Learning Evolving Patient Risk Processes for C. Diff Colonization

Jenna Wiens JWIENS@CSAIL.MIT.EDU
John Guttag GUTTAG@CSAIL.MIT.EDU
MIT, 77 Massachusetts Ave., Cambridge, MA 02139 USA

Eric Horvitz HORVITZ@MICROSOFT.COM
Microsoft Research, One Microsoft Way, Redmond, WA 98052 USA

Abstract

Predictions of adverse events during hospitalizations can be used in programs aimed at improving patient outcomes. A patient's risk for adverse events may be biased by temporal processes influenced by diagnostic and therapeutic activities, as well as by the overall evolution of the patient's pathophysiology over time. Representing and reasoning about temporal process promises to enhance the accuracy of inferences about risk. However, understanding temporal influences is challenging for a number of reasons, including the large number of variables, the large class imbalance, and the difficulty of defining ground truth for risk over time. We explore such challenges in the context of predicting an inpatient's daily risk of becoming colonized with *Clostridium Difficile*. We present and evaluate different methods for extracting *risk processes* from medical records. These results highlight the benefit of including a temporal dimension when modeling patient risk.

1. Introduction

We consider the problem of predicting the risk of an inpatient becoming colonized with Clostridium Difficile (C. Diff) during a stay in the hospital. Although many of the risk factors are well known (*e.g.*, exposure, age, underlying disease, use of antimicrobial agents *etc.*), C. Diff continues to be a significant problem in many US hospitals. From 1996 to 2009, C. Diff rates for hospitalized patients aged $\geq$ 65 years increased by 200% (CDC, 2011).

Appearing in the *ICML 2012 Workshop on Clinical Data Analysis*, Edinburgh, Scotland, UK, 2012.

There are well-established clinical guidelines for predicting whether a test for C. Diff is likely to be positive (Katz et al., 1996). Such guidelines are largely based on the presence of symptoms associated with a C. Diff infection, and thus are not useful for predicting whether a patient will *become* infected. In contrast, risk stratification models aim to identify patients at high risk of becoming infected. The use of these models could lead to a better understanding of the risk factors involved and ultimately provide information about how to reduce the incidence of C. Diff in hospitals.

Tanner *et al.* developed a risk score to risk stratify patients for contracting C. Diff (Tanner et al., 2009). The score measures risk only at the time of admission. Dubberke *et al.* points out that the risk of a C. Diff infection may change from the time of admission, and sets out to develop and validate a C. Diff risk prediction model that could be used for predicting risk during hospitalization (Dubberke et al., 2011). It presents a model based on 11 variables developed using logistic regression. Although the authors consider a number of important variables when calculating risk on a given day, they do not include risk on previous days, *i.e.*, the temporal evolution of a patient's risk is ignored.

We hypothesize that incorporating a patient's evolving risk profile can lead to a more accurate model for predicting future infections. More generally, we propose and motivate the study of patient *risk processes* to model the evolution of risk over the course of a hospital admission.

To the best of our knowledge, representing and studying the risk of acquiring C. Diff (or any other infection) as a time series has not previously been done. There are a number of challenges associated with time series capture and analysis in this setting.

One major problem is the lack of ground truth about

risk in advance of a confirmed colonization. There are two issues here,

- Though it is likely that, on average, patients who contract C. Diff were at higher risk at some point during their hospital stay than patients who did not become colonized, we cannot assume that they were at a higher risk during each day of their stay.
- In practice, most risk stratification models are built using retrospective data. Because of this, it is challenging to acquire accurate daily labels for the data. In our case, not all patients get tested ever, let alone daily, for C. Diff.

Another difficulty is that there are often hundreds if not thousands of correlated variables, all changing at possibly different rates, that could affect a patient's risk. Finally, as with many risk stratification tasks the number of positive examples is low relative to the number of negative examples. In this paper, we describe and evaluate different approaches to grappling with these challenges.

1.1. Related Work

Researchers working in machine learning have proposed many different ways for taking into account the temporal dynamics of a problem.

Many argue that Hidden Markov Models (HMMs) are the most natural tool for dealing with sequential data. However, because the method is unsupervised one has limited control over what the hidden states correspond to (Rao & Hong, 2010). When working with high dimensional imbalanced data it can be especially difficult to learn relevant states in an unsupervised manner. Thus, despite the fact that HMMs seem to naturally fit the sequential data problem, many prefer discriminative methods such as support vector machines (SVMs) when working with complex high dimensional data.

When dealing with temporal data for classification, researchers commonly apply SVMs in one of two ways. The first technique maintains order by concatenating features from the last n observations (Shoeb et al., 2011). This approach works well when the notion of time does not vary tremendously across examples, *e.g.*, if the event of interest always occurs within a predictable window. For the challenge of predicting colonization with C.Diff, this is not the case, a patient may test positive any number of days into the admission and incubation periods can vary. The second technique involves *time-dependent features*. For example in financial time series forecasting, where the goal is to predict a stock price for the next day, features are commonly based on metrics calculated using data from the last t days (Rao & Hong, 2010). We consider a variation on this approach, including both static and dynamic variables in our feature vectors.

Our approach, described in Section 3, employs SVMs. Here we use SVMs first to reduce the high dimensional feature space. The outputs of the SVM when concatenated together for a given admission produce a time series, which we refer to as an approximate *risk process*. We refine this risk process using different weighted averages, smoothing the observations and refining the risk processes. In Section 4 we compare this approch to the currently prevalent approach of classifying patients based solely on their current observed state.

2. The Data

Our dataset comes from a large urban hospital in the US. We consider a study population of all in-patient admissions from one year. We restrict ourselves to stays $>= 7 days$, since we are interested in a patient's risk over time.

To ensure that we are in fact predicting the acquisition of C. Diff during the current admission, we remove patients who tested positive for C. Diff in the 60 days preceding or, if negative, following the current admission. In addition, we remove patients who tested positive for C. Diff before day 7 of the admission. Positive cases are those patients who test positive on or after 7 days in the hospital. Negative patients are all remaining patients (patients who either test negative or are not tested at all).

Typically, patients are first tested for the C. Diff antigen, and if a patient tests positive for the antigen this is followed up with a C. Diff toxin assay (Chapin, 2012), which is treated as definitive. We define the start of the risk period of a patient as the time of admission and the end of the risk period, according to the following rule: if the patient tests positive the first positive test marks the end of risk period, otherwise the patient is considered at risk until discharge. The final population consisted of 9,751 hospital admissions and 8,166 unique patients. Within this population 181 admissions had a positive test result for C. Diff antigen and 76 admissions had a positive test result for the C. Diff toxin.

3. Methods

Dubberke *et al.* developed a system for predicting risk in a real-time manner using logistic regression but they

did not perform a longitudinal evaluation of their system. In contrast, we consider risk at each day of a patient's admission. By analyzing the entire risk process of a patient we hope to get a more accurate prediction than if we had just considered the patient at a single point in time.

This section describes our methods in detail and is subdivided into three parts: 3.1 Feature Extraction, 3.2 Extracting Approximate Patient Risk Processes, and 3.3 Refining Patient Risk Processes.

First, in Section 3.1, we describe how we transform a patient's medical record into a sequence of feature vectors, one feature vector for each day. Medical records contain many categorical variables, *e.g.*, procedures performed and medications prescribed. We represent each possible categorical value as a binary feature. This results in a high dimensional feature space with approximately 10,000 features. Furthermore, many of the features are highly correlated with others. For example, a patient who had a cardioversion procedure is likely to be on a blood thinner. Given the high dimensionality and the potential correlations among the variables, we investigated a two-step approach to risk stratification.

In Section 3.2, we describe how we employ SVMs to reduce the dimensionality of the data. As a supervised learning method, SVMs require labels for each day of a patient's hospital admission. Since ground truth is unknowable in our context, we will discuss the assumptions made when generating these labels and our reasoning behind these assumptions. The daily outputs of the SVM when concatenated together for a hospital admission produce an approximate risk process. In Section 3.3 we refine these approximate patient risk processes, exploring different weighted averages to smooth out the predictions. In doing so, a patient's current risk is redefined as an average over previous risk approximations.

3.1. Feature Extraction

We extracted more than 10,000 variables for each day of every hospital admission. Approximately half of the features, 5,055 features, are based on data collected at the time of admission. These features remain constant throughout the stay. The remaining 5,542 features are collected over the course of the admission and may change on a daily basis. The types of features are listed in Table 1.

The majority of the features are represented by binary variables. We exploded categorical features such as financial class, and marital status into binary features.

Table 1. Static features are collected at the time of admission and do not change during the hospital admission while dynamic features can change on a daily basis. This table lists the features by type, and in parentheses the number of features associated with each variable.

Static Features	Dynamic Features
• prev. ICD9 codes (2513)	• lab results (2012)
• home medications (761)	• procedures (1293)
• prev. admission medications (655)	• location room (1209)
• patient's city (535)	• medications (872)
• attending MD (443)	• vitals (95)
• hospital service(39)	• location unit (61)
• admission source (22)	• day of admission (1)
• financial class code (19)	• unit CP (1)
• admission complaint (18)	• hospital wide CP (1)
• admission procedure (16)	
• patient's race (10)	
• patient's age (9)	
• patient's marital status (5)	
• patient's sex (1)	
• expected surgery (1)	
• ER admission (1)	
• dialysis (1)	
• diabetic (1)	
• history of C. Diff (1)	
• num hospital visits (90 days) (1)	
• avg., max., total los (90 days) (1)	

We binned and then exploded discrete features such as age. In the database, lab tests and vitals were flagged as: normal, low, high or critical based on reference values. We used these flags to represent the result of each test/vital as a binary variable. Admission complaint and admission procedure were both recorded as free text entries in the database. These entries were automatically mapped to categories corresponding to ICD9 codes (NCHS, 2008) using look-up tables, and then exploded into binary features. The remaining features: colonization pressure (CP), and statistics from the hospital admissions in the last 90 days are left as continuous variables. Given the number of binary features, it is not surprising that the feature vectors are quite sparse. On average less than 1% of the feature vector has a non-zero entry.

CP aims to the measure the proportion of patients, in a unit or hospital, colonized or infected with a particular disease. In our analysis, the contribution a patient, p, makes to the CP on day, t, depends on when the patient tested positive for the first and last time, t_f and t_l, and finally when the patient is discharged from the hospital, t_d. While the patient continues to test positive he or she contributes a constant amount to the

CP. After the last positive test result (which can be the first positive test result) a patient contributes to the CP for no more than 14 days. During this time period, the patient is assumed to be treated or in isolation, and we assume a linearly decreasing relationship. This function is summarized by Equation 1.

$$C_p(t) = \begin{cases} 1 & t \in [t_f, t_l] \\ -\frac{1}{14}t + \frac{(t_l+14)}{14} & t \in [t_l, min(t_d, t_l + 14)] \\ 0 & otherwise \end{cases} \quad (1)$$

Since we have time-stamped locations for each patient, we can calculate a colonization pressure for each unit, Equation 2. The colonization pressure of a unit depends on each patient's contribution to the colonization pressure on that day, $C_p(t)$ and each patient's length of stay in that unit on that day, $los(unit, p, t)$.

$$CP_{unit}(t) = \sum_p C_p(t) * \frac{los(unit, p, t)}{24} \quad (2)$$

When computing the *unit CP* feature for a patient on day t, we consider the CP in all of the units that patient spent time in on that day, i.e., $\sum_{unit \in los(unit,p,t)>0} CP_{unit}(t)$. We calculate the hospital wide CP by summing across all units. As a result, the unit CP varies across patients for a given day, while the hospital wide CP does not.

3.2. Extracting Approximate Patient Risk Processes

3.2.1. Labeling the Data

SVMs require labeled training data. Here, each day of a patient's admission is associated with its own feature vector. However, we do not have ground truth labels for each day of a patient's admission. We only know whether or not a patient eventually tests positive for C. Diff. Though it is safe to assume that, on average, patients who contract C. Diff. had been at higher risk at some point during their hospital stay than patients who are not colonized, we cannot assume that the patients were at a higher risk over each day of their stay. A patient's risk profile changes from day to day as he or she receives various treatments, contracts other diseases, or has increased or decreased exposure to the C. Diff bacillus.

We briefly explored a labeling in which we assumed that a patient was only at high risk x number of days before a positive index event, but choosing x proved difficult since this approach assumes that x is constant across patients. What if a critical event influencing the patient's risk actually happened at the time of admission? Looking only x days back may not capture this event. It is possible that any part of a patient's admission may contribute to their risk of later acquiring C. Diff.

Recognizing that the purpose of the SVM is to provide an approximate risk, which will be refined in the next stage of the analysis, we assign each day of an admission in which the patient eventually tests positive as positive, and negative otherwise. In doing so, we hope to identify high risk patients as early as possible. Since we do not expect a patient's risk to remain constant during an entire admission, there will be some noise in the training labels. For example, the may be some days that look almost identical in the feature space but have different labels. To handle this noise we use a soft-margin SVM, one that allows for misclassifications. As long as our assumption does not propose more incorrect labels than correct labels it should be possible to learn a meaningful classifier, despite the approximate labels.

3.2.2. Learning the Decision Boundary

From these approximately labeled feature vectors we learn a linear SVM $f(\mathbf{x})$. Applied to a test patient, $\mathbf{x_i}$ the SVM could classify each day as positive or negative, *i.e.*, the patient will eventually test positive or not. However, we do not use the SVM as a classifier but instead consider the continuous predictions made by the SVM: $\mathbf{w} \cdot \mathbf{x}_i - b$, *i.e.* the distance to the decision boundary. From the approximate binary training labels, we can derive continuous approximate risk scores for each day. The distribution of these risk scores on the training data is shown in Figure 1. As one can see, these scores do a reasonable job of separating positive and negative cases.

In the clinical literature, a common approach to validation of risk stratification metrics involves computing the patient's risk at a constant distance from the index event, and classifying the patient as high risk or low-risk according to some threshold. We refer to this risk stratification approach as the *Current State* approach as, for the prediction task of classifying a patient as high risk or low risk, one uses only the risk score of the most recent or current day and does not consider risk on previous days. Sweeping this threshold at two days prior to the index event results in an area under the receiver operating characteristic curve (AUROC) of approximately 0.7912 with a 95% confidence interval of 0.7495-0.8275 (on the training data). The leftmost red point in the plot in Figure 2 shows this result.

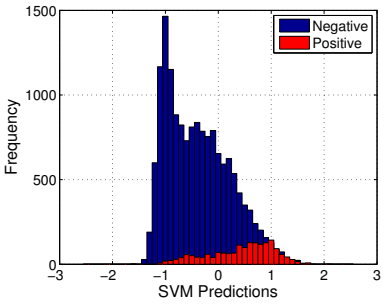


Figure 1. The continuous output of the SVM, the distance each example falls from the decision boundary, can be concatenated to produce approximate risk processes for each patient.

To get a sense of how our ability to predict positive test results changes as we move further from the index event, we compute the AUROC of a classifier based on data collected at different points during the admission. We consider the same group of patients at each point, but we use a different day to score the patient: the most recent day relative to the classification problem. For example, given a patient who tests positive on day 7, we consider their risk score on day 1 to compute the AUROC of the classifier 6 days in advance of the index event. We note that the classifier's ability to predict high risk events is greater closer to the index event. More on these results will be presented in Section 4. But, at this point nothing regarding the patient's previous risk, aside from the patient's state upon admission, is incorporated into the calculation of the patient's risk, since the SVM scores each day independently. The next step is to combine these predictions.

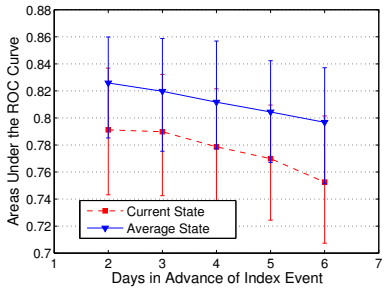


Figure 2. One could measure a patient's risk considering only the current state of the patient, however taking into account the patient's risk on previous days (*Average State*) leads to a significant increase in performance on the training data.

3.3. Refining Patient Risk Processes

When concatenated for a given admission, the continuous SVM predictions result in a risk process for each patient. Examples of these risk processes for the training data are plotted in Figure 3. As shown in the examples, the SVM predictions can fluctuate greatly from day to day. We quantify the extent of these fluctuations over the course of an admission as the expected absolute second difference, or Φ. The plot in Figure 4 is a distribution of Φ across all patients in the training set, while Figure 3 gives examples of patients with different Φ. As Figure 4 shows, the SVM predictions can greatly fluctuate from day to day. Half of the admissions are associated with a $\Phi > 0.42$.

Though large fluctuations in actual risk over time are not impossible, it seems unlikely that they occur as often as our model suggests. Recall in this initial calculation that the variables from time of admission are included in the prediction, but the previous day's risk is not. Taking into account the past risk when calculating current risk could eliminate large fluctuations in the data.

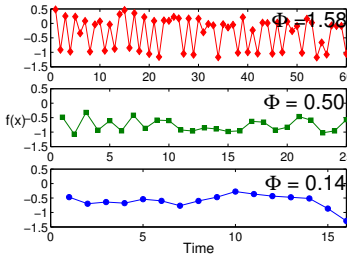


Figure 3. Three examples of approximate risk processes with different amounts of fluctuation, Φ.

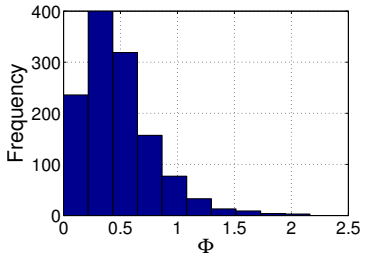


Figure 4. For some patients, the SVM predictions, $f(\mathbf{x})$ fluctuate greatly from day to day.

We consider a variety of weighted trailing averages for which the risk of a patient on a given day incorporates the risks on all previous days. We consider constant, linear, and quadratic weightings. The linear and quadratic weighting schemes weight predictions closer to the index event more heavily. Applied to the train-

ing set, each weighted average outperforms a classifier that uses only the current prediction for that day (See Section 4). Figure 2 compares the average AUROC of a classifier based on a constant weighted average, *Average State*, to a classifier that does not incorporate past risk, *Current State*.

4. Experiments & Results

This section describes a set of experiments used to compare several methods for predicting patient risk over time. We start by describing the experimental setup, which is maintained across all experiments.

4.1. Experimental Setup

In order to reduce the possibility of confusing the risk of becoming colonized with C. Diff with the existence of a current infection, for patients from the positive class, we consider only data collected up to two days before a positive test result. This reduces the possibility of learning a classifier based on symptoms or treatment.

For patients who never test positive, researchers typically use the discharge day as the index event (Dubberke et al., 2011). However, this can lead to deceptively good results because patients nearing discharge are typically healthier than patients not nearing discharge. To avoid this problem, we define the index event for negative examples as either the halfway point of their admission, or 5 days into the admission, whichever is greater. We consider a minimum of 5 days for a negative patient since 5 days is the minimum amount of data we have for any positive patient (e.g., a patient who tests positive on day 7).

To handle class imbalance, we randomly subsampled the negative class, selecting 10 negative examples for each positive example, and we employ asymmetric cost parameters (Morik et al., 1999). Additionally, we removed outliers, those patients with admissions longer than 60 days. Next, we split the data into stratified training and test sets with a 70/30 split. The training set consisted of 1,251 admissions (127 positive), while the test set was composed of 532 admissions (50 positive). This split was maintained across all experiments. In all of the experiments, the training data was used for training purposes and cross-validation parameter selection, and the test set was used for evaluation purposes. For training and classification we employed SVM^{light} (Joachims, 1999).

Table 2. Predicting a positive test result two days in advance by averaging over all previous SVM predictions: the result of different weighting schemes

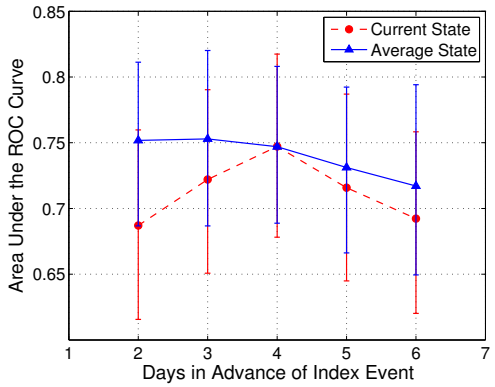
Weighting Scheme	Training AUROC (95%CI)	Testing AUROC (95%CI)
Constant	0.8259 (0.78-0.86)	0.7518 (0.69-0.81)
Linear	0.8242 (0.78-0.86)	0.7444 (0.67-0.80)
Quadratic	0.8214 (0.78-0.86)	0.7360 (0.67-0.80)
Current State	0.7912 (0.74-0.82)	0.6870 (0.61-0.77)

4.2. Results

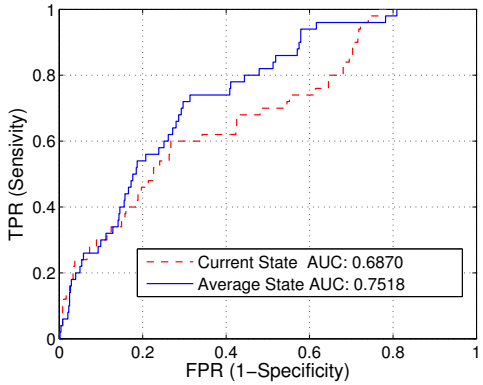
Table 2 compares the performance of four different classifiers applied to the training and test datasets. We consider the problem of predicting whether a patient will have a positive C. Diff test result two days before the index event. Each classifier averages the predictions of the SVM using a different weighting scheme, and based on this average classifies patients as: high risk *i.e.*, will test positive, or low risk *i.e.*, will be discharged without testing positive. *Current State* classifies the patient based only on the most recent prediction. *Constant* is a simple average of the predictions up to the current day, while the other methods weight predictions that occur closer to the index event more heavily.

We showed in Section 3.3 that, when evaluated with respect to training error, a classifier based on the patient's evolving risk outperforms a classifier based only on a patient's current risk. We see the same effect on the held-out data. In all cases, the other methods outperform the *Current State* classifier. There does not appear to be a great difference in performance among the three remaining classifiers. Thus, in the remainder of the results section, we consider only the *Constant* weighting, since it is the simplest weighting scheme that takes into account a patient's evolving risk.

In the plots in Figure 5, we give the results of the *Average State* approach compared to the *Current State* approach on the test set. Prior work on risk stratification methods for predicting C. Diff performed evaluation at only one point in time during an admission (Tanner et al., 2009) (Dubberke et al., 2011). This type of evaluation makes it difficult to interpret how the classifier will perform in practice. Here, we evaluate our risk stratification methods daily from 2 to 6 days before the index event. We cannot evaluate our method at any earlier point across all patients as some patients have only 7 days worth of data. Figure 5(a) gives the AUROC for both of the methods applied to the test set at different points during the hospital admission. The AUROC is calculated by sweeping the decision threshold from 0 to 1. We generated 100 bootsrap replicas



(a) AUROC curve x days before the index event.



(b) ROC Curves for predicting the index event 2 days in advance.

Figure 5. Results of two different methods for predicting a patient's risk of testing positive for C. Diff during their hospital admission in the held-out test set. The *Average State* method consider that patient's average risk up to the time of prediction, while the *Current State* method does not.

by sampling the test set with replacement to compute the pointwise 95% confidence intervals, represented by the error bars. Figure 5(b) plots the ROC curves for both methods.

It is difficult to interpret the performance of a classifier based on these results alone, especially since the classes are imbalanced. Table 3 gives the confusion matrix for the *Average State* classifier.

To further convey the ability of the classifier to risk stratify patients, we split the test patients into quintiles (as is often done in clinical studies) based on the *Average State* prediction two days in advance of the index event. Each quintile contains approximately 106 patients. For each quintile we calculated the probability of a positive test result, based on those patients who eventually test positive for C. Diff. Figure 6 shows that the probability increases with each quintile. The difference between the 1^{st} and 5^{th} quintiles is striking; relative to the 1^{st} quintile, patients in the 5^{th} quintile are at more than a 20-fold greater risk.

Table 3. Confusion Matrix

		Predicted Outcome p	**Predicted Outcome** n
Actual Outcome	p′	TP:26	FN:24
	n′	FP:89	TN:393

5. Discussion & Conclusion

The test results in Table 2 confirm the initial results we achieved on the training data in Section 3.3. At two days before the index event, all methods we investigated, that incorporate a patient's evolving risk, outperform the previous standard *Current State* approach. When we consider the performance of the *Average State* classifier across multiple days, we observe an increase in performance as we move closer to the index event. This is expected since as we move closer to the index event we gain more information about the patient. The performance of the *Current State* classifier has no obvious trend.

That the computed risk changes from day to day highlights the importance of evaluating one's risk stratification method at different points in time, especially if one is classifying patients based solely on the current state of the patient.

In general, the *Average State* classifier outperforms the *Current State* classifier. This difference is most pronounced closer to the index event. However, because of the small number of positive cases, there is large overlap in the confidence intervals, so we cannot say that the difference is significant. The ROC curves cross, indicating that one approach does not dominate the others at all thresholds. Still, we are encouraged by these results.

One source of error in these models is the initial label-

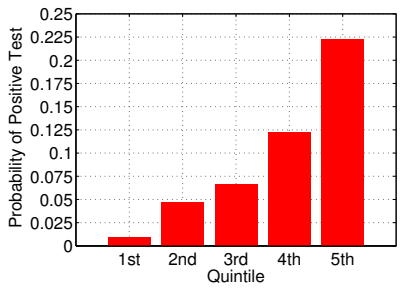


Figure 6. Test patients with *Average State* predictions in the 5^{th} quintile are 20 times more likely to test positive for C. Diff than those in the 1^{st} quintile.

ing of patient days when training the SVM. Incorrect labels at this stage could affect analysis at later stages. Ideally, one would have expert provided labels for each day of whether or not the patient was at high risk. However, these labels would be difficult (if not impossible) to obtain. In Section 3.2.1, we use a labeling scheme to address this challenge. The high dimensionality of the data was another challenge. This, in addition to the large class imbalance, makes learning a useful classifier directly from the data very difficult. Instead, we employed a two-step approach, first reducing the dimensionality of the data in a supervised manner. This step summarizes the patient's state and how it relates to the outcome, but in doing so valuable information may become lost about the precise state of the patient. In future work we will consider other supervised dimensionality reduction techniques.

Despite these challenges, we successfully extracted patient risk processes. Although only approximate, these risk processes are a first step in analyzing how patient risk for C. Diff may evolve during a hospitalization. From our results we conclude that methods that take into account temporal aspects of risk can provide more accurate predictions of the likelihood of adverse outcomes. In this initial work, we consider only weighted combinations of prior risk, but we hypothesize that more sophisticated methods for time series classification could reveal temporal trends and further augment performance.

We believe that gaining an understanding about the dynamics of patient risk over time and its influence on outcomes can have real clinical impact. Having access in clinical settings to an up-to-date risk score and overall prediction about outcomes for each patient could guide actions aimed at reducing patient risk. Also, gaining a deeper understanding of risk processes could shed light on sources of risk, helping clinicians to develop more effective programs to reduce risk.

References

CDC. Rates for clostridium difficile infection among hospitalized patients. *Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report*, 60(34):1171, 2011.

Chapin, Kimberle. Discrepancies in testing recommendations for clostridium difficile infection: updated review favors amplification test systems. *Expert Review of Molecular Diagnostics*, 12(3):223–226, April 2012.

Dubberke, E.R., Yan, Y., Reske, K.A., Butler, A.M., Doherty, J., Pham, V., and Fraser, V.J. Development and validation of a clostridium difficile infection risk prediction model. *Infect Control Hosp Epidemiol*, 32(4):360–366, 2011.

Joachims, T. Making large-scale svm learning practical. advances in kernel methods - support vector learning, 1999.

Katz, D.A., Lynch, M.E., and Littenber, B. Clinical prediction rules to optimize cytotoxin testing for clostridium difficile in hospitalized patients with diarrhea. *American Journal of Medicine*, 100(5):487–95, 1996.

Morik, K., Brockhausen, P., and Joachims, T. Combining statistical learning with a knowledge-based approach - a case study in intensive care monitoring. *Proc. 16th International Conference on Machine Learning*, 1999.

NCHS. International classification of diseases, 9th revision, clinical modification. National Center for Health Statistics. http://icd9cm.chrisendres.com, 2008.

Rao, S. and Hong, J. Analysis of hidden markov models and support vector machines in financial applications. Technical Report UCB/EECS-2010-63, Berkeley, May 2010.

Shoeb, A., Kharbouch, A., Soegaard, J., and Guttag, J. An algorithm for detecting seizure termination in scalp eeg. In *International Conference of the IEEE Engineering in Medicine and Biology (EMBS)*, 2011.

Tanner, J., Khan, D., Anthony, D., and Paton, J. Waterlow score to predict patietns at risk of developing clostridium difficile-associated disease. *Journal of Hospital Infection*, 71(3):239–244, 2009.