A User Study on Data-Driven Heparin Dosing

Joe Rowley

A thesis presented for the degree of

Bachelor of Science

Advised by:

Sri Kurniawan

Mohammad M. Ghassemi

University California, Santa Cruz

December 2015

Abstract

The intention of this study was to determine the feasibility of using a statistical drug dosing tools in a mock clinical setting. We wanted to learn about both the technical implementation and user experience challenges involved with creating and using computer-aided dosing tools. First, a web application was developed using existing models for dosing heparin in ICU patients. This application was next developed into an interactive user survey. We found that doctors were generally more confident when dosing using the tools, but also expressed concerns about their understanding of the tool. In closing, we recommend some best practices for future development of statistical drug dosing tools.

Acknowledgements

The models used in this tool were introduced Mohammad M. Ghassemi, Stefan E. Richter, Ifeoma M. Eche, Tszyi W. Chen, John Danziger, and Leo A. Celi in their paper *A data-driven approach to optimized medication dosing: a focus on heparin*. This project would not have been possible without Mohammad's support.

Additionally, data used in these models was sourced from the MIMIC-II and MIMIC-III (Saeed et al. (2011)) medical databases, a project supported by the MIT Lab for Computational Physiology.

I also have many advisors who helped me with this project:

- Mohammad M. Ghassemi
- Sri Kurniawan
- Matthew Guthaus

Additionally the following people were very influential in preparing me for this project:

- James Davis
- Phokion Kolaitis
- Philip Strong

Finally, this project would not have been possible without the unwavering moral and technical support of my parents Howard Rowley and Carol Diamond.

Table of Contents

1 Introduction

The primary goal of this study was to determine the technical implementation challenges and user experience challenges involved with using a statistical dosing tool in a clinical setting. Specifically, we implemented a heparin dosing tool that was first introduced in the paper *A data-driven approach to optimized medication dosing: a focus on heparin* (Ghassemi et al. 2014). In the paper a statistical model is generated for predicting an optimal infusion rate of heparin for patients in intensive care units (ICUs). In this study, we reimplemented their existing models into an interactive web application, which allows the user to generate dosage guidelines based on a patient's features such as age, weight, and other lab results. Next a user study web application was created, which presents the doctors with a patient case. The doctor initially provides their own suggestion for the infusion rate and is then presented with the model's prediction for an optimal infusion rate. After the doctor views the model, they have the opportunity to adjust their initial dose and provide a reasoning for their dose adjustment. Their interaction with the tool yields many metrics, which when combined with their text based comments provides us insight into how they used the tool. What follows is a description of how we implemented the application, conducted the survey, and some proposed future areas for research.

2 Background

2.1 Traditional Dosing of Heparin

Heparin is an anticoagulant medication administered intravenously to patients with blood clots. Specifically, it is often used for strokes, pulmonary embolism (blockage in an artery in the lung), acute coronary syndrome (e.g. heart attack), and many other conditions. Each hospital has it's own set of protocols for how to dose heparin, but the general protocol for dosing heparin is to initially order baseline lab results for the patient. These values combined with context of the patients condition allow a doctor to determine the infusion rate and if a bolus (a single large initial dose) is required. The primary metric used for monitoring patient progress is activated partial thromboplastin time (aPTT). aPTT is medical test specifically used to determine how well blood is coagulating. It is measured in seconds, with a therapeutic range defined as between 60 and 100 (Ghassemi et al. 2014). After the infusion is started the patient is monitored and the dose is adjusted periodically. Most often new labs are ordered every 6 hours (Lisa Gryttenholm 2004), (Washington Pharmacy Services 2014).

2.2 Predicting Therapeutic Outcome

The objective of the dosing tool is determine an ideal initial infusion rate, such that a patient has the highest probability of reaching a therapeutic state in 6 hours time. A therapeutic state is defined as an aPTT value of between 60 seconds and 100 seconds. To predict an optimal dose Ghassemi

et al. employed the use of a multi-variant logistic regression.

2.2.1 Logistic Regression Approach

The objective of the model was to determine the probability that a patient would end in one of three categories: sub-therapeutic, therapeutic, or supra-therapeutic. A patient can by definition only be in one category at a time. This lead Ghassemi et al. to the following equation for describing the probability of being a given state:

(1) $1 = P(sub{-}therapeutic) + P(therapeutic) + P(supra{-}therapeutic)$

The challenge of using logistic regression to predict therapeutic outcomes is that the probability of a therapeutic outcome is not a strictly monotonic function. So using logistic regression directly to predict P(therapeutic) will not work. That being said, the probability of sub-therapeutic and supra-therapeutic outcomes are monotonically decreasing and increasing functions, respectively. Given these properties, we can use multinomial logistic regression to determine functions for sub-therapeutic and supra-therapeutic aPTT. Using Equation 1 we can then determine the probability of a therapeutic outcome.

At this point a case has been made that multinomial logistic regression may be an appropriate tool, but there are a few considerations that Ghassemi et al. took with regards to choosing the form of regression. There are many forms of multinomial logistic regression with subtle but important differences. There is a natural order to the classifications (sub-therapeutic, therapeutic, supra-therapeutic), so ordinal logistic regression could be considered. The challenge with ordinal logistic regression, however, is that it makes the as-

sumption of proportional odds, so regression intercept terms vary across classes while regression coefficients are shared across classes. In the words of the original authors *"it assumes that explanatory features maintain identical effects across varying ranges of the outcome."* Instead of the ordinal regression they used a modified version of standard multinomial logistic regression, which utilized a flexible reference category. That is instead of using a single class to predict the outcome that is not being modeled, they used all classes. So when modeling supra-therapeutic aPTT, standard multinomial logistic regression would reference either therapeutic or sub-therapeutic aPTT classes, while their approach referenced both classes. This choice was demonstrated to have significant effects on the validity of their models (Ghassemi et al. 2014).

2.2.2 Feature Selection

There are innumerable potential features one could consider as predictors of patient outcome. In their paper, Ghassemi et al. determined the salient features as follows: age, SOFA score, Elixhauser, Heparin dose, aPTT Measurement time, Creatinine, Ethnicity, Gender, ICU type, presence of pulmonary embolism, obesity, and presence of end stage renal failure.

2.3 Summary of Prior Work

The original paper and associated appendix document the efforts taken to validate their models. To summarize, they partitioned their dataset, dedicating 70% to training and 30% for validation. They then compared the

predicted classification for each patient using both a full featured model and a weight-only model. The full featured model was more accurate than the weight-only approach as measured by Volume Under the Receiver Operating Characteristic Surface (the multiple class version of Area Under Receiver Operating Characteristic Curve), with values of 0.48 vs 0.42 (Ferri et al. 2003). In closing, they noted that a randomized controlled trial would be necessary to determine if this mechanism is more effective for dosing than the traditional guidelines. This thesis project represents the initial steps towards this trial.

3 Heparin Dosage Calculator

This project is composed of two key components. The first is a "dose calculator" which contains all the functionality required to determine the optimal dose for a given patient. The second component is a survey that incorporates the calculator and is used to study how doctors interact with the dosing tool. We now describe the calculator.

3.1 Functional Requirements

The purpose of the calculator is to determine the probability functions for sub-therapeutic, therapeutic, supra-therapeutic aPTT for a range of infusion amounts for a given patient. The calculation can be broken into three steps:

1. Access and refine the dataset.
2. Calculate static models.
3. Make a prediction based on a given patient's features.

3.1.1 Data Processing

The data source for the original study was the MIMIC II database. MIMIC is an open access database consisting of over 40,000 deidentified patient encounters from Beth Israel Deaconess Hospital's ICUs. It is hosted at MIT's Lab for Computational Physiology. The majority of patients in an ICU do not receive heparin, so filtering by medication was needed. This was accomplished by a series of SQL queries on a local machine. Furthermore, we worked to maintain an identical dataset as was used in the original dataset. This required filtering out all transfer patients and all patients missing required features. Additionally some features viz. Elixhauser comorbidity index, mean Sequential Organ Failure Assessment (SOFA), required manual calculation on a per patient basis, as they were not available directly from MIMIC. This yielded a dataset of approximately 1,600 patients. This number varies depending on the features supplied for a given prediction.

This data munging was primarily done in Python using the Pandas library, but also required PostgreSQL for hosting the database. SQL queries were composed by hand. To prepare the data for model generation we needed to code several variables. Specifically, we needed to add binary classifiers for sub-therapeutic and supra-therapeutic aPTT to each patient's dataset. Additionally, a binary classifier was needed for ethnicity, and we followed the original coding used in the paper (white and non-white) (Ghassemi et al. 2014).

3.1.2 Model Generation

Once the dataset was prepared, generating the models was fairly simple. We used the Python library Scikit-learn's (Pedregosa et al. 2011) implementation of multinomial Logistic Regression to compute the two logistic regressions. Once the two logistic regressions for sub-therapeutic and supra-therapeutic aPTT are generated they can be stored for later use as long as the dataset has not changed.

3.1.3 Model Prediction

To generate the final dose curves for a specific patient, we use Scikit-learn's logistic regression predict_proba function . A function was created to output the probability of sub-therapeutic, sup-therapeutic aPTT given a patients features and a dose. For graphical purposes, we call this function repeatedly for varying infusion rates in the range of 0 to 34 units/Kg/HR in steps of 0.5 units/Kg/HR. Predicting the specific infusion rate which maximizes the probability of a therapeutic outcome can also be thought of as finding the point which minimizes the sum of the probabilities of sub-therapeutic and supra-therapeutic (see equation 1). Using this fact we used the Python library Scipy's minimize_scalar (Jones et al. 2001) function which implements Brent's Method (Jones et al. 2015) and let it solve for the highest probability with a tolerance of .01% . The infusion rate curves and the optimal dose are then cached in Postgres for future queries that match the same patient features, given that the initial dataset has not changed.

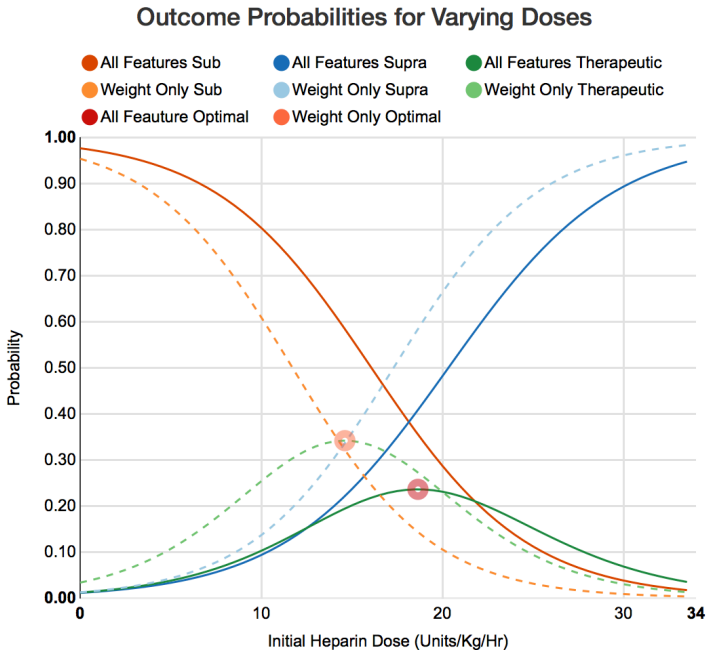


Figure 1: Graph generated by the client software displaying an all features model vs. a weight-only model.

3.2 Architecture

3.2.1 Front-end

The front-end for the application was built in Angular.js (Google 2015). Angular.js is client-side MVC (Model View Controller) framework that was chosen for it's strengths in quick prototyping. The user interface was built on Twitter Bootstrap (Twitter 2015), as it was familiar to the authors, is relatively easy to use and offers built in responsive design. Additionally, it has many CSS components such as nice forms and progress bars that are useful for the design.

For graphing the D3.js (Bostock 2015), NV-D3 (Partners 2015) and Angular-nvd3 (Skipor 2015) libraries were used which enables inserting graphs using simple directives. Figure 1 shows an example output of the standalone cal-

culator as seen in Appendix 2 or online. The two prominent dots represent the optimal doses for each model. Note that in this example, the weight-only model predicts that the probability that the patient will reach therapeutic state is higher than the all features model.

3.2.2 Back-end

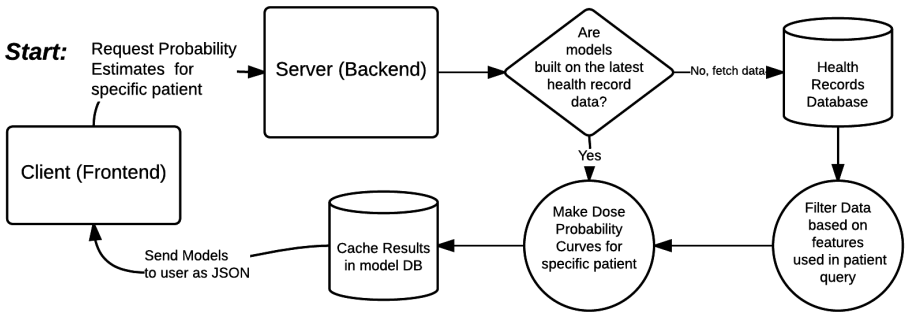


Figure 2: Example of a typical client server interaction.

The back-end for the application utilizes Python's Flask (Ronacher n.d.) webframe work. Data is stored in postgreSQL (Group 2008) and accessed through the sqlalchemy object relational mapping library (Bayer 2015). The front-end interfaces with the server via a stateless RESTful API which enables the exchange of JSON data. Figure 2 shows an example of how a client interacts with the back-end server currently.

4 Heparin Survey

4.1 Survey Design

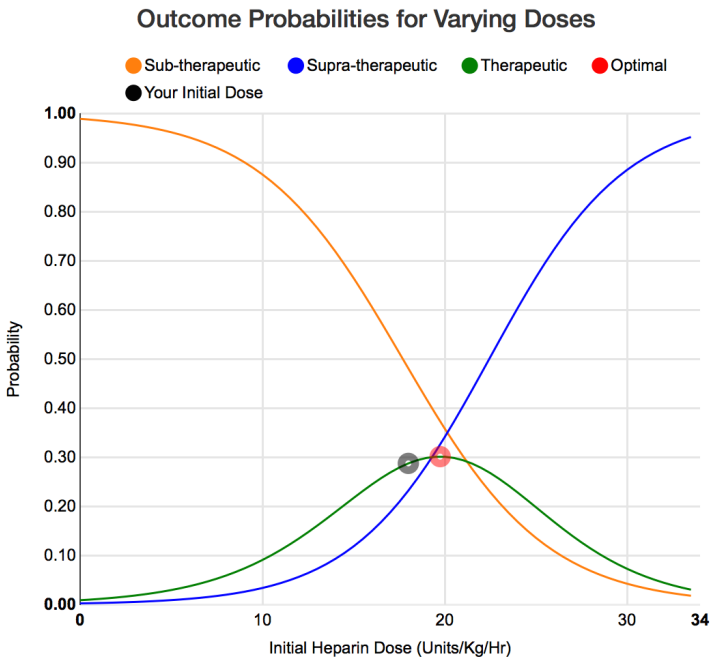


Figure 3: Dosage curves with doctors initial dose plotted in black on the therapeutic curve.

The primary objective of the survey was to learn about the user experience challenges involved with data-driven dosage techniques. To do this, a test interface was developed and several medical doctors of varying disciplines attempted to use it. You can view the interface online or in Appendix 2. The survey consisted of five patient cases developed by domain experts to cover multiple different use cases for heparin. A doctor would first agree to a research disclosure, then fill out a short questionnaire about their background and relevant experience (Figure 5). Next an instruction page was displayed explaining the survey (Figure 6, 7). Once the doctor read over the instructions and guidelines, they were presented with their first case which

consists of some lab results and a short admission note (Figure 8). It was their task to determine a potential infusion rate based on traditional heparin dosing guidelines that were available to them. After choosing an infusion rate, the optimal dose model was displayed (Figure 9). Unlike the standalone calculator, the survey version of the graph does not present the weight-only model. It does, however, plot the user's initial infusion rate on the estimated therapeutic curve as a black dot, so they can visualize the estimated dose versus their own dose (Figure 3). The doctor then had the option of either adjusting the dose or sticking with their initial dose. Whenever dosing the doctor was able to leave notes to explain their reasoning. This task then was repeated for the remaining four patients. After completing the test, there was a short follow-up questionnaire consisting of a few questions to gauge receptiveness to the tool and measure any change in confidence when using the tool. Additionally, participants were asked for other places they thought the tool could be used (Figure 10). Due to financial and scope limitations, we could not compensate participants for their time taking the survey. As a result all volunteers were acquaintances or friend's of friends of the author.

4.2 Architecture and Implementation

The survey was implemented using the same technological stack as the calculator (Angular.js and Bootstrap on the front-end, Python and Flask on the back-end). By utilizing a SPA (single page application) methodology we were able to quickly iterate and make many revisions. There were several candidate versions of the survey developed and tested with domain experts for feedback. One initial version had the user adjusting the heparin infu-

sion rate over time, to more accurately simulate a patient's care (Figure 11). This technique was not used in the final version because we did not want to introduce the challenge of generating lab results dynamically overtime as a function of heparin infusion rate, and the patient's current and past conditions. The final iteration does not require the simulation of patient lab results and allows for the most direct input of comments from doctors, instead of relying on proxy metrics such as time to complete cases that are similar too each other.

5 Survey Results

5.1 Primary Findings

Participants adjusted their doses 72% of time after reviewing the model, which resulted in a higher probability of therapeutic outcome in 82% of the adjustments. Additionally, participants reported greater confidence when using the tool but also expressed some hesitation with adopting the tool, often due to a lack of understanding of the underlying models. More information follows. Access to the extended survey results can be found in Appendix 1.

5.2 Participant Profiles

In total nine participants, all medical doctors, completed the survey. They range in experience from 4 to 32 years practicing. Of the nine, one participant is a practicing physician in Ethiopia, and the rest practice in the United States. One participant completed the study from a mobile device and the

others from a traditional laptop or desktop computer. Five of the nine reported that they frequently prescribe heparin (at least once a week) in their current position. One participant reported prescribing it infrequently (once a month), two reported prescribing it rarely and one reported prescribing never. All participants reported that at sometime in the past they prescribed it frequently. The participant medical specialties can be found in Table 2, but ranged from hematopathology to neuroendovascular surgery. Although nine participants took the survey, one participant's (participant 9) infusion rates were disqualified from the data analysis because they were 100 times greater than the mean. The user is a practicing radiologist in the United States, who rarely prescribes heparin. Their comments in Table 6 note that they would like to see more feedback from the calculator.

Suggested IR (Units/Kg/HR)	**Patient 1** 19.7		**Patient 2** 13.3		**Patient 3** 15.9		**Patient 4** 14.4		**Patient 5** 17.5	
Participant	**IR 1**	**IR 2**	**IR 1**	**IR 2**	**IR 1**	**IR 2**	**IR 1**	**IR 2**	**IR 1**	**IR 2**
1	18	25	18	19	18	20	18	26	20	24
2	18	19.8	18	13.9	13	16	12	14.7	13	17.6
3	18	20.3	80	14.1	18	16.4	18	15	18	18
4	12	12	15	15	12	15.9	15	15	12	17.3
5	18	19.6	18	13.7	12	15.9	18	14.5	12	17.4
6	18	18	18	18	13	13	12	14.8	0	0
7	18	19.6	18	13.6	17	16	15	15	18	18
8	18	19.7	18	13.8	13	16	12	14.7	18	18
Min Dose	12	12	15	13.6	12	13	12	14.5	0	0
Max Dose	18	25	80	19	18	20	18	26	20	24
Standard Deviation	2	4	22	2	3	2	3	4	6	7
Mean Dose	17	19	25	15	15	16	15	16	14	16
Median Dose	18	20	18	14	13	16	15	15	16	18

Table 1: Infusion rates before and after optimal dosage model was displayed. IR1 refers to the initial infusion rate, and IR2 refers the infusion rate chosen after viewing the model.

5.3 Dose Differences

Each of the eight participants prescribed heparin twice to five different patients, yielding a total of 80 infusion rates. In Table 1, for each patient there is a suggested optimal infusion rate (IR) in the top header. This is the number that was displayed to the user after they entered their initial infusion rate *IR1*. *IR2* is the value they prescribed after viewing the optimal dosage curve and the suggested infusion rate.

In the 40 trials, participants choose only 11 times to maintain their initial infusion rate. In other words, 72% of the time participants adjusted their dose after viewing the model's suggested dose. In five of the 29 times that a change was made, the change resulted in lower probability of a therapeutic outcome than the original dose. All five of these events occurred with participant number one, which may reflect the user struggling to understand how to use the tool. Unfortunately, due to the nature of this study, we cannot determine what an ideal dose should be, as many guidelines exist for recommending infusion rates, and there may exist multiple infusion rates that result yield therapeutic results.

As for the reasoning for adjusting doses, they vary greatly as can be seen in the appendix in Table 9. Those who choose to adjust towards the model values generally expressed trust in the model, while some users, in particularly those who currently prescribe heparin frequently were more wary of it. Some users explained they wanted to adjust the infusion rate based on features that were not considered in the model. Others expressed they wanted to adjust dosage using the Anti-Xa Assay (an assay designed to monitor anticoagulant therapy (Reka G Szigeti 2014).) Additionally, others expressed

that their initial dose was close enough, and that they preferred round numbers, since administering doses in non-integer units is not always possible. These results break down into two points of consideration. One is the user's lack of faith in the model, which is understandable considering it is not yet mature. The other set of concerns are perceived faults in the model, which come from a lack of understanding. This indicates an issue with usability and the communication of the model to the user.

5.4 User Feedback

Eight of the nine participants reported they felt more confident dosing with the tool than without. One participant said it did not affect their confidence. Although in general users reported greater confidence with the tool, the majority expressed concern with adopting it due to a lack of understanding of the models. As one participant noted the models are "a bit of a 'black box'." Other users noted that they were concerned the tool was not adjusting for some specific complications. Another user noted they were concerned about adopting it given there has not been a validation trial for the technique. These concerns and others will be addressed in the following section.

6 Retrospective and Next Steps

6.1 User Concerns

Although most participants expressed that the dosing tool gave them greater confidence, the majority also had reservations with adopting the tool in their

practice. Their common concern involved their understanding of how the tool worked. This concern is appropriate, as it reflects a conservative approach to experimenting on best practices in patient care. One reason for their lack of understanding can be attributed to the nature of this study, in that participants completed the survey on their own time and were given no incentives for completing the survey. As such, many users may have read the system explanation quickly.

Additionally, the technique used for dosing, although not terribly complex, is not trivial to understand. Physicians in general cannot be asked to learn how a logistic regression works or how these models are built. Simply speaking, their time is very valuable and limited. Thus, experimental techniques need to be validated in formal clinical setting with physicians who are invested in understanding and participating in a trial of the new tools. Then once the techniques are validated others can adopt them with greater confidence.

Another concern expressed was that the tool did not consider other potential features when dosing patients. Although these features can be added overtime, as the data becomes available and coded appropriately, this concern speaks to the greater issue of user interaction with the model. In its current state, the standalone calculator allows physicians to view the weight-only model and the all-features model (age, SOFA score, creatinine level, ethnicity, gender, etc.) This gives them a better understanding of how these features change the therapeutic curve. Future versions of this tool should allow the user to enable and disable a given feature, so they can have a better understanding of the effects of the feature on the model outcome. This understanding could potentially come in the form of additional visualizations like the original graph or arguably better in the form of written descriptions,

e.g. "The SOFA score causes a deviation of 3% from normal dosing." An initial version of this project included a crude nearest-neighbor visualization, but that feature was not incorporated in the final version due to time limitations. Increasing user interaction with the tool will hopefully address some knowledge and trust issues with the current system.

6.2 Shortcomings in the Survey

This study has no shortage of problems. The problems generally come from a lack of resources, in terms of time and money. A funded study would allow for a greater number of participants with specific ICU expertise. Additionally, the participant selection process would not be limited to friends of the researchers, which may have skewed the results. Funding would also allow for time for researchers to work with participants to better understand the model. Given the problems with the survey, and limited number of participants, the results are still very valuable as they clearly reveal areas for future improvement in the system.

One place for future research would be creating a similar survey that periodically provides sub-par dosage estimates for doctors. Then one could measure how frequently doctors correct the estimate. If doctors do not consistently catch dosing errors, that would indicate that increased dose validation software should be implemented.

6.3 Future Implementations

The experimental nature of this project has lead to a software implementation that is limited in several ways. The codebase served its purpose as a tool for a research study fairly well, but it is not "real world" ready in its current form. In future versions we'd like to see improved logging, system documentation, automated unit testing on builds, and better scaling performance. Additionally, the overall design should be more modular, such that it is easier to try out and compare different candidate models (e.g. Ensemble Learning, Neural Networks, SVMs, etc.) on different drugs. Automated testing could even run trial comparisons and compute statistics on models. From a data processing perspective, an entire subsystem should be designed and built for selecting data from the medical records database and translating it into a suitable form for models. There should also be tools for validating the cohort selection. This is not a trivial task when working with a specific medical record database, and becomes much more challenging when attempting to interface with other databases. Defining the generic interface used for selecting the cohort from the database will be challenging. Finally, the data will also need to be validated and then translated (coded) to appropriate values for a given model.

On the front-end, user analytics could be improved, as it would be ideal to know more about the user's device and how they interact on a per-click basis. One final improvement on the client would be modifying the data model, such that instead of requesting an estimate for a specific patient, the client requests a static model that has been precomputed on the server using the latest dataset. Then the user can manually modify features and the es-

timated probability curves are computed client side. This should allow for a much more responsive experience for the user, providing them instantaneous feedback for each feature change made, instead requiring a request to the server for each change. The dosage curves could be computed in a separate thread using web workers. Additionally, by implementing the predictions client side, patient data does not need to be transfered to an external server (easing HIPAA compliance), and the client can be fully functional offline which is ideal for low resource environments such as developing countries.

7 Conclusion

The central focus of this study was to implement an interactive data-driven dosage tool and test its usability with medical doctors. While doing this, we learned about domain specific, technical, and usability challenges. The goal of our work was to provide sufficient information such that a new version could be built and tested in a formal clinical setting. While the majority of participants felt more confident prescribing heparin using the tool, they also overwhelmingly expressed a need for a greater understanding of the underlying models making the predictions. This speaks to a usability problem that will be addressed in future versions by creating a more interactive, learning based tool. The future architecture will also open the door to dosing other drugs and trying out other models beyond logistic regression. These techniques will provide personalized insight in complex cases involving unique patient cohorts. Without question, in time, the computer's roll in prescribing drugs will grow, leading to improved quality of care for patients while decreasing hospital stay times and the cost of care.

Appendix 1: Full Survey Results

- Project Resources
  - Link to Client Side and Server Side Code Archive (Password Protected)
  - Link to Raw Survey Results Data

7.1 Presurvey Questionnaire:

1. What are your credentials? - MD - MD/PhD - NP - Other

2. What is primary medical specialty? - (e.g. Obstetrics)

3. How long have you been practicing in a clinical setting? - (e.g. 10 years)

4. In your current position how often do you prescribe heparin or other anticoagulants?
   - Frequently - (at least once a week)
   - Infrequently - (once a month)
   - Rarely - (once a year or so)
   - Never

5. In the past did you ever prescribe heparin or other anticoagulants more frequently? If so, how often?
   - Frequently - (at least once a week)

- Infrequently - (once a month)
- Rarely - (once a year or so)
- Never - I haven't prescribed heparin or other anticoagulants

6. Have you ever worked in an ICU? If so, to what extent?
    - (e.g During my residency I had 2 rotations lasting roughly 10 weeks each in different ICUs.)

Participant	Credential	Specialty	Years Practicing	Current Heparin Experience	Past Heparin Experience
1	MD	Pediatric Hematology	30	Infrequently	Frequently
2	MD	Hematopathology	32	Never	Frequently
3	MD	Neuroradiology	30	Rarely	Frequently
4	MD	neurosurgery, neurocritical care, clinical informatics	16	Frequently	Frequently
5	MD	Neurology	15	Frequently	Frequently
6	MD	Pediatric Critical Care	28	Frequently	Frequently
7	MD	Neurology	4	Frequently	Frequently
8	MD	Neuroendovascular Surgery	12	Frequently	Frequently

Table 2: Presurvey Results Questions 1-5

Participant	ICU Experience
1	I have served as a consultant in the PICU for 30 years
2	Surgery internship: vascular surgery patients - 2 mos
3	About 10 months of ICU during training - mixture of medical, surgical, and cardiac
4	In residency training x 7 years and as faculty x 9.5 years
5	Approximately 1 to 5 patients per month in ICU.
6	Residency (3 month long rotations), Fellowship (3 years), Attending (13 years)
7	I have bimonthly rotations there
8	No

Table 3: Presurvey Results Question 6

7.2 Follow Up Survey

1. Did you feel more confident prescribing heparin when using the models?

- I felt more confident using the models.

- I felt less confident using the models.
- The models did not affect my confidence.

2. What hesitations do you have about adopting statistical dosing tools?

3. Are there any specific areas you'd like to see a tool like this being used?

4. If you have any further comments, questions, suggestions, ideas for improvement, etc please enter them below.

	Q1	Question 2 - Hesitations
1	More	need a bit more data about renal function and overall status
2	More	You are stepping out of the realm of current standard practice. You are relying on 1) the data used in the model reflecting the patient group seen in your practice, 2) the statistical methods used being the most appropriate.
3	More	A bit of a 'black box' - I would want to see more about validation before fully implementing
4	Equally	would like to know more about the stats and the mechanisms that are responsible for the predictions
5	More	Lack of prospective trial using this method to provide validation of this tool
6	More	Still have to take in consideration other complications with regard to thrombus formation, bleeding risk, clearance (renal injury), other coagulation factors, etc.
7	More	The fractions in calculation makes hard to adminster we could round up for example 18.2 to 18 that way it would be more applicable
8	More	None
9	More	None; very helpful

Table 4: Follow Up Survey Responses Question 1 and Question 2

Participant	Question 3 - Other Areas
1	yes - in the ICU
2	Potentially this could be a great tool for patients on multiple drugs simultaneously - are there models that will help predict more accurately what type of dose adjustments should be made for patients on multiple heart failure meds e.g.
3	Other pharmaceuticals, e.g. antibiotics requiring levels (Vancomycin, Gentamycin, etc)
4	medical care of patients in general particularly for blood sugar, pain control & sedation, etc.
5	Warfarin dosing?
6	I think it would be helpful in the ICU as a guide.
7	Yes. Specially in fluid management in shock patients
8	protocolizeed approaches.
9	Treatment for PE, acute extremity thrombosis, stroke

Table 5: Follow up Survey Responses Question 3

Participant	Question 4 - Other Comments
2	It would be interesting to know if the dosing model works the same, or similarly among different patient groups (race, sex, BMI etc) and if not are there ways to adapt for these factors.,The dosing tool model is an excellent idea.
8	Create your own web page where people can go and use....similar to Angiocalc.com
9	Would be good to get one more feedback calculation after adjusted the dose for a confirmation since I made mistakes by a factor of 10

Table 6: Follow up Survey Responses Question 4

	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		
Participant	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**Total Time**
1	67	38	21	29	30	37	38	27	23	44	5 mins, 54 secs
2	136	42	49	45	53	33	98	26	112	19	10 mins, 13 secs
3	38	38	32	34	52	34	68	238	45	28	10 mins, 7 secs
4	23	69	50	53	237	79	96	32	86	72	13 mins, 17 secs
5	11	8	87	15	24	25	9	18	54	7	4 mins, 18 secs
6	55	60	25	84	38	21	85	46	51	22	8 mins, 7 secs
7	70	86	45	58	442	51	95	36	94	36	16 mins, 53 secs
8	135	83	54	40	29	19	29	18	44	50	8 mins, 21 secs
Min Time	11	8	21	15	24	19	9	18	23	7	
Max Time	136	86	87	84	442	79	98	238	112	72	
Standard Deviation	47	26	21	21	150	20	35	74	30	20	
Mean Time	67	53	45	45	113	37	65	55	64	35	9 mins, 39 secs
Median Time	61	51	47	42.5	45	33.5	76.5	29.5	52.5	32	7 mins, 51 secs

Table 7: Timing Results, Where T1 is amount of time in seconds between displaying the case and the doctors first dose is prescribed and where T2 is amount of time in seconds that the doctor viewed the model before submitting their second dose.

Participant	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
1	80	80	60	80	60
2	80	60	0	40	0
3	5000	5000	80	80	80
4	5000	6000	3500	6000	5400
5	80	80	0	80	0
6	80	80	0	44.5	0
7	80	80	0	60	80
8	80	80	0	60	80

Table 8: Bolus Information

Participant	Patient #	Adjust or Not	Why
1	1	Yes	need to get her there
	2	Yes	to get him there
	3	Yes	same reason as before - under dosed her
	4	Yes	need to get there
	5	Yes	to prevent further thrombosis
2	1	Yes	I'm relying on the model
	2	Yes	Model is probably better than my prediction as pt is pretty heavy
	3	Yes	model
	4	Yes	Bolus dose may well be too small
	5	Yes	I am relying on this model
3	1	Yes	model is close to my prediction and based on real data
	2	Yes	I actually made a mistake in my initial calculation!
	3	Yes	Makes sense, older pt may need less as indicated by model
	4	Yes	Graphs indicate higher likelihood of therapeutic result for this patient profile, and it seems reasonable to my initital estimate
	5	No	Close enough - no compelling reason to change
4	1	No	No PE yet. Would follow labs to see where she goes.
	2	No	has active PE. will adjust infusion based on individual response for Xa level.
	3	Yes	No clear emergency on type of dosing. Pt stroke is completed. Likely will not need a decompression unless bleeding occurs post heparinization.
	4	No	it's about the same and I'd rather be slightly high than low.
	5	Yes	if this is more likely to get to a therapeutic window quickly, would choose to do so.
6	1	No	Creatinine was elevated
	3	No	Risk of intracranial bleed
	4	Yes	Risk of continued thrombus greater than risk of bleeding
	5	No	Lack of evidence in this condition.
7	4	No	It is fairly equal. No meed to calculate the fractions
	5	No	Again only fractions
8	1	Yes	model said would be better.
	2	Yes	per model.
	3	Yes	per model.
	4	Yes	per model.
9	1	Yes	recommended
9	2	Yes	recommended

Table 9: Reasons participants choose to adjust or not to adjust their initial dose.

Appendix 2: Application User Interface

Below are screenshots of the calculator application and survey application in use. You encouraged to try them out yourself directly at https://hepstack-stage.herokuapp.com/.

7.3 Calculator Interface

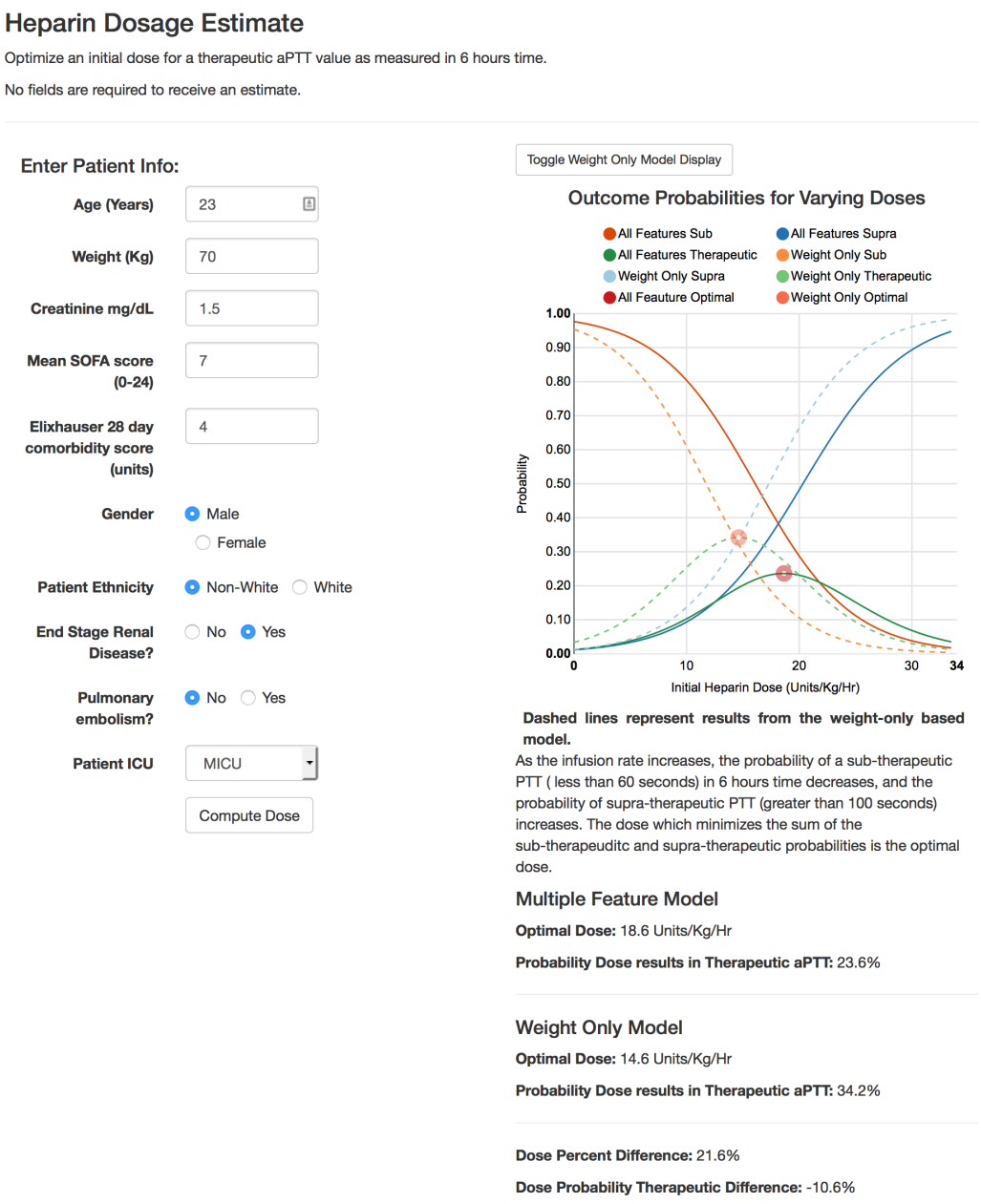


Figure 4: The standalone calculator user interface with patient data entered.

7.4 Survey Interface

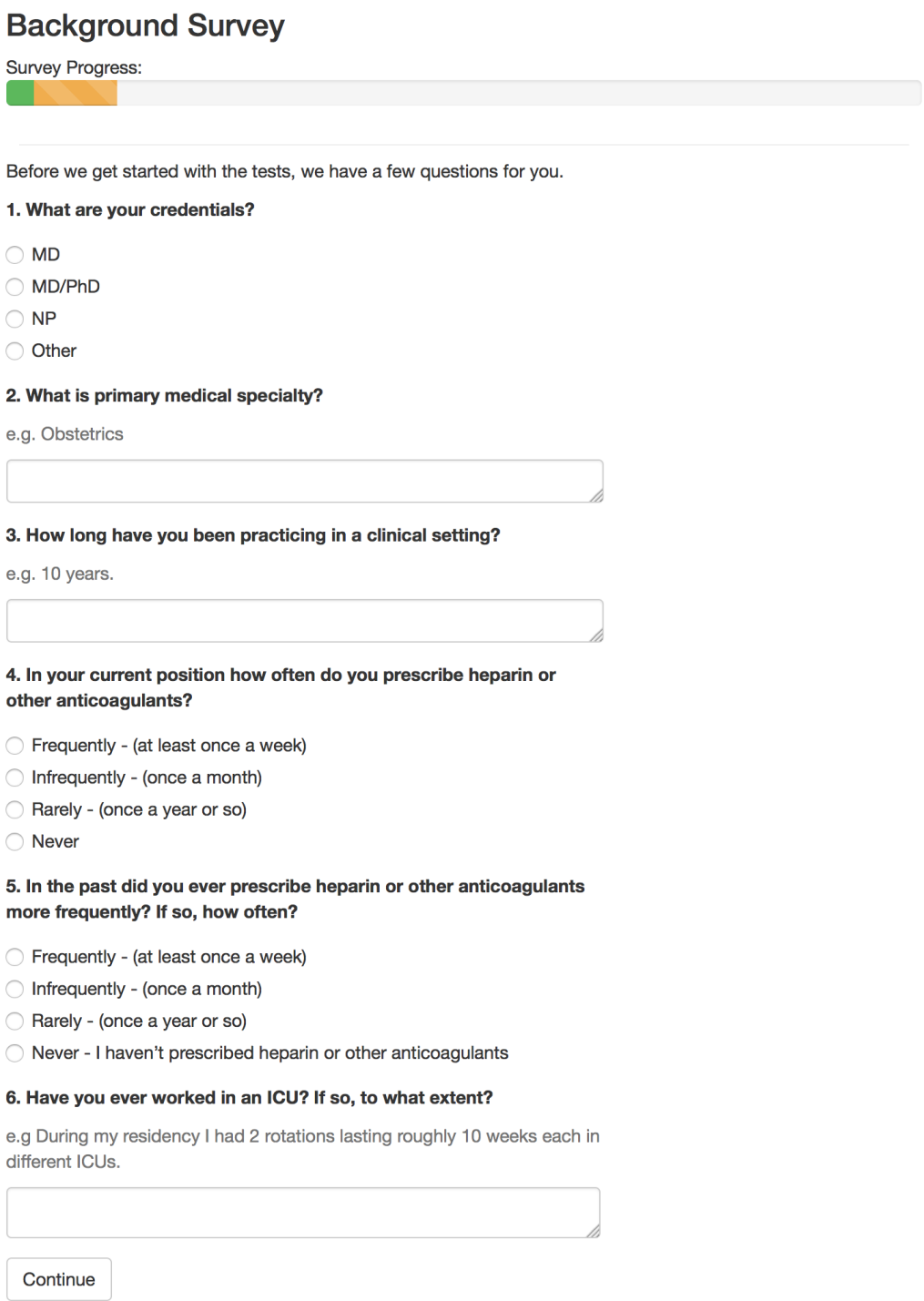


Figure 5: The questionnaire presented to study participants before they are able to dose patients.

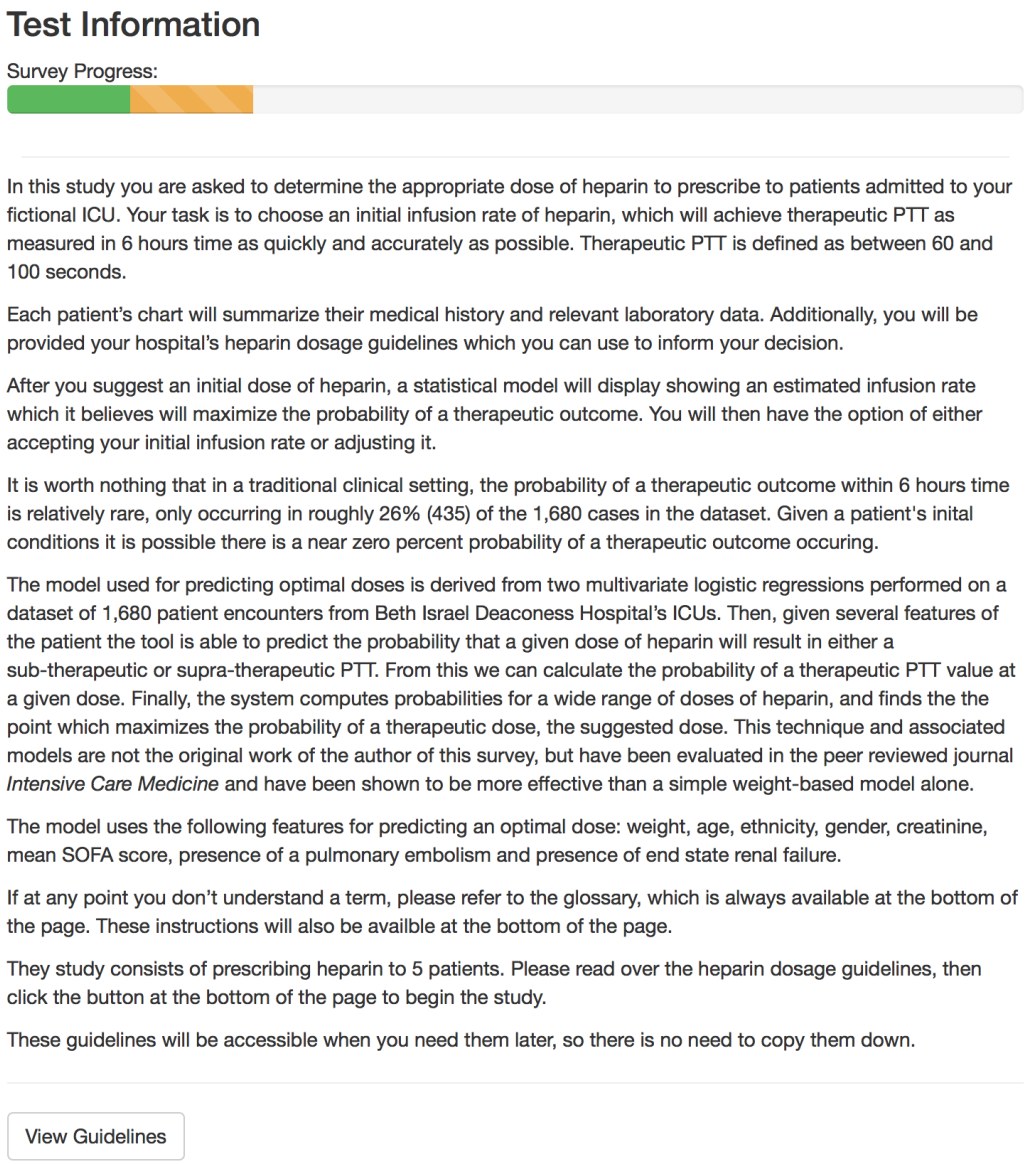
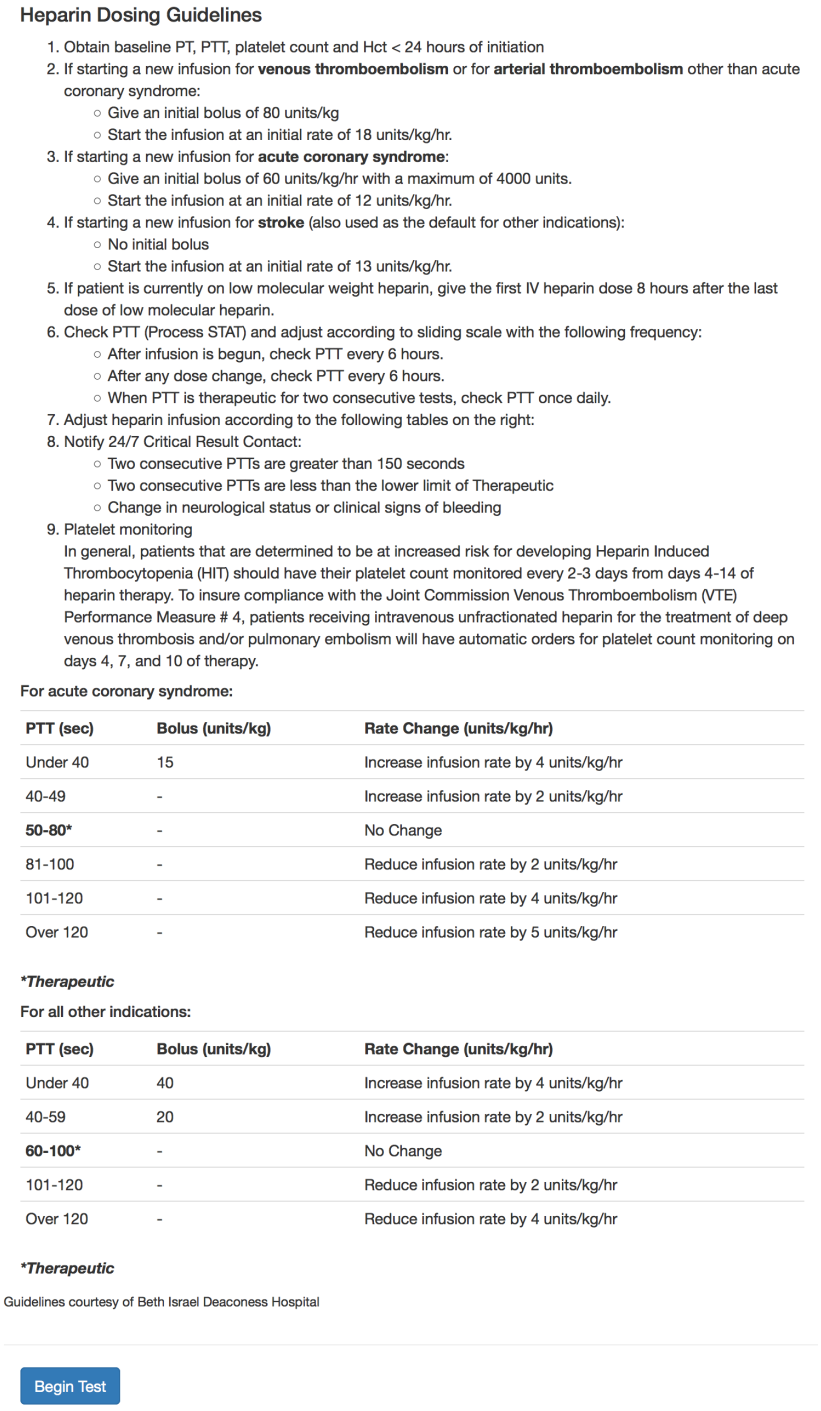


Figure 6: The first part of the information page outlining the requirements for the dosing task.



Heparin Dosing Guidelines

1. Obtain baseline PT, PTT, platelet count and Hct < 24 hours of initiation
2. If starting a new infusion for **venous thromboembolism** or for **arterial thromboembolism** other than acute coronary syndrome:
   - Give an initial bolus of 80 units/kg
   - Start the infusion at an initial rate of 18 units/kg/hr.
3. If starting a new infusion for **acute coronary syndrome**:
   - Give an initial bolus of 60 units/kg/hr with a maximum of 4000 units.
   - Start the infusion at an initial rate of 12 units/kg/hr.
4. If starting a new infusion for **stroke** (also used as the default for other indications):
   - No initial bolus
   - Start the infusion at an initial rate of 13 units/kg/hr.
5. If patient is currently on low molecular weight heparin, give the first IV heparin dose 8 hours after the last dose of low molecular heparin.
6. Check PTT (Process STAT) and adjust according to sliding scale with the following frequency:
   - After infusion is begun, check PTT every 6 hours.
   - After any dose change, check PTT every 6 hours.
   - When PTT is therapeutic for two consecutive tests, check PTT once daily.
7. Adjust heparin infusion according to the following tables on the right:
8. Notify 24/7 Critical Result Contact:
   - Two consecutive PTTs are greater than 150 seconds
   - Two consecutive PTTs are less than the lower limit of Therapeutic
   - Change in neurological status or clinical signs of bleeding
9. Platelet monitoring
   In general, patients that are determined to be at increased risk for developing Heparin Induced Thrombocytopenia (HIT) should have their platelet count monitored every 2-3 days from days 4-14 of heparin therapy. To insure compliance with the Joint Commission Venous Thromboembolism (VTE) Performance Measure # 4, patients receiving intravenous unfractionated heparin for the treatment of deep venous thrombosis and/or pulmonary embolism will have automatic orders for platelet count monitoring on days 4, 7, and 10 of therapy.

For acute coronary syndrome:

PTT (sec)	Bolus (units/kg)	Rate Change (units/kg/hr)
Under 40	15	Increase infusion rate by 4 units/kg/hr
40-49	-	Increase infusion rate by 2 units/kg/hr
50-80*	-	No Change
81-100	-	Reduce infusion rate by 2 units/kg/hr
101-120	-	Reduce infusion rate by 4 units/kg/hr
Over 120	-	Reduce infusion rate by 5 units/kg/hr

Therapeutic

For all other indications:

PTT (sec)	Bolus (units/kg)	Rate Change (units/kg/hr)
Under 40	40	Increase infusion rate by 4 units/kg/hr
40-59	20	Increase infusion rate by 2 units/kg/hr
60-100*	-	No Change
101-120	-	Reduce infusion rate by 2 units/kg/hr
Over 120	-	Reduce infusion rate by 4 units/kg/hr

Therapeutic

Guidelines courtesy of Beth Israel Deaconess Hospital

Begin Test

Figure 7: The standard heparin dosing guidelines participants used throughout the study. The guidelines were always accessible when a participant was asked to prescribe heparin.

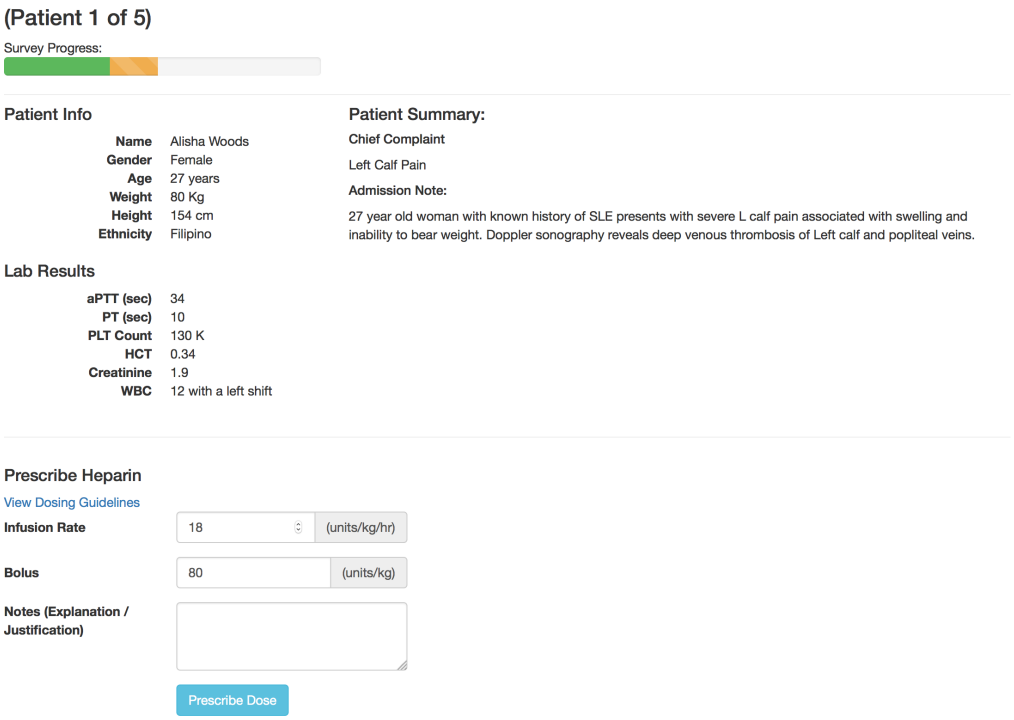


Figure 8: The patient profile and dosing page. In this image the participant has entered in values for the infusion rate and bolus.

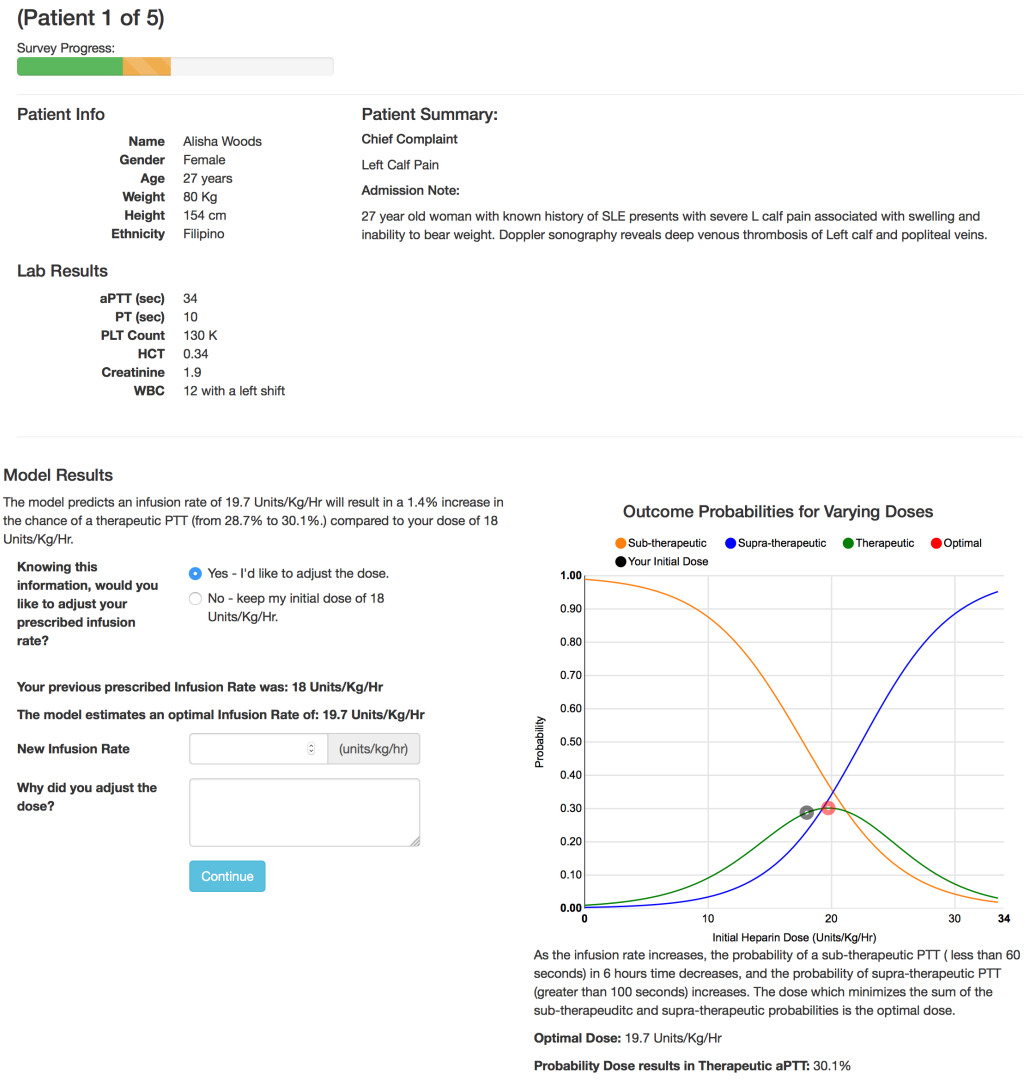


Figure 9: The page a study participant sees after selecting an initial infusion rate during the survey. Here the model is displayed and an interface allows them to adjust their initial dose. Note that the system plots the users initial dose (in black) on the therapeutic curve so the user can see how their dose compares to the best estimated dose.

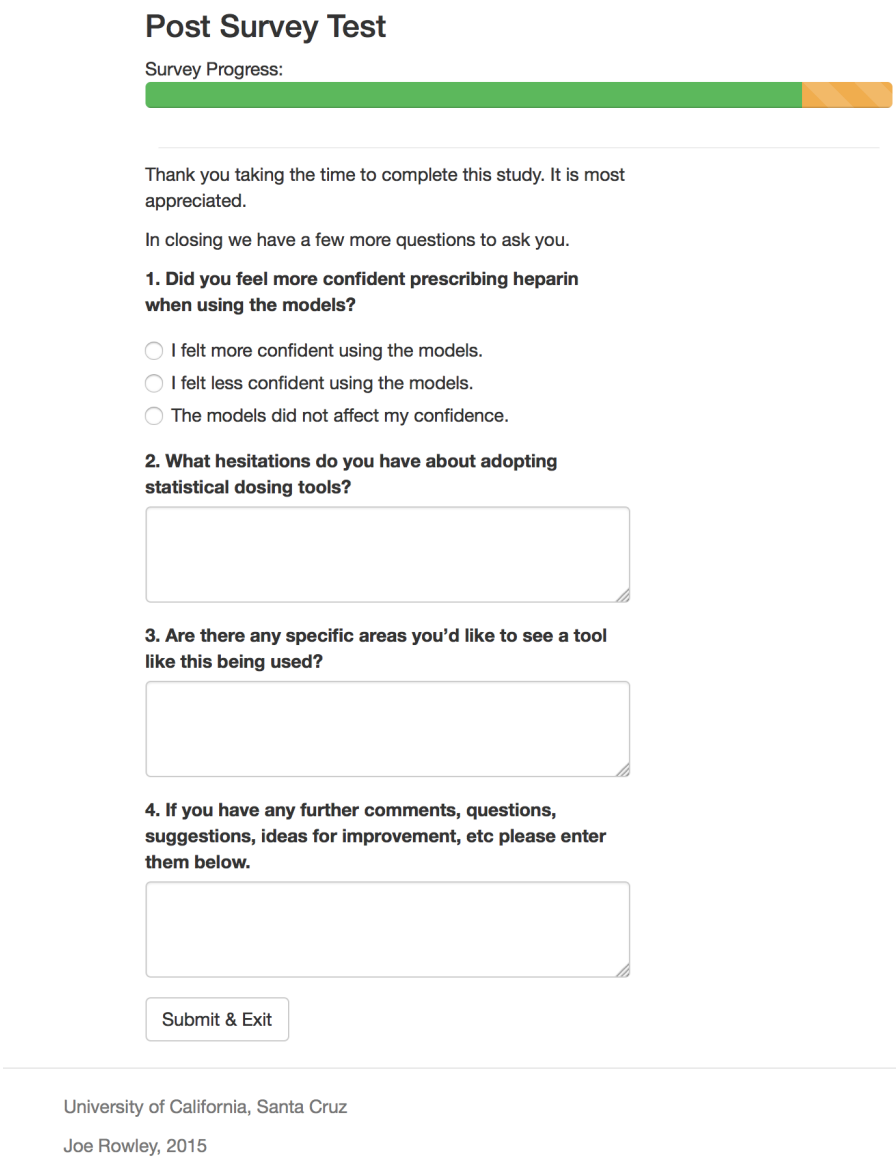


Figure 10: The questionnaire presented to study participants after dosing all patients in the previous section.

	May 01 11:00	May 01 16:00	May 01 22:00
PPT (sec)	30	45	60
PT (sec)	40	40	40
PLT Count	140 K	140 K	140 K
HCT	123	123	123
SOFA	12	12	12
Notes	This is a note 1	This is a note 1	This is a note 1
Medications	**May 1 11:00**	**May 1 16:00**	**Now: May 01 22:15**
Infusion Rate (units/kg/hr)	**25**	**20**	10
Bolus (units/kg)	**0**	**0**	3
Hold (mins)	**10**	**0**	4
Note	**test note**	**test note2**	Notes
			Prescribe Dose

Figure 11: An example of one of the interfaces created during testing. It allowed for extended testing with the same patient. This methodology was not used since a suitable technique for predicting aPTT outcomes for greater than 6 hours hasn't been studied significantly, and we wanted users to feel patient results were realistic.

References

Bayer, M., 2015. SQLAlchemy. Available at: http://www.sqlalchemy.org/.

Bostock, M., 2015. D3 jS. Available at: http://d3js.org/.

Ferri, C., Hernández-Orallo, J. & Salido, M.A., 2003. Volume under the rOC surface for multi-class problems. In *Machine learning: ECML 2003*. Springer, pp. 108–120.

Ghassemi, M.M. et al., 2014. A data-driven approach to optimized medication dosing: A focus on heparin. *Intensive care medicine*, 40(9), pp.1332–1339.

Google, 2015. Angular jS. Available at: https://angularjs.org/.

Group, 2008. PostgreSQL.

Jones, E. et al., 2015. Scipy documentation. Available at: https://docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.optimize.minimize_scalar.html.

Jones, E. et al., 2001. SciPy: Open source scientific tools for Python. Available at: https://www.scipy.org/.

Lisa Gryttenholm, et a., Dan Hendrickson, 2004. Guidelines for the therapeutic dosing of heparin. Available at: http://www.uwhealth.org/files/uwhealth/docs/pdf2/Heparin_Infusion_Guideline.pdf.

Partners, N., 2015. NVD3. Available at: http://nvd3.org/.

Pedregosa, F. et al., 2011. Scikit-learn: Machine learning in Python. *Journal*

of Machine Learning Research, 12, pp.2825–2830.

Reka G Szigeti, P., MD, 2014. Anti-xa assay. Available at: http://emedicine.medscape.com/article/2085000-overview.

Ronacher, A., Flask: A python microwebframework. Available at: http://flask.pocoo.org/.

Saeed, M. et al., 2011. Multiparameter intelligent monitoring in intensive care iI (mimic-ii): A public-access intensive care unit database. *Critical Care Medicine*, 39, pp.952–960.

Skipor, K., 2015. Angular-nVD3. Available at: https://krispo.github.io/angular-nvd3/#/.

Twitter, 2015. Bootstrap 3.0. Available at: http://getbootstrap.com/.

Washington Pharmacy Services, U. of, 2014. Heparin infusion guidelines. Available at: https://depts.washington.edu/anticoag/home/content/heparin-infusion-guidelines.