Personalized Medication Dosing Using Volatile Data Streams

Mohammad M. Ghassemi*[1], Tuka Alhanai*[1], M. Brandon Westover[2], Roger G. Mark[1], Shamim Nemati[3]
[1]Massachusetts Institute of Technology, Cambridge MA, USA
[2]Massachusetts General Hospital, Boston MA, USA
[3]Emory University, Atlanta GA, USA
*Equal contribution

Abstract

One area of medicine that could benefit from personalized procedures is medication dosing. Mis-dosing medications may incur additional morbidity, or unnecessarily increase the length of patient stay. Here we illustrate a novel approach to personalized medication dosing that is robust to missing data, a common problem in the clinical care setting. We perform dose estimation using a novel take on multinomial logistic regression where model parameters are continuously estimated, for each patient, using a weighted combination of the data from a population of other patients, and a volatile data stream available from the individual under treatment. We evaluate our approach on 4,470 patients who received anticoagulation therapy during intensive care treatment. Our approach was 29% more accurate than intensive care staff, and better able to distinguish outcomes than a non-personalized baseline (0.11 improvement in model VUS, a multiclass version of AUC). The advantages of our approach are its ease of interpretation, robustness to missing features, and extensibility to other problems with similar structure.

Introduction

Over the last decade, there have been increasing calls to develop personalized approaches to patient care that better account for the complex factors influencing health (Reuben and Tinetti 2012; Mirnezami, Nicholson, and Darzi 2012; Barry and Edgman-Levitan 2012). In many studies however, "personalization" is performed at the level of static demographic features (such as age, gender, weight etc.) that are known at the start of care, and do not change over the course of treatment (Ghassemi et al. 2014). Moving forward, personalization will require on-line approaches that begin with simple demographic-level assumptions and become more patient-specific as additional data on the individual, and their response to treatment, is collected.

One area of medicine that could immediately benefit from personalized procedures is the dosing of medication with narrow therapeutic windows. Studies estimate that medication errors are responsible for up to 400,000 preventable hospital deaths each year (James 2013), but the potential complications associated with medication mis-dosing are usually more nuanced than life or death. Patients who are over- or under- dosed can experience unnecessarily complications, extended length of hospital stay, or require additional follow-up interventions, thereby driving up costs and incurring additional morbidity. Hence, even if the consequences of misdosing are limited, procedures should minimize the total number of dose-adjustments needed to reach the therapeutic targets defined by the care provider.

In recent years, a myriad of promising medication dosing models have been proposed (Roswell et al. 2016; El-Solh 2004; Hohner et al. 2015; Levi et al. 2007; Fleischman, Shin, and Li 2016). However, most of the proposed approaches require data streams and/or features to be consistently available and this requirement may be unrealistic in actual care settings. Even within the same hospital, different care providers record different physiological signals, for different patients, at different times. This practical reality of clinical data has increased the difficulty of robustly translating many promising observational models, into practical applications that improve care at the bedside (Johnson et al. 2016a).

In this work, we propose a simple solution to the problem of personalized medication dosing when feature information is not reliably collected. Our method proposes an initial dose based on demographic-level features collected at hospital admission, and prescribes subsequent dose recommendations based on an individual patient's available real-time data. Our method avoids the problem of missing features by re-estimating the maximum likelihood model parameters when feature information disappears. For practical deployment of such a model, interpretability is essential. For this reason, we used a linear modeling framework but also compare performance against state-of-the-art techniques (neural networks).

We tested the performance of our personalized-medication procedure on unfractionated heparin (UFH) - a drug with a narrow therapeutic window. UFH is typically initiated with a weight-based bolus, followed by a continuous infusion and *repeated adjustments* based on measures from a therapeutic indicator. Prior studies have shown that overdosing of UFH is associated with increased risk of internal bleeding, while under-dosing places patients at undue risk for embolic events, including pulmonary emboli, deep venous thrombosis, and ischemic stroke (Levine et al. 2001). Moreover, even when clinicians are adhering to stan-

dardized dose adjustment protocols, UFH is often misdosed, placing patients at unnecessary risk (Ghassemi et al. 2014; Alexander et al. 2005). Due to its sensitivity, risk profile, and frequency of use, UFH is a useful test-case for our personalized dosing strategy.

Methods

Data

Raw Data and Outcomes All data in this study was collected from the MIMIC critical care database (Johnson et al. 2016b). MIMIC is a de-identified, publicly available Electronic Medical Record archive containing over 40,000 unique Intensive Care Unit (ICU) admissions from the Beth Israel Deaconess Medical center (BIDMC) in Boston, MA from 2001 - 2016. We identified 4,470 patients from MIMIC who received intravenous UFH infusions during their ICU stay. aPTT was selected as our outcome in adherence with the guidelines of the BIDMC, where the data was collected.

At the time of an aPTT draw, patients may be categorized into one of three therapeutic states: Therapeutic, Subtherapeutic or Supratherapeutic. The continuous aPTT ranges used to define these states were:

$$\begin{cases} Subtherapeutic & aPTT \leq l \\ Therapeutic & l < aPTT < u \\ Supratherapeutic & aPTT > u \end{cases}$$

Where $u = 100$ and $l = 60$ describe the upper and lower bounds of the therapeutic state respectively. The aPTT therapeutic target ranges were defined by BIDMC's UFH dosing guidelines during the time of data collection (Appendix, section 8).

Feature Selection For the first 48 hours following each patient's UFH initiation, we extracted all available aPTT measures and features that are thought to confound the relationship between UFH and aPTT (Badawi, Oyen, and Haines 2004). Selected features are shown in Table 1.

Preprocessing We computed the median value of the included features in the intervals between recorded aPTT measures to provide a one-to-one correspondence between feature values and aPTT measures. Most patients have more than one dose-adjustment over the course of their ICU stay. For model development, we considered all UFH doses, and corresponding aPTT responses across patients as independent. Additional information on the pre-processing approach may be found in section 1 and Tables E2, and E3 of the Appendix.

High-Level Modeling Approach

We propose a categorical approach to dose estimation using multinomial logistic regression (MNR) where model features and parameters are re-estimated for each patient, at each aPTT draw, using a weighted combination of the data from a population of existing patients, and the individual patient's real-time data stream. We will refer to our modeling approach as the *individual model*. We will refer to a more classical approach, with features and coefficients that do not change, as *population models*.

Features (N= 9684)	**Mean**	**Standard Deviation**	**Missing Data (%)**
Static Features			
Age	68.01	14.91	0.00
Gender (%Male)	58	-	0.00
ICU Type (%Surgical)	35	-	0.00
Ethnicity (%White)	69	-	0.00
End Stage Renal Disease (%)	3	-	0.00
Pulmonary Embolism (%)	9	-	0.00
Continuously Measured Features			
Heparin Dose (units/kg)	11.79	4.11	6.88
White Blood Cell Count	12.26	6.35	6.23
Creatinine	1.58	1.48	5.18
Carbon Dioxide	24.61	4.67	5.69
Heart Rate (Mean)	84.81	17.12	0.01
Glasgow Coma Score	12.40	3.63	0.02
Hematocrit	31.50	4.65	4.27
Hemoglobin	10.63	1.66	6.45
Platelet Count	226.76	118.29	5.10
Urea	31.72	23.45	6.03
Temperature (F)	98.28	2.71	7.05
International Normalized Ratio	1.50	1.10	7.03
Prothrombin Time	15.22	3.99	0.12
Peripheral Capillary Oxygen Saturation	97.24	2.65	0.01

Table 1: Summary statistics of selected features collapsed across dosing intervals.

By incorporating patient-specific data, we expect enhanced performance of the individual model, compared to the population model. We also expect the individual model to be more robust to missing data as it utilizes only the subset of features available for each individual patient. For example, if the Creatinine of a patient were not known, the individual model would exclude this feature, while the population model would exclude the patient.

In Figure 1, we provides an illustrative depiction of the in-

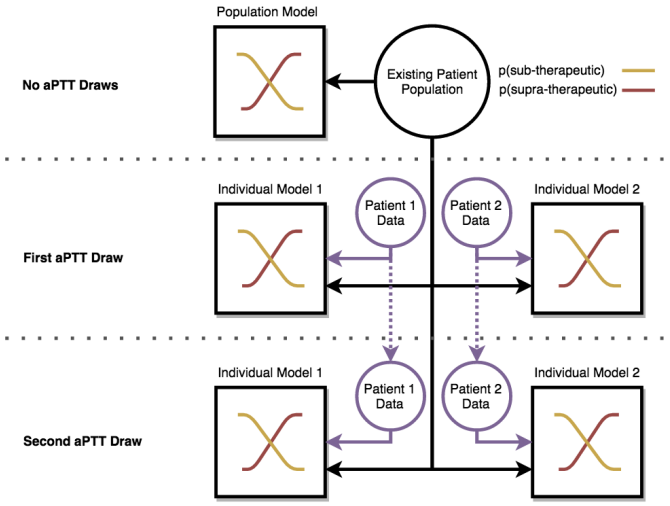

Figure 1: An illustration of our modeling approach for two patients. The population model is first computed using data from an existing patient population, before any individual aPTT draws have been measured. Following an aPTT draw, an individual model is created using a weighted combination of data from the existing population, and the individual patient's data stream.

dividual modeling approach. To promote reproducibility and extensions of this work, we will also include an implementation of our method in a publicly accessible online repository, following publication.

Formal Modeling Approach

Individualized 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 outcomes denoted by $\mathbf{y}_i^n$; 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 $r_p^n \times c_i^n$ feature data matrix with corresponding therapeutic outcomes denoted by $\mathbf{y}_p^n$; here r_p^n describes the subset of patients in 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:

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

where $\theta_{i,s}^n$ represents the 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 our case $\phi(n)$ was chosen to be sigmoidal:

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

Where the α and γ hyper-parameters control the shape and magnitude of the individual data weighting function. Finally, the weighted likelihood function is defined as:

$$\mathcal{L}(\theta_i^n) = \prod_{j=1}^{r_i^n} p(y_i^{(j)}|\mathbf{x}_i^{(j)}, \theta_i^n)^{\phi(n)} \times \prod_{k=1}^{r_p} p(y_p^{(k)}|\mathbf{x}_p^{(k)}, \theta_i^n)$$

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

$$\underset{\theta_i^n}{\text{argmax}}\ \mathcal{L}(\theta_i^n|X_i^n, \mathbf{y}_i^n, X_p^n, \mathbf{y}_p^n)$$

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. After setting the therapeutic state as the reference class for the MNR and accounting for the effects of covariates, $p(sub - therapeutic)$ will be a monotonically decreasing function of the UFH dose while $p(supra - therapeutic)$ will be a monotonically increasing function of the UFH dose:

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

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

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

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

Hence, the optimal dose at each interval, n, then corresponds to the dose value that jointly minimizes the probability of supra-therapeutic and sub-therapeutic probabilities. Given that $p(S_i^n = supratherapeutic)$ is monotonically decreasing with dose, and $p(S_i^n = supratherapeutic)$ is monotonically increasing with dose, the optimal therapeutic dose will always occur where the 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:

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

Which is the dose with the optimal probability of yielding a therapeutic state.

Performance Characterization

We compared the performance of the individual model against four alternative population models.

Baselines 1&2: Multinomial Logistic Regression

The first baseline model was a multinomial logistic regression which utilized all the selected features seen in Table 1. We will refer to this model as the *full population model*. The second baseline was a multinomial logistic regression that utilized only the continuously measured UFH dose and the six features listed under the static features heading in Table 1. We will refer to this as *static-population model*.

We expect the performance of the full population model to be superior to the static population model. Unfortunately, the improved performance of the full population model comes at the cost of the model's applicability, as individual patients may be missing one or more of the features required to evaluate the model. In our case, the full population model excludes nearly a quarter (23.6%) of all patients due to missing features. The static population model only utilizes features from Table 1 without any missing data, which improves its general applicability, but may also decrease its performance.

Baseline 3: Feed-Forward Neural Networks

A densely connected, feed-forward network with two hidden layers and a softmax output was also trained. Rectified linear units were used for the activation function of the hidden layer neurons. Network weights were initialized using Xavier initialization. The optimal network topology was identified via grid search (maximum hidden layer size of 20). Network performance was optimized on a validation set (15%) using scaled conjugant gradient backpropogation, with a cross-entropy performance metric.

Baseline 4: Reinforcement Learning

The problem of medication dosing may also be cast as as a partially observable Markov Decision Problem (POMDP), where the goal is to identify the dosing policy that prescribes optimal dosing actions, given known information on the state of the patient. To identify the policy, a reward signal must be specified. The solution for this problem may be solved via reinforcement learning. More specifically, the policy may be represented as a deep feed-forward neural network, whos weights are determined via deep deterministic policy gradient descent. For the purpose of policy identification, we defined the state, action, and rewards as follows: (1) State: The state of the patient as defined by their aPTT and laboratory measures (2) Actions: maintain dose, increase dose, decrease dose. (4) Rewards: a penalty proportional to the aPTT error. Importantly, the penalty of overdosing patients was twice the penalty of under-dosing. The value of actions were always normalized to sum to 1 to allow their comparison against other methods.

Validation

Models were validated using Leave-One-Out Cross Validation (LOOCV). That is, for each individual patient we trained a model where all other patients simulated a existing population at the hospital, and were used to train the population-component of the model, while the individual's data simulated an incoming data stream.

Model Comparison

To validate the individual model, we compared both its performance and applicability against the baseline models. The measure of performance we selected for model comparison was the Volume Under the Receiver Operator Surface (VUS), which is analogous to The Area Under The Receiver Operator Curve (AUC), but is used for evaluating the performance of classifiers with more than two states (Ferri, Hernández-Orallo, and Salido 2003). Additional performance metrics described by Cook et al. including AUC, net reclassification improvement (NRI), and integrated discrimination improvement (INI) were also computed (Cook and Ridker 2009).

Model-Clinician Comparison We also compared the performance of all models (individual and population) to the recorded predictive accuracy of the clinical staff. To do this, we assumed that the clinical staff intended to bring all patients into the therapeutic state as rapidly as possible when they administered or adjusted a dose of UFH (that is, clinicians did not intentionally over- or under-dose UFH). This assumption allows us to compare the predictive accuracy of the clinician, to that of our models.

Sensitivity Analysis There may be circumstances where the assumption that clinicians are aiming for a therapeutic dosing is invalid, particularly for patients at high risk for bleeding, where clinicians may intentionally under-dose UFH. To account for this, we repeated our analysis after excluding any patient whose final aPTT state after dose adjustment was sub-therapeutic, reflecting that the intention of the staff was in fact to under-dose the patient.

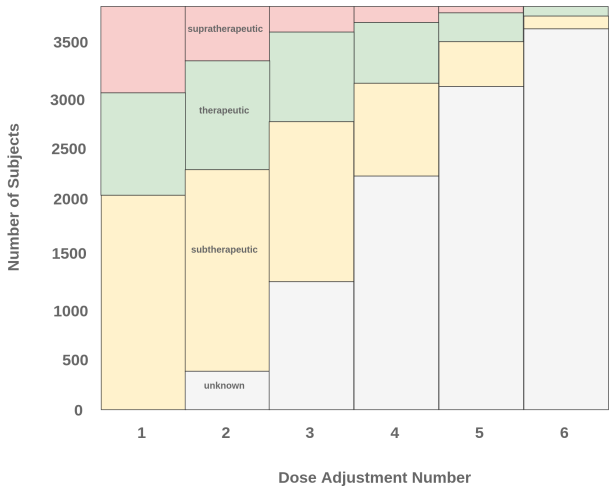

Figure 2: The aPTT outcome distribution of the patient cohort at each dose adjustment, and subsequent aPTT draw.

Results

Data Characteristics

In Figure 2 we show the cohort size at each dose adjustment interval, partitioned by aPTT therapeutic state. The figure highlights that UFH misdosing is consistently error-prone even after multiple aPTT draws (and consequent opportunities for dose adjustment). We observe that over 80% of our sample stopped receiving aPTT draws after their fifth adjustment and only 5% of the 3,883 patient with recorded aPTT values had a sixth dose adjustment. Please see section 4 in the Appendix for a detailed discussion of recorded dosing behavior with respect to other features in our data. Our final cohort for the full-population model, and neural network approaches included the data from the 2,827 patients without missing data, while the static population model and the individual model utilized all 3,883 patients with recorded aPTT values.

Main Results

In Table 2 we compare the overall accuracy, VUS and the percentage of subjects excluded by the models. Table 2 also evaluate the ability of the models to detect sub- and supratherapeutic dosing as measured by AUC. The improvement

	% Subjects Excluded (overall)	Volume Under Surface (overall)	Accuracy (overall)	AUC (sub- / supra)
Clinician	0%	0.17	31%	-
Population Models				
MultiNom Regression, static features	0%	0.35	57%	0.66 / 0.67
MultiNom Regression, all features	23.6%	0.41	58%	0.70 / 0.74
Feedforward Network (20x20), all features	23.6%	0.43	56%	0.72 / 0.74
Policy Network (20x20), all features	23.6%	0.44	58%	0.73 / 0.75
Individual Model				
MultiNom Regression, with available features	0%	0.46	60%	0.75 / 0.77

Table 2: Overall performance metrics for the models across LOOCV folds, compared to the clinician. VUS: Volume Under Receiver Operator Surface. AUC: Area Under The Receiver Operator Curve

in the VUS of the individual model compared to the full- and static-multinomial regression models was 0.05 and 0.11 respectively. The individual model had the highest overall accuracy of all tested approache and surpassed the best performing neural network by a VUS of 0.02. Compared to clinicians, the model was nearly twice as accurate.

In Table 3 we provide metrics that compare the overall improvement in the performance of the individual model, relative to both the static- and full- multinomial regression 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. 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. 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.

In addition to overall comparisons between our models, we also investigated differences in performance across dose adjustments. In Figure 3 (A-C), we compare the AUC of the individual model over time for the prediction of the sub-, supra- and therapeutic aPTT states at each dosing interval (solid lines) compared to the full-population model (dashed lines). Here again we observe improved performance of the individual model for the classification of all three states up until the sixth dose adjustment, after which the cohort size is significantly smaller (n<152) and our performance metrics become more sensitive to noise. In Figure 3(D), we compare the accuracy of the individual, and full-population model over time to the clinician where we also see that the individual model consistently outperforms the other approaches across time. Lastly, in Figure 4 we compare the dose estimation of the clinician, the full-population model, and the individual model for an exemplary patient.

	vs. Population Model with Static Features	vs. Population Model With All Features
Individual Model	(sub / supra)	(sub / supra)
NRI	2% / 7.3%	1% / 2.5%
IDI	7% / 9%	3% / 2.5%

Table 3: Performance metrics to compare performance of models at the detection of adverse events. NRI: Net reclassification improvement; IDI: Integrated discrimination improvement

Sensitivity Analysis

The results of our sensitivity analysis were similar to what we observed in the main results. The improvement in the VUS of the individual model compared to the full- and static-population models was 0.03 and 0.10 respectively. The differences in accuracy between the clinicians and our modeling approaches were significantly less pronounced in our sensitivity analysis. The individual model exhibited a 3% gain in accuracy compared to the clinicians. This result may be interpreted as evidence that clinicians are intentionally underdosing some patients, but without explicit knowledge of clinical intent, we cannot draw firm conclusions.

Discussion

Modeling Approach

Our approach is an on-line multinomial logistic regression for personalized dose-response estimation with volatile data streams. The method begins with data from a population of patients to estimate initial model parameters and sequentially, at each aPTT draw, incorporates weighted patient data to compute updated model coefficients. Importantly, this approach does not assume a constant weighting of individual patient data across dosing stages, but learns the parameters of an optimal weighting function.

Our method was designed to take an arbitrary therapeutic range, provided by a clinical expert, and to use this definition, and the available data to provide an optimal, personalized dose recommendation. Hence, it is easy to change the target of the algorithm to provide dosing recommendations for any desired therapeutic range (including a lower target for patients with high bleeding risk). In reality, dosing protocols for UFH can vary by institution, diagnostic group (e.g., venous thromboembolism vs acute coronary syndromes) and even monitoring parameters (e.g., targeting anti-Xa levels vs activated partial thromboplastin time (aPTT)). We have attempted to provide evidence that our method is capable of accounting for these differences.

Utilizing the most stringent form of model validation (LOOCV), we demonstrated that our approach enabled a

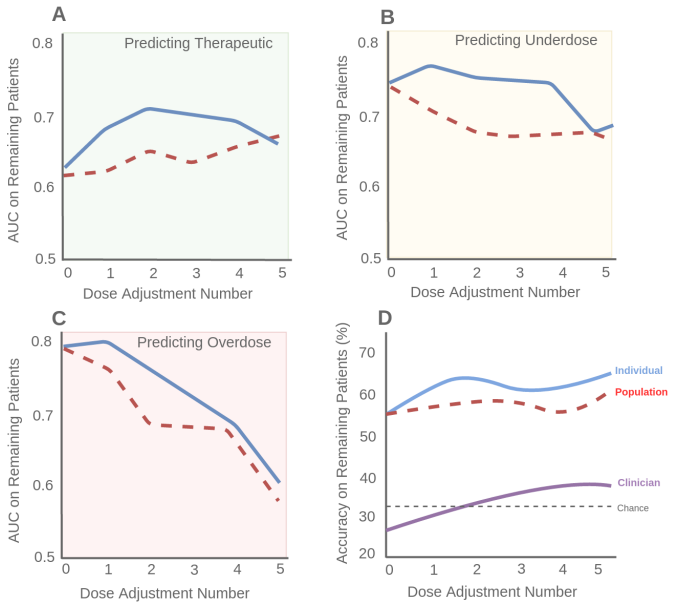

Figure 3: A comparison of predictive performance over sequential dose adjustments. Blue lines represent the performance of the individual model. Red dashed lines represent the performance of the population model. (A), (B), and (C), compare the AUC of the individual and population model's prediction of therapeutic dosing, underdosing and overdosing respectively. (D) compares the accuracy of the population model, the individual model and clinicians on patients after each dose adjustment.

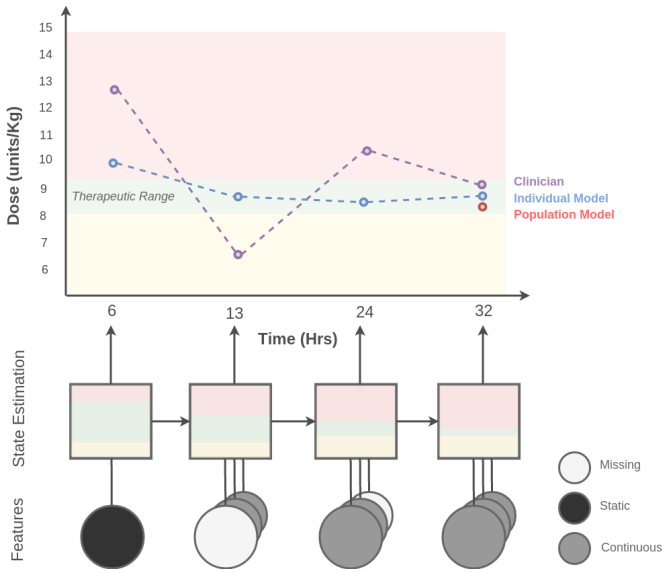

Figure 4: A comparison of the doses prescribed by a clinician, the population model, and the individual model for an exemplary patient. The clinician (purple line) alternates between over- and under-dosing the patients until 32 hours into the ICU stay. The population model (red dot) is unable to provide a therapeutic dose recommendation until 32 hours due to missing data. The individual model (blue line) identifies a correct dose after a single dose adjustment (hour 13), is robust to missing data, and continuously improves it's state estimates.

more robust estimation of the dose-response relationship compared to two population models. According to all computed metrics, the individual model is superior to the population models, and recorded performance of clinicians. The ultimate conclusion we drew from our main analysis is that the individual model is able to provide improvements in estimation of the dose-response relationship, without sacrificing applicability due to missing data, or interpretability due to model complexity (neural networks).

We believe that modeling approaches with the potential to be patient-specific are of interest to the medical community, and beyond. We are hopeful that this work will aid others in the development or deployment of their own individualized models for other problems. We stress here that the methodological contribution of the present work is in the sequential weighting of incoming data to inform model parameters, and not the choice of the model's form. Indeed, an SVM, Neural Network, or other technique may be applied using the same principal. To aid in the ease of replicating the work presented in this study, we have released a code repository that includes SQL queries for extracting data from the MIMIC database and Matlab code that run the individual and population models described in this work[1].

[1] https://github.com/deskool/Sequential-Regression-Heparin.git

Heparin

Heparin is the world's most commonly used anticoagulant. The precise dosing of heparin however is not as simple as once believed. It is now known that many individualized patient factors including race, gender, age, weight, severity of illness and ailment all confound the dose-response relationship (Ghassemi et al. 2014). These complications increase the difficulty of rapidly bringing patients to therapeutic levels of anticoagulation, and may (in the case of over-dose) increase the risk of bleeding.

Given the challenges of developing a unified dosing protocol (Hirsh 1991), many recent heparin-dosing studies are narrowly focused, investigating dosing regiments in the context of specific patient conditions such as renal replacement therapy, atrial fibrillation (Roswell et al. 2016), obesity (El-Solh 2004; Hohner et al. 2015), Sepsis (Levi et al. 2007) among others (Fleischman, Shin, and Li 2016). While it is clear from the recent literature that a singular approach to UFH dosing is insufficient, a fragmented set of protocols across a variety of conditions is also sub-optimal. To solve this problem, we proposed the use of the individual model described in this paper.

Limitations

The validity of these results hinge on an assumption that clinicians were dosing patients with an intention to achieve

the therapeutic aPTT (as defined by the institution). We acknowledge that this could be untrue in some cases. Patients with a high propensity for bleeding, for instance, may receive more conservative doses of UFH, leading to deliberate sub-therapeutic aPTT values. Whether the observed misdosings were intentional actions on the part of the clinician, mistakes, or simply a lack of adherence to hospital guidelines was beyond our ability to investigate with the dataset at hand and we acknowledge this as a weakness of the study. We attempted to address this weakness of the study by repeating our analysis on a subset of patients whose final state was not sub-therapeutic where we observed comparable results in the general performance of the models. Interestingly, the relative improvement of the individual model when compared to the population models in the sensitivity analysis was even stronger than what we observed in our main analysis, providing evidence that our results are reliable.

Deployment Details

A discussion on the challenges and opportunities for deployment of our algorithm in an actual clinical setting is an important component of this work. 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 (Miller et al. 2015). 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 (Russ et al. 2009). The issues of limited usability and inadequate algorithms represent a major barrier to system adoption and meaningful utilization and our method was designed to address these challenges. Additional discussion on deployment may be found in section 7 of the Appendix[2].

Conclusion

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 academic interest to the research community, and of practical interest to clinicians. We hope that our work will aid others in the development or deployment of their own individualized models for this, and other clinically relevant problems.

[2] https://github.com/deskool/Sequential-Regression-Heparin.git

Acknowledgements

M.M. Ghassemi would like to acknowledge the Salerno foundation, The NIH Neuroimaging Training Grant (NTP -T32 EB 001680), the Advanced Neuroimaging Training Grant (AMNTP T90 DA 22759). T. Alhanai acknowledges the Al-Nokhba Scholarship and the Abu Dhabi Education Council for their support. S. Nemati would like to acknowledge the support of the NIH early career development award in biomedical big data science (K01ES025445). R.G. Mark would like to acknowledge the support of the NIH Critical Care Informatics Grant (R01-EB017205) and the NIH Research Resource for Complex Physiologic Signals Grant (R01-GM104987).

References

Alexander, K. P.; Chen, A. Y.; Roe, M. T.; Newby, L. K.; Gibson, C. M.; Allen-LaPointe, N. M.; Pollack, C.; Gibler, W. B.; Ohman, E. M.; Peterson, E. D.; et al. 2005. Excess dosing of antiplatelet and antithrombin agents in the treatment of non–st-segment elevation acute coronary syndromes. *Jama* 294(24):3108–3116.

Badawi, O.; Oyen, L. J.; and Haines, S. T. 2004. Dosing of unfractionated heparin in acute coronary syndromes. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 24(12):1681–1691.

Barry, M. J., and Edgman-Levitan, S. 2012. Shared decision makingthe pinnacle of patient-centered care. *New England Journal of Medicine* 366(9):780–781.

Cook, N. R., and Ridker, P. M. 2009. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Annals of internal medicine* 150(11):795–802.

El-Solh, A. A. 2004. Clinical approach to the critically ill, morbidly obese patient. *American journal of respiratory and critical care medicine* 169(5):557–561.

Ferri, C.; Hernández-Orallo, J.; and Salido, M. A. 2003. Volume under the roc surface for multi-class problems. In *European Conference on Machine Learning*, 108–120. Springer.

Fleischman, J. W. E.; Shin, J.; and Li, F. 2016. Development of heparin protocols for patients receiving ventricular assist device. In *D45. CRITICAL CARE: CIRCULATORY HEMODYMANICS, SHOCK, CARDIOVASCULAR DISEASE, AND FLUID MANAGEMENT*. Am Thoracic Soc. A7065–A7065.

Ghassemi, M. M.; Richter, S. E.; Eche, I. M.; Chen, T. W.; Danziger, J.; and Celi, L. A. 2014. A data-driven approach to optimized medication dosing: a focus on heparin. *Intensive care medicine* 40(9):1332–1339.

Hirsh, J. 1991. Heparin. *New England Journal of Medicine* 324(22):1565–1574.

Hohner, E.; Kruer, R.; Gilmore, V.; Streiff, M.; and Gibbs, H. 2015. Unfractionated heparin dosing for therapeutic anticoagulation in critically ill obese adults. *Journal of critical care* 30(2):395–399.

James, J. T. 2013. A new, evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety* 9(3):122–128.

Johnson, A. E.; Ghassemi, M. M.; Nemati, S.; Niehaus, K. E.; Clifton, D. A.; and Clifford, G. D. 2016a. Machine learning and decision support in critical care. *Proceedings of the IEEE* 104(2):444–466.

Johnson, A. E.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L. A.; and Mark, R. G. 2016b. Mimic-iii, a freely accessible critical care database. *Scientific data* 3.

Levi, M.; Levy, M.; Williams, M. D.; Douglas, I.; Artigas, A.; Antonelli, M.; Wyncoll, D.; Janes, J.; Booth, F. V.; Wang, D.; et al. 2007. Prophylactic heparin in patients with severe sepsis treated with drotrecogin alfa (activated). *American journal of respiratory and critical care medicine* 176(5):483–490.

Levine, M. N.; Raskob, G.; Landefeld, S.; and Kearon, C. 2001. Hemorrhagic complications of anticoagulant treatment. *Chest Journal* 119(1_suppl):108S–121S.

Miller, A.; Moon, B.; Anders, S.; Walden, R.; Brown, S.; and Montella, D. 2015. Integrating computerized clinical decision support systems into clinical work: a meta-synthesis of qualitative research. *International journal of medical informatics* 84(12):1009–1018.

Mirnezami, R.; Nicholson, J.; and Darzi, A. 2012. Preparing for precision medicine. *New England Journal of Medicine* 366(6):489–491.

Reuben, D. B., and Tinetti, M. E. 2012. Goal-oriented patient carean alternative health outcomes paradigm. *New England Journal of Medicine* 366(9):777–779.

Roswell, R. O.; Greet, B.; Shah, S.; Bernard, S.; Milin, A.; Lobach, I.; Guo, Y.; Radford, M. J.; and Berger, J. S. 2016. Intravenous heparin dosing strategy in hospitalized patients with atrial dysrhythmias. *Journal of thrombosis and thrombolysis* 42(2):179–185.

Russ, A. L.; Zillich, A. J.; McManus, M. S.; Doebbeling, B. N.; and Saleem, J. J. 2009. A human factors investigation of medication alerts: barriers to prescriber decision-making and clinical workflow. In *AMIA Annual Symposium Proceedings*, volume 2009, 548. American Medical Informatics Association.