Personalized Medication Dosing via Sequential Regression
Appendix

1. Data Extraction and Pre-processing

Many features in the MIMIC database are nonuniformly sampled, which can complicate analysis. To address this, we generated a uniformly sampled approximation of the features by collapsing all measures to their nearest hourly value. Multiple measurements within the same time bin were then replaced by the median value of the set. We interpolated any missing hourly values using the *sample and hold* method, which we consider the most practical form of interpolation at the bedside, given the unknown nature of missing data in most clinical settings.

The intervals between aPTT measures characterize the response of the patient to the heparin dose within that interval. Hence, features were further collapsed from hourly measures to dosing interval measures. Specifically, all measures falling between aPTT draws were condensed into their median values in the intervals to provide a one-to-one correspondence between feature values and aPTT measures. We selected a minimum interval length of 6 hours.

2. Model Definitions

The ***population model*** is a single multinomial logistic regression with coefficients and features that apply to all patients. The ***individual model*** is a *set* of multinomial logistic regressions with *unique coefficients and features for each individual patient*. Both modeling approaches leverage the existing retrospective data to identify initial parameter values. However, the features and coefficients of the individual model are continuously updated as additional patient-specific data and features are collected. This property of the individual model improves its performance, and robustness to missing data, relative to the population model.

2.1 Weighted Multinomial Regression

We use a MNR model to estimate the probability of each patient, *i* , being in state *s* at a given dosing interval, *n*. Let X_i^n denote a $r_i^n \times c_i^n$ feature matrix with corresponding vector of outcomes denoted by $\boldsymbol{y}_i^n$; here c_i^n describes the subset of available features for an individual patient at dosing interval *n* and r_i^n represents the $n-1$ previous dosing interval data for the individual. Next, let X_p^n denote a $r_p^c \times c_i^n$ feature data matrix with corresponding therapeutic outcomes denoted by $\boldsymbol{y}_p^n$; here r_p^n describes the subset of patients in the population with the c_i^n features matching the individual patient. With this data, we estimate an individualized MNR model for each patient, at each dosing interval:

$$p(y_i^n = s \mid \boldsymbol{x_i^n}, \boldsymbol{\theta_i^n}) = \frac{\boldsymbol{e}^{\boldsymbol{x_i^n}^\top \boldsymbol{\theta_{i,s}^n}}}{\sum_{\boldsymbol{k=1}}^{\boldsymbol{3}} \boldsymbol{e}^{\boldsymbol{x_i^n}^\top \boldsymbol{\theta_{i,k}^n}}}$$

Where $\theta_{i,s}^n$ represents the maximum likelihood parameters of the model for state *s*. To

individualize each model, we define a function $\phi(n)$ that weights the importance of samples from the individual and population data when computing the maximum likelihood parameters of the MNR. In out case $\phi(n)$ was chosen to be sigmoidal

$$\phi(n) = \frac{\alpha}{1 + e^{-(\gamma_0 + \gamma_1 n)}}$$

Where the α and γ parameters control the shape and magnitude of the individual data weighting function. The precise values of α and γ are determined during model training using global optimization techniques. In our case, simple scatter search was deployed [8]. Finally, the weighted likelihood function is defined as:

$$\mathcal{L}(\theta_i^n) = \prod_{j=1}^{r_i^n} p(y_i^{(j)} \mid \boldsymbol{x_i^{(j)}}, \theta_i^n)^{\phi(n)} \times \prod_{k=1}^{r_p^n} p(y_p^{(k)} \mid \boldsymbol{x_p^{(k)}}, \theta_i^n)$$

Where $\boldsymbol{x}_i^{(j)}$ and $y_i^{(j)}$ represent the individual patient's j'th dose interval features and therapeutic outcome and $\boldsymbol{x}_p^{(k)}$ and $y_p^{(k)}$ represent the k'th dose interval features and therapeutic outcome from the population. The likelihood function is then maximized via gradient descent to yield the optimal parameter values:

$$argmax_{\theta_i^n} \mathcal{L}(\theta_i^n \mid X_i^n, y_i^n, X_p^n, y_p^n)$$

2.2 From Probabilities to Dose Estimates

Recall that the therapeutic state of the patient can fall into one of three classes: sub-therapeutic, therapeutic, and supra-therapeutic. By setting the therapeutic state as the

will be a monotonically decreasing function of the UFH dose while *p(supra-therapeutic)* will be a monotonically increasing function of the UFH dose:

$$p(S_i^n = supra) = \frac{1}{1 + e^{-(\beta_{i,o}^n d_i^n + \kappa_{i,o}^n)}}$$

$$p(S_i^n = sub) = \frac{1}{1 + e^{-(\beta_{i,u}^n d_i^n + \kappa_{i,u}^n)}}$$

From these equations, the probability of a therapeutic dose is simply:

$$p(S_i^n = therapeutic) = 1 - p(S_i^n = supra) - p(S_i^n = sub)$$

In the above equations, $\beta_{i,o}^n$ and $\beta_{i,u}^n$ are the maximum likelihood parameters from θ_i^n that model the effects of an individual patient's medication dose, d_i^n, on the probabilities of supra-therapeutic and sub-therapeutic states respectively. $\kappa_{i,o}^n$ and $\kappa_{i,u}^n$ are scalars that reflect the cumulative effects of the other $c_i^n - 1$ selected features on the probabilities of supra-therapeutic and sub-therapeutic states respectively.

The optimal dose at each dose adjustment interval, *n,* then corresponds to the dose value that jointly minimizes the probability of supra-therapeutic and sub-therapeutic probabilities. Given that $p(S_i^n = supra)$ is monotonically increasing with dose, and $p(S_i^n = sub)$ is monotonically increasing with dose, the optimal therapeutic dose will always occur where the sigmoidal curves intersect with respect to d_i^n:

$$\frac{1}{1 + e^{-(\beta_{i,o}^n d_i^n + \kappa_{i,o}^n)}} = \frac{1}{1 + e^{-(\beta_{i,u}^n d_i^n + \kappa_{i,u}^n)}}, \quad find\ d_i^n$$

Providing

$$d_i^n = \frac{\kappa_{i,u}^n - \kappa_{i,o}^n}{\beta_{i,o}^n - \beta_{i,u}^n}$$

3. Performance Metrics

We utilized two measures to evaluate the performance of our models. When comparing our models against the performance of clinicians, we utilized accuracy, and when comparing our models against each other, we utilized VUS.

We believe that accuracy is the correct metric to compare our models to the clinicians because we expect that the "prediction" of clinicians when they provide a UFH dose is always a therapeutic patient outcome. Given this assumption, the VUS of the clinician would always evaluate to chance, although their accuracy could be much higher. Hence, we compared of the *accuracy* of the clinicians against both the individual and population models.

A major goal of this paper was to highlight the advantages of an individual model compared to a population model. For general multi-class model-to-model comparison, VUS is an established, and useful metric. Hence, we believe that VUS and accuracy are the best-suited performance metrics for comparing our models to each other, and the clinicians in the *main text*.

4. Properties of the Collected Data

In Table E2 we provide additional information on Figure 1, including the total cohort size at each dose adjustment, after removing subjects with missing outcomes and features. In Table E4 we provide the probabilities of our patient's final recorded aPTT state given their first recorded aPTT state. In Table E5 we provides the joint probabilities of each starting and ending state combination.

In exploring our data, we found that the final aPTT recordings of 60% of patients were sub-therapeutic, 32% were therapeutic, and 8% were supra-therapeutic. We also found that while only 38% of all patients had a first and last aPTT measure that was sub-therapeutic, 72.2% of patient's that started with a sub-therapeutic aPTT also ended with a sub-therapeutic aPTT. Additionally, many patients with an initially therapeutic aPTT ended with a sub-therapeutic aPTT (45.5%). Interestingly, a majority of patient's with an initially supra-therapeutic aPTT also ended with a sub-therapeutic aPTT (46.8%). Another interesting observation was that while only 3.6% of all patients had a supra-therapeutic initial and final aPTT, 18% of patient's who started with a supra-therapeutic aPTT also ended with supra-therapeutic aPTT. These results provide some evidence that clinicians may not be adhering to the BIDMC guidelines, but are employing personal judgment to determine the preferred aPTT of the patient. This finding also motivates our proposed sensitivity analysis, where we removed patients whose final endpoint was sub-therapeutic, and re-validated the performance of the algorithm.

In further analysis we found that 61% of patient's whose final aPTT state was therapeutic were not so initially, 36% of patient's whose final aPTT state was sub-therapeutic were not so initially and 51% of patient's whose final aPTT was supra-therapeutic were not so initially. Hence, even if we assume that that the final recorded aPTT of the patient's is the "correct" value, these results indicates that the UFH dosing procedure

could be improved such that patients reach their intended individual endpoints more rapidly. These results are particularly important, as the goal of the individual model is to bring patients to the aPTT value defined by the clinician, regardless of what that definition might be. This is unlike the population model, which necessarily assumes a single definition of therapeutic aPTT for all patients.

In Table E6a we partitioned the patients according to their first and last aPTT state and provide the proportions of our dichotomous features for each patient subgroup. In Table E6b-c we partitioned the patient's according to their first and last aPTT state and provide the mean value of our continuously measured features for each patient subgroup. These 2s provide several interesting insights into the behavior of clinicians when dosing UFH. For instance, in Table E6a we see that 3% of patient's whose first and last recorded aPTT was subtherapeutic had pulmonary emboli compared to 14% of patients whose first and last aPTT was therapeutic, and 17% of patients who first and last aPTT was supratherapeutic. We also observe that 43% of patients whose first and last recorded aPTT was subtherapeutic were post-surgical patients, compared to 23% of patients whose first and last aPTT was therapeutic and 13% of patient's who first and last aPTT was supratherapeutic. This result reflects caution on the part of the clinicians when dosing post-surgical patients with UFH.

5. Model Comparison for Main Analysis

The results of the HL-test (Table 2) indicate poor calibration of the population and individual models, but we are not inclined to draw solid conclusions from these results. The HL-test can be an unreliable measure of model calibration because the results of the test vary dramatically with the number of groups selected by the investigator. Here we are reporting results for a group size of 22, but note that we observed large fluctuations in the outcome of the HL-test for various selections of group size (many of which indicated proper calibration). If nothing else, the HL-test indicates that the supratherapeutic component of the individual model is better calibrated than either population model.

The BIC of the individual model relative to the static-population model indicates that, for some patients, our selected continuously measured features may be unnecessary for the classification task. However, the improvement in the BIC of the individual model relative to the full-population model indicates that, given the same number of features, the individual model more effectively tunes model parameters to individual patients than the full-population model.

In Table 2 we provide metrics that compare the overall improvement in the performance of the individual model, relative to both the static- and full-population models. The NRI measure indicates that, when compared to the static-population model, the individual model is 2% more likely to correctly detect sub-therapeutic doses and 7.3% more likely to detect supratherapeutic doses. When compared to the full-population model, the individual model exhibited smaller NRI values, being only 1% more likely to detect sub-therapeutic doses, and 2.5% more likely to detect supra-therapeutic doses. Once again, while these improvement in performance are modest, they are significant because in a real care setting, the individual model would be applicable to all patients, while the full-population model would exclude nearly one quarter of all patients. The results of the IDI further validate

the utility of the individual model, relative to the population models. The difference in the average predicted probability of an overdose, if the dose was indeed too high, increased by 9% using the individual model relative to the static-population model and 2.5% when using the individual model relative to the full-population model. The difference in the average predicted probability of an underdose, if the dose was indeed too low, increased by 7% when using the individual model relative to the static-population model and 3% when using the individual model relative to the full-population model.

6 Model Comparison For Sensitivity Analysis

The HL-test indicated that both the static-population model and individual model were well calibrated ($p > 0.05$) while the full-population model was not ($p < 0.05$). The relative BIC values mirror those presented in our main analysis. The NRI measures from our sensitivity analysis indicate that, when compared to the static-population model, the individual model is 19% more likely to correctly detect sub-therapeutic doses and 11% more likely to detect supratherapeutic doses. When compared to the full-population model, the individual model was 13% more likely to detect sub-therapeutic doses, and 6% more likely to detect supratherapeutic doses. Gains were also observed for the IDI of the individual model, relative to the population models. The difference in the average predicted probability of an overdose, if the dose was indeed too high, increased by 10% when using the individual model relative to the static-population model and 4% when using the individual model relative to the full-population model. The difference in the average predicted probability of an underdose, if the dose was indeed too low, increased by 9% when using the individual model relative to the static-population model and 4% when using the individual model relative to the full-population model. The relative improvements in performance between the population and individual model were stronger than what we observed in our main analysis.

7. Deployment Challenges and Opportunities

Cases other than heparin dosing where our approach may be helpful are readily available. For example, warfarin therapy for atrial fibrillation relies on serial dose adjustments in response to international normalized ratio (INR) values, with a typical targeted therapeutic range of INR between 2 to 3. However, maintaining therapeutic INRs is notoriously difficult, with one recent study finding that INR values are out of range more than 1/3 of the time[1]. As another example, phenytoin is among the most commonly prescribed anti-seizure medications, and serial monitoring of serum phenytoin levels is part of standard care. Phenytoin pharmacokinetics is well known to be complex, and exhibit non-linear dose response patterns. Serum levels can also vary dramatically depending on nutritional status, hydration, age, and other factors, particularly in vulnerable populations such as the elderly and critically ill hospitalized patients[2]. For both these cases, we predict that our methodology may be useful in reducing dosing errors and bringing patients to a therapeutic state more rapidly.

A discussion on the challenges and opportunities for deployment of our algorithm in an actual clinical setting is an important component of this work. There are three ways such a technology could be realistically implemented in a care setting. The first approach to deploying this technology is to build a publically accessible website, which would allow clinicians to provide feature information, and receive a dose-recommendation in response. We have already implemented a prototype of this type of dose recommendation algorithm (based on our prior work), which is publically at the following web-address https://hepstack-stage.herokuapp.com/#/calc]. Unfortunately, the speed of deployment for this first approach comes at the cost of its usability, as it requires clinicians to manually transfer a potentially long list of features from their EMR system into a web application in order to receive the

The second approach to deploying such technologies is to integrate the application directly into an existing healthcare IT ecosystem such that the dose-recommendation is automatically generated for each patient. Following integration, the clinical institution can then host a website on the secure institutional intranet, which contains a web interface for automated dosing that populates required features directly from the EMR. This second approach is unlikely to be rapidly deployed, but is significantly more usable, which is of key importance for the meaningful adoption of this kind of technology in a real clinical environment.

A third approach is to deploy the technology on an existing cross-institutional platform, such as SMART Health IT (SMART), which enables developers to generate applications that may securely interface with the EMR. The challenge with this approach is that there are no guarantees that a given institution will be SMART compatible. However, for projects in the research and development domain we believe that platforms like SMART provide a useful compromise between the cost deployment, and the ease of usability.

Prior work suggests that computer aided decision support systems can have a meaningful impact on quality of care. A meta-analysis by Gillaizeau et al. explored the effects of computer-aided medication dosing (the latest paper of a series of papers on the topic) and reported reduced lengths of hospital stay and improved cost effectiveness[3]. With regards to anticoagulants specifically, the authors reported that computer-aided procedures tended to decrease the time until therapeutic stability was attained, and significantly reduced thromboembolism events, although no significant effects on mortality were reported. Indeed, there are conflicting reports on the precise utility of decision support systems for medication dosing. A recent meta-analysis performed by Bayoumi et al. compared the effects of physician dosing and computer-aided dosing of anticoagulants across 14 randomized controlled trials[4]. The authors found no difference in the number of medication

errors, or duration of hospitalization, although a statistically significant improvement in the time spent in the therapeutic range was reported, which has been associated with clinical outcomes in other works[5]. Motivated by the conflicting reports on the efficacy of many computer-aided decision support systems, a recent meta-analysis by Miller et al. identified a host of important practical issues that have inhibited the proper utilization of computer aided decision support systems[6]. The most notable of these issues were (1) limited usability and (2) inadequate algorithms. Regarding usability, the authors found that many clinical trials evaluating decision support systems failed to reliably combine patient-data, system knowledge and clinician experience. The authors also explicitly highlighted the need for 'better algorithms', stressing that many tested systems utilize overly simplistic approaches, which in turn leads to mistrust of the recommendation by care providers, and diminished reliance. The authors also reported that many care providers find the output of statistical models difficult to interpret, preferring algorithms that utilize categories instead[7].

The issues of limited usability and inadequate algorithms represent a major barrier to system adoption and meaningful utilization. These issues were taken into considering when performing this study. To our knowledge, all existing computer-aided heparin dosing guidelines are based on population models that do not take advantage of the incoming data streams to improve performance. For these reasons, we believe that the patient-specific modeling approaches outlined in this study will be of interest to the research community, and will aid others in the development or deployment of their own individualized models for this, and other clinically relevant problems. Additionally, our method is designed to respect the judgment and intuitions of clinical experts by allowing them to specify an arbitrary aPTT range, for a given patient. Lastly, our algorithm can be leveraged to provide categorical outputs, to aid in interpretability.

8. BIDMC UFH 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:
- o Give an initial bolus of 80 units/kg
- o Start the infusion at an initial rate of 18 units/kg/hr.

3. If starting a new infusion for **acute coronary syndrome**:
- o Give an initial bolus of 60 units/kg/hr with a maximum of 4000 units o 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):
- o No initial bolus
- o 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:
- o After infusion is begun, check PTT every 6 hours.
- o After any dose change, check PTT every 6 hours.
- o When PTT is therapeutic for two consecutive tests, check PTT once daily.

7. Adjust heparin infusion according to the following sliding scale:

For **acute coronary syndrome**:

PTT (sec)	Bolus (units/kg)	Hold (min)	Rate Change (units/kg/hr)
Under 40	15	-	
40 - 49	-	-	Increase infusion rate by 4 units/kg/hr
50 - 80*	-	-	Increase infusion rate by 2 units/kg/hr
81 - 100	-	-	No change
101 - 120	-	30	Reduce infusion rate by 2 units/kg/hr
Over 120	-	60	Reduce infusion rate by 4 units/kg/hr
			Reduce infusion rate by 5 units/kg/hr

**Therapeutic*

For **all other indications**:

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

**Therapeutic*

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.

Features (N = 9684)	Mean	Standard Deviation	Missing Data (%)
Albumin	2.81	0.66	88.69
Blood Pressure - Diastolic	59.50	12.08	37.56
Blood Pressure – Systolic	118.44	21.44	37.56
Bilirubin	1.75	3.10	81.94
Blood PH	7.40	0.07	39.18
Respiration Rate	18.75	4.67	97.61
Sequential Organ Failure Assessment Score	5.92	4.00	11.81
Troponin	1.87	3.31	79.47
Obese (%)	35	-	40.66

Table E1: Summary statistics for features which were excluded due to high incidence of missing data (missing in more than 8% of the patients).

Dose Adjustment Interval	Number of Patients (Total)	Number of Patients (with complete data)	Patients by Category: (Sub/Therapeutic/Supra)
1	3883	2827	1458 / 749 / 651
2	3535	2727	1524 / 809 / 391
3	2643	2059	1188 / 692 / 173
4	1639	1308	721 / 480 / 106
5	747	593	320 / 229 / 41
6	197	151	84 / 57 / 9
7	22	18	12 / 5 / 1
8	1	1	0 / 1 / 0

Table E2: The size of our sample at over the eight dose adjustment intervals, before and after excluding patients with incomplete data.

Features	p-value
Static Features	
Age	0.37
Gender (Male)	0.77
ICU Type (Surgical)	0.77
Ethnicity	0.45
End Stage Renal Disease	0.99
Pulmonary Embolism	0.27
Continuously Measured Features	
Dose/Weight (units/kg)	0.99
Carbon Dioxide	0.42
Heart Rate	0.16
Creatinine	0.58
Glasgow Coma Scale	0.12
Hematocrit	0.91
Hemoglobin	0.92
International Normalized Ratio	0.98
Platlet Count	0.18
Prothrombin Time	0.80
Peripheral capillary oxygen saturation	0.41
Temperature (F)	0.67
Urea	0.11
White Blood Cell Count	0.92

Table E3: The univariate differences between the feature distributions of our sample before and after application of the exclusion criteria. A two-sample t-test was utilized for all continuous features while a Wilcoxon rank sum test was utilized for binary features. A p-value of < 0.05 indicates statistically significant differences.

Within Class Transition Matrix	Final aPTT Sub-therapeutic	Final aPTT Therapeutic	Final aPTT Supra-Therapeutic
Starting aPTT Sub-therapeutic	72.2%	23.4%	4.4%
Starting aPTT Therapeutic	45.5%	48%	6.5%
Starting aPTT Supra-Therapeutic	46.8%	34.8%	18.3%

Table E4: The probabilities of the Final aPTT state (columns), given the first recorded aPTT values (rows).

Overall Transition Matrix	Final aPTT Sub-therapeutic	Final aPTT Therapeutic	Final aPTT Supra-Therapeutic
Starting aPTT Sub-therapeutic	38.2%	12.4%	3.4%
Starting aPTT Therapeutic	11.9%	12.5%	1.6%
Starting aPTT Supra-Therapeutic	9.9%	7.2%	3.9%

Table E5: The joint probabilities of every possible starting, and final aPTT state.

Gender (% Male)	Ending Subtherapeutic	Ending Thereapeutic	Ending Supratherapeutic
Starting Subtherapeutic	62	57	57
Starting Thereapeutic	59	59	59
Starting Supratherapeutic	49	52	52
ICU Type (%Surgical)			
Starting Subtherapeutic	43	31	31
Starting Thereapeutic	31	23	23
Starting Supratherapeutic	25	13	13
Ethnicity (%White)			
Starting Subtherapeutic	70	66	66
Starting Thereapeutic	67	67	67
Starting Supratherapeutic	67	64	64
End Stage Renal Disease (%)			
Starting Subtherapeutic	1	3	3
Starting Thereapeutic	3	2	2
Starting Supratherapeutic	3	3	3
Pulmonary Embolism (%)			
Starting Subtherapeutic	3	11	11
Starting Thereapeutic	8	14	14
Starting Supratherapeutic	9	17	17

Table E6a: The table partitions the patients according to their first and last aPTT state and provides the proportions of our dichotomous features for each patient subgroup

Age	Ending Subtherapeutic	Ending Thereapeutic	Ending Supratherapeutic
Starting Subtherapeutic	66.02(14.74)	67.44(14.84)	71.38(14.87)
Starting Thereapeutic	69.64(14.95)	67.62(15.07)	72.2(11.72)
Starting Supratherapeutic	70.76(15.25)	70.51(13.59)	69.23(15.86)
Dose/Weight (units/kg)			
Starting Subtherapeutic	10.99(3.95)	13.11(4.18)	13.08(5.02)
Starting Thereapeutic	11.48(3.54)	13(3.77)	11.95(4.16)
Starting Supratherapeutic	11.42(4.5)	11.59(3.98)	12.5(4.14)
Carbon Dioxide			
Starting Subtherapeutic	24.78(4.21)	24.94(4.18)	24.89(5.22)
Starting Thereapeutic	24.65(4.47)	24.77(4.71)	24.98(5.29)
Starting Supratherapeutic	24.02(5.26)	24.19(5.88)	22.92(4.94)
Heart Rate			
Starting Subtherapeutic	84.78(16.79)	85.5(16.94)	87.23(17.23)
Starting Thereapeutic	82.94(16.55)	84(17.56)	83.1(19.24)
Starting Supratherapeutic	84.8(16.87)	86.57(18.22)	86.21(17.95)
Creatinine			
Starting Subtherapeutic	1.47(1.4)	1.62(1.68)	1.7(1.36)
Starting Thereapeutic	1.49(1.4)	1.57(1.31)	1.72(1.65)
Starting Supratherapeutic	1.62(1.42)	1.77(1.64)	2.02(1.74)
Glasgow Coma Scale			
Starting Subtherapeutic	12.3(3.73)	12.79(3.31)	12.94(3.27)
Starting Thereapeutic	12.64(3.5)	12.73(3.46)	12.96(3.08)

Starting Supratherapeutic	11.82(3.73)	11.83(3.85)	12.09(3.77)
Hematocrit			
Starting Subtherapeutic	31.28(4.51)	32.07(4.7)	32.35(4.58)
Starting Thereapeutic	31.31(4.64)	32.15(4.87)	31.77(4.21)
Starting Supratherapeutic	30.81(4.37)	31.76(5.01)	30.69(4.63)

Table E6b : The table partitions the patient's according to their first and last aPTT state and provide the mean value of our continuous features for each patient subgroup.

Hemoglobin	Ending Subtherapeutic	Ending Thereapeutic	Ending Supratherapeutic
Starting Subtherapeutic	10.63(1.63)	10.75(1.68)	10.78(1.64)
Starting Thereapeutic	10.63(1.67)	10.81(1.74)	10.66(1.58)
Starting Supratherapeutic	10.36(1.57)	10.58(1.68)	10.21(1.64)
International Normalized Ratio			
Starting Subtherapeutic	1.37(0.53)	1.36(0.34)	1.79(3.38)
Starting Thereapeutic	1.43(0.45)	1.48(0.58)	1.54(0.44)
Starting Supratherapeutic	1.71(1.33)	1.69(0.9)	2.26(3.75)
Platlet Count			
Starting Subtherapeutic	226.82(126.32)	246.38(123.67)	238.16(104.95)
Starting Thereapeutic	210.46(89.65)	237.85(113.73)	224.81(107.78)
Starting Supratherapeutic	205.69(101.43)	232.3(131.48)	213.25(128.58)
Prothrombin Time			
Starting Subtherapeutic	14.54(3.31)	14.6(2.38)	15.68(5.99)
Starting Thereapeutic	15.06(2.99)	15.2(3.13)	15.45(3.02)
Starting Supratherapeutic	16.46(5.49)	16.48(4.88)	18.35(7.69)
Peripheral Capillary Oxygen Saturation			
Starting Subtherapeutic	97.27(2.09)	97(2.51)	97.24(2.15)
Starting Thereapeutic	97.21(3.53)	97.15(2.2)	97.35(2.27)
Starting Supratherapeutic	97.69(2.03)	97.06(4.2)	97.19(3.01)
Temperature (F)			
Starting Subtherapeutic	98.29(3.93)	98.43(1.25)	97.93(3.63)
Starting Thereapeutic	98.34(2.5)	98.35(1.31)	98.11(1.13)
Starting Supratherapeutic	98.2(1.36)	98.05(1.35)	98.14(1.4)
Urea			
Starting Subtherapeutic	29.59(22.11)	31.09(22.62)	32.57(20.71)
Starting Thereapeutic	29.56(21.72)	32.73(23.3)	33.85(27.79)
Starting Supratherapeutic	34.02(25.31)	36.88(25.55)	40.26(30.74)
White Blood Cell Count			
Starting Subtherapeutic	12.32(6.36)	11.97(4.94)	12.02(6.18)
Starting Thereapeutic	11.78(5.89)	11.94(5.45)	12.85(6.03)
Starting Supratherapeutic	12.24(5.66)	13.29(9.65)	13.12(7.5)

Table E6c: The table partitions the patient's according to their first and last aPTT state and provide the mean value of our continuous features for each patient subgroup.

References

1. Kvalseth TO. Cautionary note about R^2. The American Statistician. 1985: 39: 279-285.
2. Pokorney SD, Simon DN, Thomas L, Fonarow GC, Kowey PR, Chang P, Singer DE, Ansell J, Blanco RG, Gersh B, Mahaffey KW. Patients' time in therapeutic range on warfarin among US patients with atrial fibrillation: Results from ORBIT-AF registry. Am Heart J. 2015:170.1:141-8.
3. Durieux P, Trinquart L, Colombet I, Niès J, Walton R, Rajeswaran A, Rège WM, Harvey E, Burnand B. Computerized advice on drug dosage to improve prescribing practice. The Cochrane Library. 2013.
4. Bayoumi I, Al Balas M, Handler SM, Dolovich L, Hutchison B, Holbrook A. The effectiveness of computerized drug-lab alerts: A systematic review and meta-analysis. Int J Med Inform. 2014: 406-415.
5. Ageno W, Gallus AS, Wittkowsky A, Crowther M, Hylek EM, Palareti G; American College of Chest Physicians. Oral anticoagulant therapy: antithrombotic therapy and prevention of thrombosis: American College of Chest Physicians evidence-based clinical practice guidelines. *CHEST*. 2012: e44S-e88S.
6. Miller A, Moon B, Anders S, Walden R, Brown S, Montella D. Integrating computerized clinical decision support systems into clinical work: a meta-synthesis of qualitative research. Int J Med Inform. 2015: 1009-1018.
7. Russ AL, Zillich AJ, McManus MS, Doebbeling BN, Saleem JJ. A human factors investigation of medication alerts: barriers to prescriber decision-making and clinical workflow. *AMIA Annual Symposium Proceedings*. 2009.