defined as:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \underbrace{\max}_{a} Q(s_{t+1}, a_t) - Q(s_t, a_t)] \quad (1)$$

In this equation, r_{t+1} is the reward observed after performing action a_t in state s_t, α is the learning rate, and γ is the discount factor (Sutton & Barto, 1998). The learning rate determines how much the new learning signal will outweigh past learning signals. The discount factor determines the significance of future rewards ($\gamma = 0$ will only take into account immediate reward, and $\gamma = 1$ will seek a higher cumulative reward). The term

$\underbrace{\max}_{a} Q(s_{t+1}, a_t)$ is an estimate of the optimal future action value. If the policy is represented as a lookup table, this can be a long and arduous process, as the whole state-action space should preferably be explored during learning, including suboptimal actions. Depending on the problem dimensionality, there can be a large number of possible states and actions and the learning process may be computationally expensive.

Because of the limitations of traditional RL and Q-Learning, a method known as Q- Learning with Linear Functional Policy Approximation can help eliminate the need for unnecessary exploration, and simplify calculations. This method translates the state into a set of features and actions into a set of symbolic parameters (Irodova & Sloan, 2005), represented by the equation:

$$\pi(s, a) = \theta_1^a f_1 + \cdots \theta_n^a f_n \tag{2}$$

In this equation, $f_1 \ldots f_n$ represent the translated set of states, $\theta_1^a \ldots \theta_n^a$ represent the symbolic parameters, and $\pi(s, a)$ is the policy. Using the Q-learning equation (1), the following update rule for each parameter (θ_k^a) can be derived:

$$\theta_k^a = \theta_k^a + \alpha \left[r + \gamma \underbrace{\max}_{a'} Q^a(s', a') - Q^a(s, a) \right] \frac{dQ^a(s,a)}{d\theta_k^a} \tag{3}$$

In this equation, (a) represents the most recent action taken, (s) represents the most recent state observed, a' represents future action, and s' represents the future state (Irodova & Sloan, 2005). Q-Learning with Linear Functional Policy Approximation eliminates the

need to visit many state action pairs during the learning phase, some of which may be infeasible or even impossible.

B. Warfarin Q-Learning With Linear Policy Approximation

Q-Learning with linear functional policy approximation is the control method used to determine the optimal dose change, $\Delta\mathrm{Dose}$, and the optimal time between INR measurements, $\Delta\mathrm{Time}$. In this work, the clinical goal of warfarin therapy is to achieve a patient INR value of 2.5 represented by $\mathrm{INR}_{\mathrm{target}}$(this value is chosen because it is the midpoint of the warfarin therapeutic range of INR = 2.0-3.0). The control method uses the difference between the measured INR value and $\mathrm{INR}_{\mathrm{target}}$, defined by INRdiff , as the output of the system, where the desired state is for INRdiff = 0, defined by $\mathrm{INRdiff}_{\mathrm{target}}$. The elements of this method are detailed in FIGURE 1, and the goal of this method is for the environment to achieve the desired state, $\mathrm{INRdiff}_{\mathrm{target}}$.

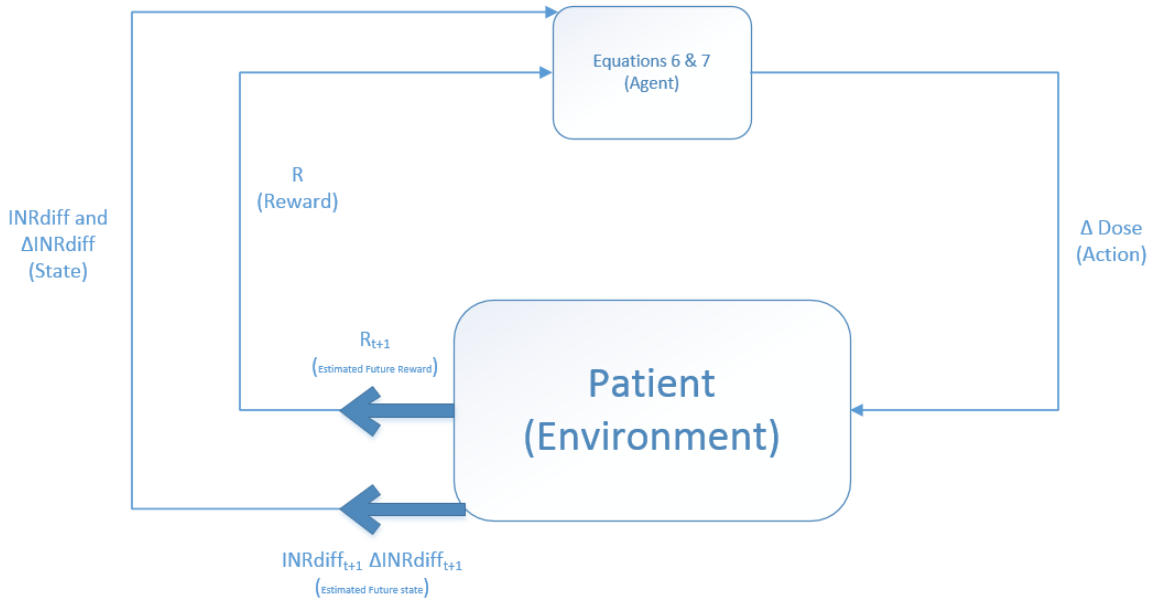

FIGURE 1 - Block Diagram of RL based Warfarin Dosing

The patient's body is the environment being affected, and more specifically the mechanisms that involve thrombus formation (Hirsh, Fuster, Ansell, & Halperin, 2003; Porter, 2010). In this work, two different patient models are used to represent the environment (patient), the control-relevant patient model, and the simulation model. The control-relevant patient model is used to design the controllers and is represented as follows:

$$\mathrm{INR}(t) = K * D(t) * t * e^{\left(-\frac{t}{T_s}\right)} \quad (4)$$

In control-relevant patient response model, K represents the gain (INR increase/mg/day of warfarin administered), t represents time (in days), and T_s represents the time constant

(time until the administered warfarin dose reaches 32% of its full pharmacodynamic (INR) effect). This model is used for controller design because it can represent the properties of the more complex process with sufficient accuracy, while minimizing the number of parameters. The simulation model used for simulation is represented by the equation:

$$\mathrm{INR}(t) = \left(K_1 * D(t) + K_2 * D^2(t)\right) * t * e^{\left(-\frac{t}{T_s}\right)} \quad (5)$$

In the simulation model, K_1 represents the linear gain (INR increase/mg/day of warfarin administered), K_2 represents the nonlinear (quadratic) gain, t represents time (in days), and T_s represents the time constant (time until the administered warfarin dose reaches 32% of its full pharmacodynamic (INR) effect). The agent is the body that governs the actions that are to be taken. In this work the agent is represented by the equations:

$$\Delta\mathrm{Dose} = a_0 + a * (\mathrm{INRdiff}) + c * (\Delta\mathrm{INRdiff}) \quad (6)$$

$$\Delta\mathrm{Time} = h_0 + h * (\mathrm{INRdiff}) + f * (\Delta\mathrm{INRdiff}) \quad (7)$$

The actions, governed by the agent, are $\Delta\mathrm{Dose}$ and $\Delta\mathrm{Time}$, where $\Delta\mathrm{Dose}$ is the dose change and $\Delta\mathrm{Time}$ is the time between INR measurements. The state is represented by the observed values INRdiff and ΔINRdiff. These values are determined by the equations:

$$\text{INRdiff} = \text{INR} - \text{INR}_{\text{target}} \tag{8}$$

$$\Delta\text{INRdiff} = \text{INRdiff} - \text{INRdiff}_{\text{Previous}} \tag{9}$$

INR is the patient's current INR value and $\text{INRdiff}_{\text{Previous}}$ is the INRdiff from the previous INR measurement. The reward is a value assigned to the action based on whether or not the environment moves closer to the desired state, $\text{INRdiff}_{\text{target}}$. Here, the equation used to assign the reward is as follows:

$$R = \left(\frac{1}{1+(\text{INRdiff})^2}\right) \tag{10}$$

As the INR moves farther away from the target value, $\text{INRdiff}_{\text{target}}$, the reward grows smaller, and as the INR moves closer to the target value, the reward grows larger. The maximum reward is R = 1.

C. Reinforcement Learning Warfarin Dosing Algorithm

The Overall Reinforcement Learning warfarin dosing algorithm is separated into 2 phases:

1) Learning phase (off-line)
2) Dosing phase (on-line)

The learning phase is the design phase and uses the principles of RL to "learn" the optimal parameter values that are implemented to calculate the proper ΔDose and ΔTime

during the dosing phase. The dosing phase uses the optimal parameter values extracted from the learning phase to calculate the proper ΔDose and ΔTime based on INRdiff and ΔINRdiff observed in the individual patient.

During the learning phase, 9 different controllers are created to optimize the warfarin dose for 9 different patient responder types. These different patient responder types range from extremely hypo-responsive (patient has a marginal INR increase compared to the dose administered) to extremely hyper-responsive (patient has a significant INR increase compared to the dose administered). To create the different controllers, the control-relevant patient response model is used. Nine different K values are used to create 9 different controllers, and are listed as follows:

TABLE I

CONTROL REVELANT K VALUES

K
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

These values are listed in order of increasing patient responsiveness ranging from extremely hypo-responsive (K = 0.1) to extremely hyper-responsive (K = 0.9). Three hundred learning episodes, each 120 days in length, are performed for each value of K. Before the first learning episode for each K value, initial values for the learned parameters are set: $a_0 = 0$, a = -1, c = 0, f = .01, h = -4, and $h_0 = 4$. Initial values are

arbitrary as the optimal parameter values will be determined over the course of the learning episodes. RL parameters γ and α, which are the discount factor and the learning rate respectively, are also initiated before the first learning episode and have initial values of $\gamma = 0.9$ and $\alpha = 0.1$. The values of the RL parameters are determined heuristically to help reduce simulation time. Once the parameter values and RL variables are initiated, the first learning episode begins.

Step 1: The algorithm calculates the Q values for all of the possible ΔDose values (-5mg/day, -4.9mg/day, -4.8mg/day…+4.8mg/day, +4.9mg/day, +5mg/day), and the calculations are made using the equation:

$$Q_{\Delta\text{Dose}} = e^{(-(a_0+a*(\text{INRdiff})-(\Delta\text{Dose})+c*(\Delta\text{INRdiff}))^2)} \quad (11)$$

The ΔDose that yields the highest $Q_{\Delta\text{Dose}}$ value is selected as the optimal dose change. The same thing is done for ΔTime, the algorithm calculates the Q values for all possible ΔTime values (1-6 weeks) using the equation:

$$Q_{\Delta\text{Time}} = e^{(-(h_0+h*(\text{INRdiff})+f*(\Delta\text{INRdiff})-(\Delta\text{Time}))^2)} \quad (12)$$

The ΔTime that yields the highest $Q_{\Delta\text{Time}}$ is then selected, and no INR measurements are made until the selected number of weeks has passed. Step 2: The new ΔDose and ΔTime are simulated using the control-relevant patient model, with the addition of random noise to compensate for external factors. A reward is then determined based on the difference between the most recent INR measurement and the target INR as governed

by equation (10). Step 3: The reward is then used to calculate value functions for both ΔDose and ΔTime using the equations:

$$V_{\Delta\text{Dose}} = R + \gamma\left(\underbrace{\max}_{\Delta\text{Dose}_{t+1}} Q_{\Delta\text{Dose}}(\text{INRdiff}_{t+1}, \Delta\text{Dose}_{t+1})\right) \tag{13}$$

$$V_{\Delta\text{Time}} = R + \gamma\left(\underbrace{\max}_{\Delta\text{Time}_{t+1}} Q_{\Delta\text{Time}}(\text{INRdiff}_{t+1}, \Delta\text{Time}_{t+1})\right) \tag{14}$$

In these equations, $\max Q_{\Delta\text{Dose}}(\text{INRdiff}_{t+1}, \Delta\text{Dose}_{t+1})$ and $Q_{\Delta\text{Time}}(\text{INRdiff}_{t+1}, \Delta\text{Time}_{t+1})$ are estimates of optimal future values. Step 4: The parameters a_0, a, c, f, h, and h_0 are then updated based on the value functions following the form listed in equation (3). The update equations are described as follows:

$$a = a + \alpha(V_{\Delta\text{Dose}} - Q_{\Delta\text{Dose}})\left(\frac{\partial Q_{\Delta\text{Dose}}}{\partial a}\right) \tag{15}$$

$$c = c + \alpha(V_{\Delta\text{Dose}} - Q_{\Delta\text{Dose}})\left(\frac{\partial Q_{\Delta\text{Dose}}}{\partial c}\right) \tag{16}$$

$$f = f + \alpha(V_{\Delta\text{Time}} - Q_{\Delta\text{Time}})\left(\frac{\partial Q_{\Delta\text{Time}}}{\partial f}\right) \tag{17}$$

$$h = h + \alpha(V_{\Delta\text{Time}} - Q_{\Delta\text{Time}})\left(\frac{\partial Q_{\Delta\text{Time}}}{\partial h}\right) \tag{18}$$

$$h_0 = h + \alpha(V_{\Delta Time} - Q_{\Delta Time})\left(\frac{\partial Q_{\Delta Time}}{\partial h_0}\right) \quad (19)$$

$a_0 = 0$ remains constant.

Steps 1-4 are repeated until the length of the simulation meets or exceeds 120 days ($time \geq 120$), and once this occurs, the learning episode is finished. The learning rate is then reduced using the equation:

$$\alpha = \alpha(0.99) \quad (20)$$

Once the learning rate is updated, the learning episode is completed, and a new learning episode begins. Each new learning episode exploits the previously learned parameter values a_0, a, c, f, h, and h_0, so that they are continuously updated during each learning episode. When all 300 learning episodes are completed, the learned optimal parameter values a_0, a, c, f, h, and h_0 are extracted.

The Dosing Phase can begin once the learning phase is complete. During the initiation of dosing, an initial dose of 5mg/day of warfarin is given to the patient. Patient's INR measurements are taken on a weekly basis (days 7, 14, and 21), and after the measurement is taken, a new dose is determined by the algorithm and administered to the patient until it is time for the next INR measurement. To adjust the dose, a control-relevant K value is estimated based on the most recent dose using the equation:

$$K_{ss} = \frac{1.5}{Dose} \quad (21)$$

The K value from table 1 that is closest to the calculated K_{ss} value is then selected as the control-relevant K value. The value of 1.5 is used in the first equation because that is the

difference between the patient's initial INR value (1) and the target INR value. Once the control-relevant K value is determined, the parameters associated with that control-relevant K value, INRdiff, and ΔINRdiff are used to determine the new dose:

$$\text{Dose} = \Delta\text{Dose} + \text{Dose}_{\text{Previous}} \tag{22}$$

In this equation, $\text{Dose}_{\text{Previous}}$ is the most recent dose given to the patient before the INR measurement. The calculated new dose is then given to the patient until the next INR measurement, and this process is repeated on day 7, 14, and 21 (initiation phase). After day 21 (maintenance phase), ΔTime is determined by the equation (7). After day 21, INR is only measured when the algorithm suggests. For maintenance dosing, the patient's INR is measured when ΔTime suggests, the control-relevant K value is estimated, and the new Dose and ΔTime are calculated.

To prove the efficacy of the new RL warfarin algorithm, three simulations were performed on each of 100 artificial patients (with varying warfarin responses), and each patient was simulated for 180 days of therapy. The first simulation was performed as an ideal case scenario (random noise with CV = 0%) for verification of the RL warfarin algorithm in the presence of no intra-patient variability. It is, however, impossible to eliminate intra-patient variability in a real world scenario, so two more simulations were performed to compare the industry standard of care to the RL warfarin algorithm. The second simulation was performed following the guidelines for dosing and INR measurements stipulated in the current standard of care, and a CV = 25% was used to simulate extreme intra-patient variability. The third simulation was performed following

guidelines stipulated by the RL warfarin algorithm, and this simulation also used a CV = 25% to simulate extreme intra-patient variability. Simulations two and three were then used to compare the industry standard of care to the RL warfarin algorithm using %TTR (time in therapeutic range) and the number of INR measurements as performance criteria.

VI. RESULTS AND DISCUSSION

All simulations, calculations, and graphs were performed and made using MATLAB software. Before simulation using the RL could take place, the learning phase to determine the optimal parameter values associated with each control-relevant K value had to be performed, and the results were as follows:

TABLE 2

PARAMETER VALUES

K	a_0	a	c	f	h	h_0
0.1	0	-3.00367	0.607573	0.413057	-3.79183	4.407258
0.2	0	-2.25922	0.562711	0.737318	-3.35005	4.774702
0.3	0	-0.82387	0.813361	0.573093	-3.47947	4.598232
0.4	0	-0.4069	0.737377	0.057894	-4.13071	3.924247
0.5	0	-0.58846	0.699238	0.835709	-3.10336	4.885482
0.6	0	-0.26811	0.564213	0.123133	-4.27049	3.988894
0.7	0	-0.25454	0.571812	0.274075	-4.01829	4.225859
0.8	0	-0.1582	0.503708	0.154036	-4.37039	3.982359
0.9	0	-0.19254	0.446211	0.232318	-4.13157	4.19317

The control relevant K values are listed in the first column of TABLE 2, and the remaining columns list the parameter values associated with the control relevant K value in the same row. The values for parameter "a" trended towards 0 as the control relevant K value increased. When the values for parameter "a" were looked at in the context of equation (6), it indicated that when the patient became more responsive to the warfarin dose (control relevant K value increased), the same INRdiff values would result in a smaller dose change. This means that if a patient was determined to have a high control relevant K value, a smaller dose change is necessary to achieve the desired effect.

The results for parameter value "c" varied, which indicated that ΔINRdiff had a varying effect depending on the responder type. The parameter values for "h_0", when taken in context of equation (7), were determined to be the optimal times between INR measurements when the INRdiff and ΔINRdiff values were both 0, which would mean the patient had reached the target values for INRdiff and ΔINRdiff. The parameter values for both "f" and "h" varied, which indicated that the values for INRdiff and ΔINRdiff had varying effects depending on the control relevant K value, respectively, when taken in the context of equation (7).

The first simulation was performed using only the RL algorithm in the presence of no intra-patient variability (CV = 0%) and the results were as follows:

TABLE 3

SIMULATION 1 RESULTS

Simulation	**Dosing Algorithm**	**CV**	**%TTR**	**%TTR standard deviation**	**mean # of INR Measurements**
1	RL	0%	92.1%	4.2%	7.94/patient

The first simulation, represented in TABLE 3, yielded a mean %TTR = 92.1% over all 100 artificial patients, with a standard deviation of 4.2%, and had a mean number of INR measurements = 7.94 measurements/patient. These results indicated, that in the presence of no intra-patient variability, the RL algorithm did an exceptional job of keeping the simulated patients' INR values within the therapeutic range. The reason the RL algorithm was not able to attain a higher %TTR was due to 2 factors. First, each patient started off with an INR value of 1, meaning that there was always a time when the patient's INR values were not in therapeutic range. Second, the initial dose given to the

patient was 5mg/day (the American Society of Hematology recommended starting dose (Cushman, Lim, & Zakai, 2011)), which, in the cases of the medium and hyper responder types (see FIGURES 3 and 4), can cause an overshoot of the therapeutic range due to an incorrect initial dose and not due to controller action. FIGURES 2, 3, and 4 are sample plots of the results that were attained from single patient types (hypo, medium, and hyper responders). A comparison to the current standard of care was still necessary to prove that the RL algorithm was viable.

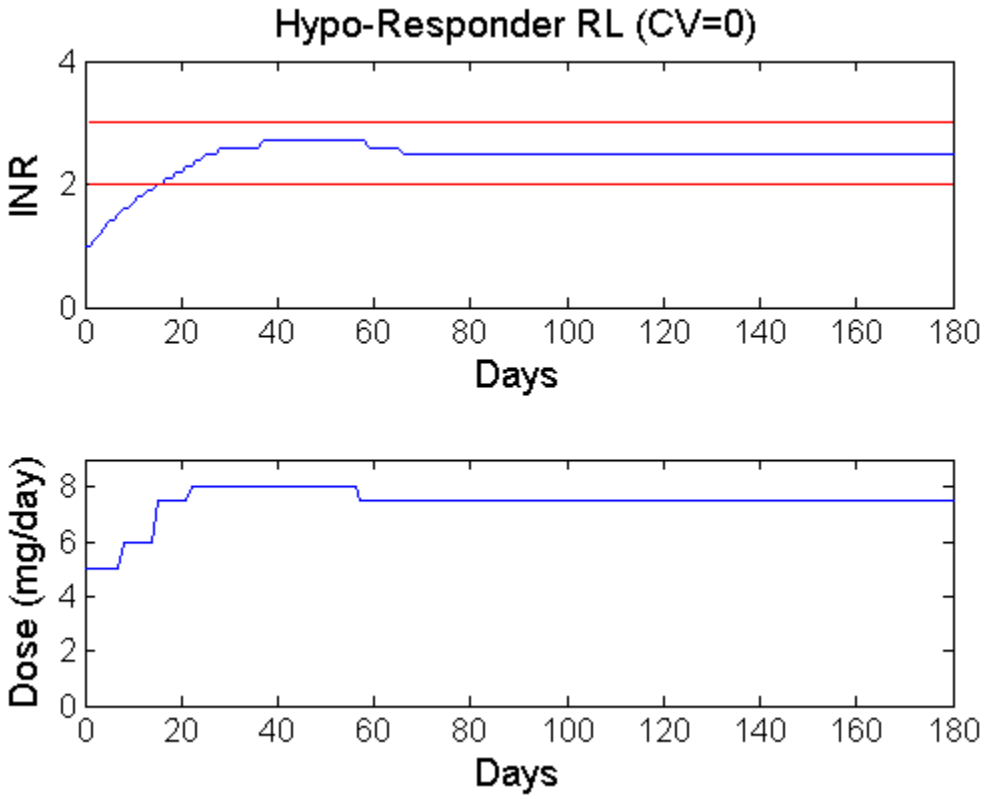

FIGURE 2 - Sample plot of a single patient - Hypo-Responder. The top plot is the patient's INR response, where the blue line is INR and the red lines indicate therapeutic range. The bottom plot is the dose.

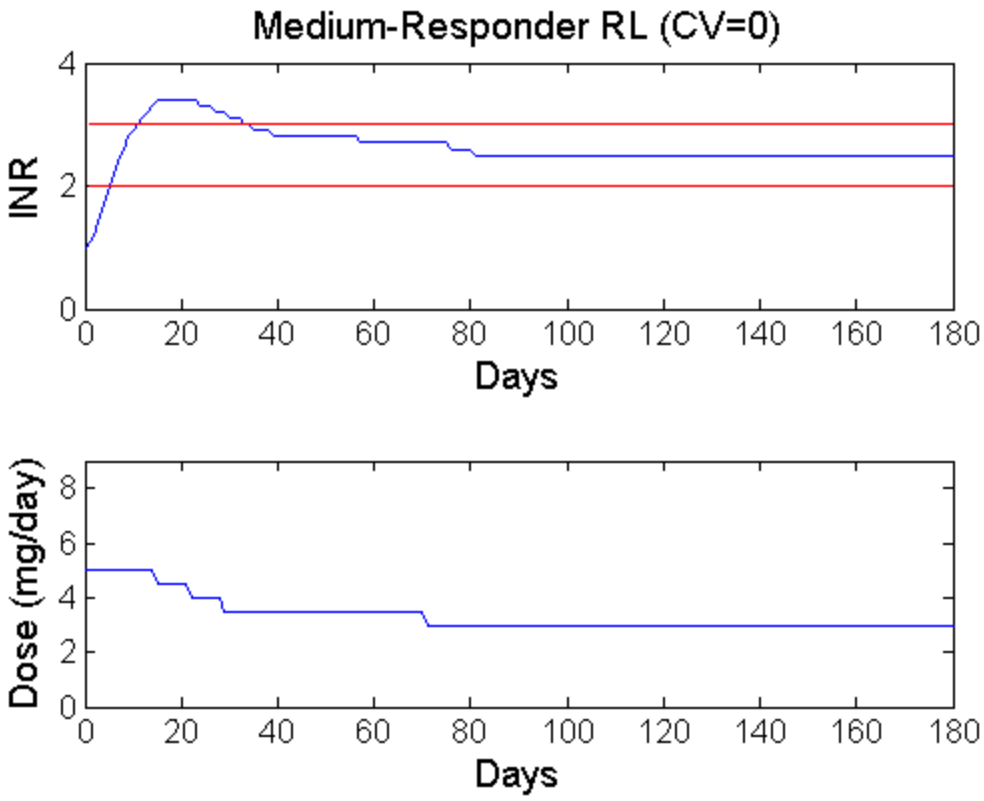

FIGURE 3 - Sample plot of a single patient - Medium-Responder. The top plot is the patient's INR response, where the blue line is INR and the red lines indicate therapeutic range. The bottom plot is the dose.

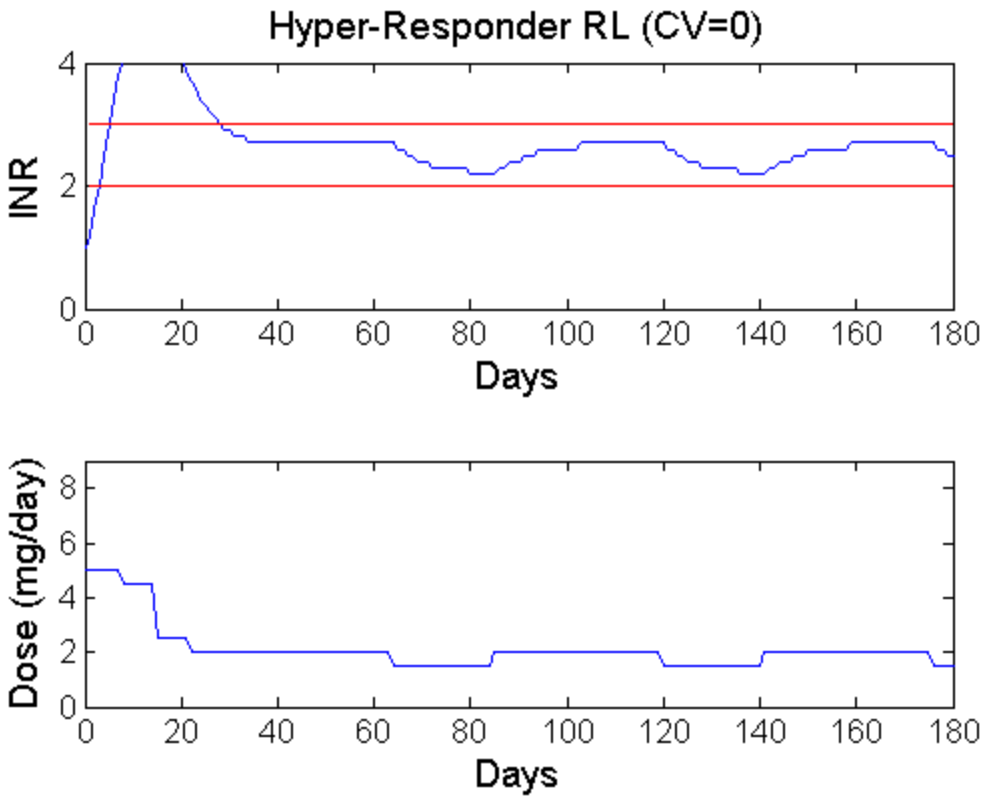

FIGURE 4 - Sample plot of a single patient - Hyper-Responder. The top plot is the patient's INR response, where the blue line is INR and the red lines indicate therapeutic range. The bottom plot is the dose.

The second simulation was performed with all 100 artificial patients using the American Society of Hematology dosing algorithm (represented as ASH in TABLE 4), which is the current standard of care, and the third simulation was performed with all 100 artificial patients using the warfarin RL algorithm (indicated in TABLE 4 as RL). Simulations two and three were both performed in the presence of heavy intra-patient variability (CV = 25%) and the results were as follows:

TABLE 4

SIMULATIONS 2 AND 3 RESULTS

Simulation	Dosing Algorithm	CV	%TTR	%TTR standard deviation	mean # of INR Measurements	RL vs ASH pvalue
2	ASH	25%	45.3%	16.4%	12.30/patient	
3	RL	25%	51.8%	10.8%	8.05/patient	<.001

The second simulation, represented in TABLE 4, yielded a mean %TTR = 45.3% over all 100 artificial patients, with a standard deviation of 16.4%, and had a mean number of INR measurements = 12.3 measurements/patient. The third simulation, represented in TABLE 4, yielded a mean %TTR = 51.8% over all 100 artificial patients, with a standard deviation of 10.8%, and had a mean number of INR measurements = 8.05 measurements/patient. There was determined to be a statistically significant difference ($P<.001$) between the American Society of Hematology algorithm and the warfarin RL algorithm. Figures 5, 6, and 7 are sample plots of the results that were attained from single patient types (hypo, medium, and hyper responders).

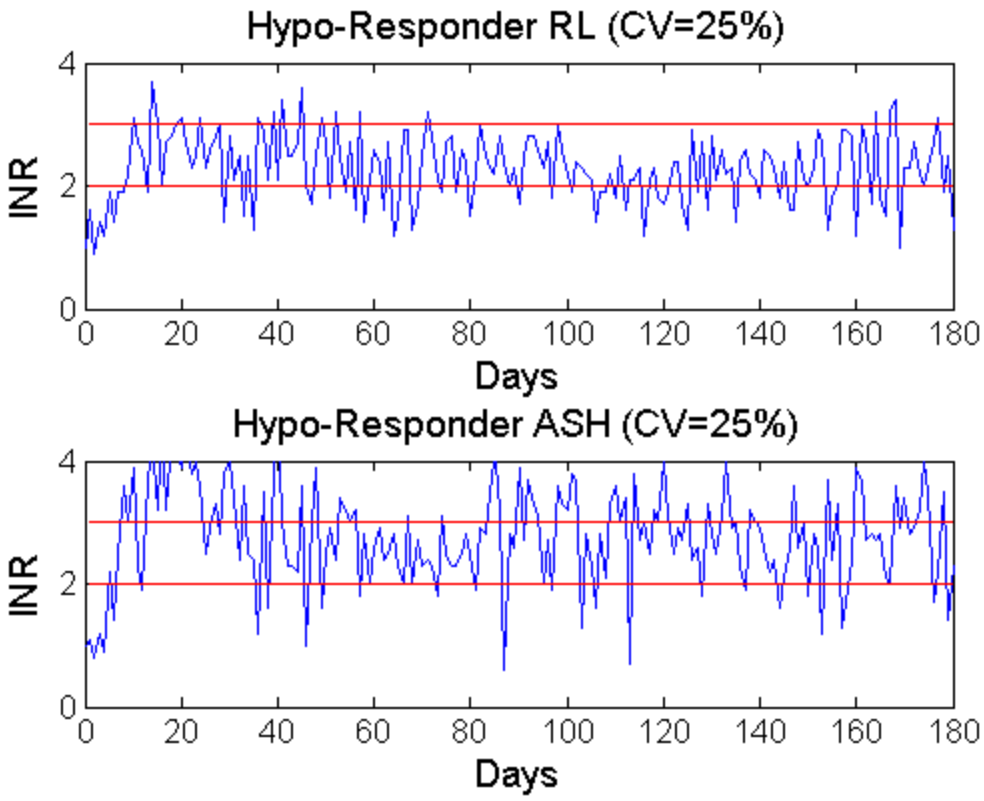

FIGURE 5 - Sample plot of a single patient - Hypo-Responder. The top plot is the patient's INR response using the RL algorithm, where the blue line is INR and the red lines indicate therapeutic range. The bottom plot is the patient's INR response using the ASH algorithm.

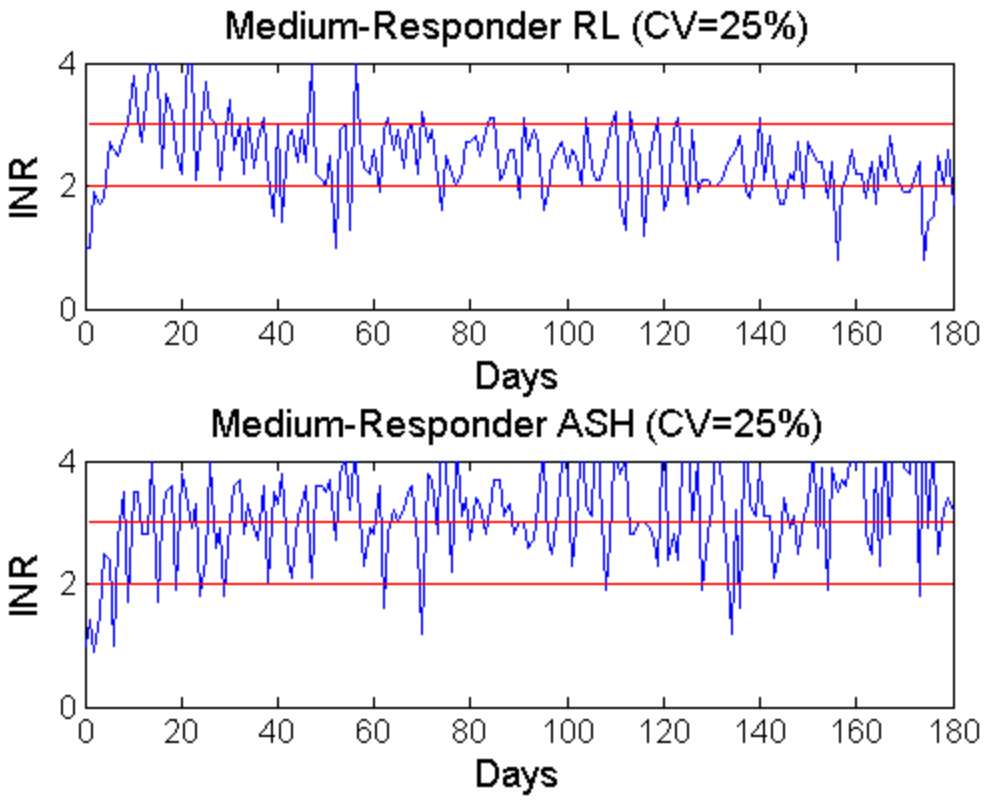

FIGURE 6 - Sample plot of a single patient - Medium-Responder. The top plot is the patient's INR response using the RL algorithm, where the blue line is INR and the red lines indicate therapeutic range. The bottom plot is the patient's INR response using the ASH algorithm.

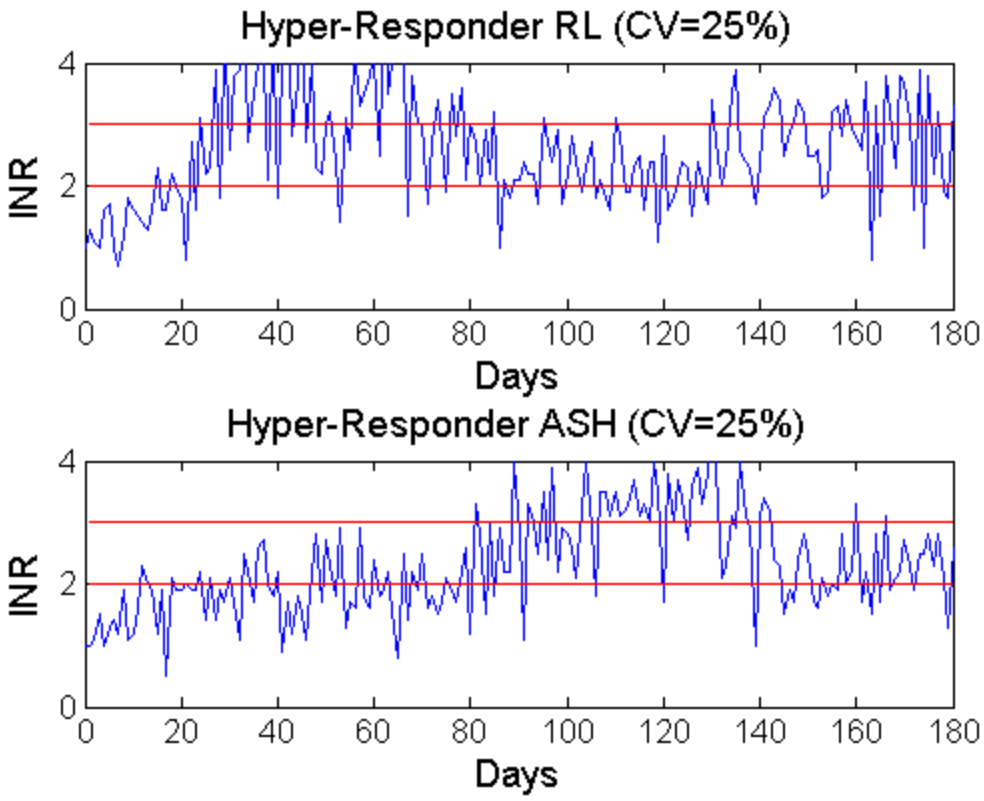

FIGURE 7 - Sample plot of a single patient - Hyper-Responder. The top plot is the patient's INR response using the RL algorithm, where the blue line is INR and the red lines indicate therapeutic range. The bottom plot is the patient's INR response using the ASH algorithm.

The results listed in TABLE 4 indicated that the warfarin RL algorithm not only did a better job at keeping the patient's INR in therapeutic range, but also reduced the number of INR measurements that were required per patient. When translated into a real world scenario, the data from TABLE 4 indicated that the warfarin RL algorithm would result in greater patient safety and therapeutic efficacy by keeping the patient's INR values in therapeutic range for a greater amount of time than the current standard of care.

The warfarin RL algorithm was also able to reduce the mean number of INR measurements that were necessary over all 100 artificial patients, which indicated that the overall cost of the warfarin therapy when using the warfarin RL algorithm would be less than the cost of warfarin therapy when using current standard of care.

The %TTR values listed in TABLE 4 correspond closely with %TTR values that are typically seen in clinical practice (Wilson, Costantini, & Crowther, 2007). The warfarin RL algorithm performed better than the current standard of care (ASH algorithm) at keeping patient's INR values in the therapeutic range due to its ability to switch between controllers to match each patient's response instead of slowly titrating the dose until the desired effect is achieved. When a patient is matched to their respective control relevant K value, the algorithm is able to make smaller or larger dose changes to match the possible patient response, whereas the ASH algorithm uses titration to achieve the desired INR value. This means that the warfarin RL algorithm is able to respond faster and better when a patient's response to warfarin dose changes, which can occur due to intra-patient variability factors (diet, drug interactions, and disease state), than titration based algorithms.

The mean number of INR measurements that were necessary, listed in TABLE 4, are another important factor to consider when comparing the warfarin RL algorithm with the current standard of care. As previously stated, utilizing the warfarin RL algorithm resulted in a fewer number if INR measurements that were necessary when compared to the current standard of care, and that would result in a reduction of the overall cost of the therapy. The rising cost of healthcare and the uncertain changes that are occurring in the U.S. healthcare market are important considerations when evaluating this metric. If a

patient could receive superior healthcare at a lower cost, the patient would be, overall, more satisfied. Also, the cost of warfarin therapy is more than just monetary, and patient convenience is an important factor to consider. If fewer INR measurements were necessary over the course of warfarin therapy, this would greatly increase patient convenience and quality of life.

Drug dosing as a whole, even outside of the realm of warfarin and anticoagulants, could greatly benefit from the use of engineering methods like the one presented in this study. For most drugs, outdated dosing methodologies are used to slowly titrate the dose until a desirable effect is achieved. These methodologies can be inefficient, and can even have the potential to be dangerous if the proper dose is not determined fast enough or the methodology used is slow to respond to inter and intra patient variability. There is a need for the development of new dosing methodologies that utilize engineering methods to improve patient safety, and reduce the cost of different types of therapy by more "intelligently" dosing patients.

These methods could be applied to drugs like Plavix, which is an antiplatelet agent, and also other unrelated drugs such as erythropoiesis stimulating agents, which are used to stimulate red blood cell production in patients with End Stage Renal Disease. In fact, there is evidence in the literature which suggests that control systems engineering methodologies have been effective in real life clinical settings (Gaweda, Jacobs, Arnoff, & Brier, 2008). If the medical community were to develop and adopt "smarter" dosing protocols based on control systems engineering techniques, patient's receiving a variety of different drug therapies would benefit greatly.

VII. CONCLUSIONS

Based on the data from the simulations that were performed, the patient population as a whole could greatly benefit from using the RL warfarin algorithm as an alternative to the current standard of care. Using the RL warfarin algorithm could help keep patient's INR in therapeutic range in the presence of heavy intra-patient variability while also greatly reduce the cost of the therapy by optimizing the number of INR measurements that are required. The RL Warfarin Algorithm offers distinct advantages compared to the industry standard dosing methods, and while there are other computational and evidence based algorithms in practice, no other algorithm optimizes monitoring frequency. Next, a human study of the RL warfarin algorithm should be performed to ensure patient safety and efficacy.

LIST OF REFERENCES

Ansell, J., Hirish, J., Hylek, E., Jacobson, A., Crowther, M., & Palareti, G. (2008). Pharmacology and Management of the Vitamin K Antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Antithrombotic and Thrombolytic Therapy 8th Edition: ACCP Guidelines, 160s-198s.

Chambers, S., Chadda, S., & Plumb, J. M. (2009). How much does international normalized ratio monitoring cost during oral anticoagulation with a vitamin K antagonist? A systematic review. International Journal of Laboratory Hematology, 427-442.

Chappell, J. C., Dickinson, G., Mitchell, M. I., Haber, H., Jin, Y., & Lobo, E. D. (2012). Evaluation of methods for achieving stable INR in healthy subjects during a multiple-dose warfarin study. European Journal of Clinical Pharmacology, 239-247.

Cushman, M., Lim, W., & Zakai, N. A. (2011). 2011 Clinical Practice Guide on Anticoagulant Dosing and Management of Anticoagulant-Associated Bleeding

Complications in Adults. American Society of Hematology.

Dimberg, I., Grzymala-Lubanski, B., Hägerfelth, A., Rosenqvist, M., Svensson, P., & Själander, A. (2012). Computerised assistance for warfarin dosage — Effects on treatment quality. European Journal of Internal Medicine, 742-744.

Gage, B. F., Fihn, S. D., & White, R. H. (2000). Management and Dosing of Warfarin Therapy. The American Journal of Medicine, 481-488.

Grzymala-Lubanski, B., Själander, S., Renlund, H., Svensson, P. J., & Själander, A. (2013). Computer aided warfarin dosing in the Swedish national quality registry AuriculA – Algorithmic suggestions are performing better than manually changed doses. Thrombosis Research, 130-134.

Hanley, J. P. (2004). Warfarin Reversal. Journal of Clinical Pathology, 1132-1139.

Harrington, A. R., Armstrong, E. P., Nolan, P. E., & Malone, D. C. (2013). Cost-Effectiveness of Apixaban, Dabigatran, Rivaroxaban, and Warfarin for Stroke Prevention in Atrial Fibrillation. Stroke: Journal of The American Heart Association.

Hirsh, J., Fuster, V., Ansell, J., & Halperin, J. L. (2003). American Heart Association/American College of Cardiology Foundation Guide to Warfarin Therapy. Journal of the American College of Cardiology, 1633-1652.

Kangelaris, K. N., Bent, S., Nussbaum, R. L., Garcia, D. A., & Tice, J. A. (2009). Genetic Testing Before Anticoagulation? A Systematic Review of Pharmacogenetic Dosing of Warfarin. Journal of General Internal Medicine, 656-

664.

Lader, E., Martin, N., Cohen, G., Meyer, M., Reiter, P., Dimova, A., & Parikh, D. (2012). Warfarin therapeutic monitoring: is 70% time in the therapeutic range the best we can do? Journal of Clinical Pharmacy and Therapeutics, 375-377.

Lafata, J. E., Martin, S. A., Kaatz, S., & Ward, R. E. (2000). The Cost-Effectiveness of Different Management Strategies for Patients on Chronic Warfarin Therapy. Journal of General Internal Medicine, 31-37.

Meehan, R., Tavares, M., & Sweeney, J. (2013). Clinical experience with oral versus intravenous vitamin K for warfarin reversal. Transfusion, 491-498.

Merli, G. J., & Tzanis, G. (2009). Warfarin: what are the clinical implications of an out-of-range-therapeutic international normalized ratio? Journal of Thrombosis and Thrombolysis, 293-297.

Nieuwlaat, R., Barker, L., Kim, Y.-K., Haynes, R. B., Eikelboom, J. W., Yusuf, S., & Connolly, S. J. (2010). Underuse of evidence-based warfarin dosing methods for atrial fibrillation patients. Thrombosis Research, 128-131.

Porter, W. R. (2010). Warfarin: history, tautomerism and activity. Journal of Computer-Aided Molecular Design, 553-573.

Siguret, V., Gouin, I., Debray, M., Perret-Guillaume, C., Boddaert, J., Mahe, I., . . . Pautas, E. (2005). Initiation of warfarin therapy in elderly medical inpatients: A safe and accurate regimen. The American Journal of Medicine, 137-142.

The International Warfarin Pharmacogenetics Consortium. (2009). Estimation of the

Warfarin Dose with Clinical and Pharmacogenetic Data. The New England Journal of Medicine, 753-764.

Wardrop, D., & Keeling, D. (2008). The story of the discovery of heparin and warfarin. British Journal of Haematology, 757-763.

White, P. J. (2010). Patient Factors That Influence Warfarin Dose Response. Journal of Pharmacy Practice, 194-204.

Wilson, S. E., Costantini, L., & Crowther, M. A. (2007). Paper-based dosing algorithms for maintenance of warfarin anticoagulation. Journal of Thrombosis and Thrombolysis, 195-198.