Journal of Biomedical Informatics 59 (2016) 209–217

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Classification of hospital acquired complications using temporal clinical information from a large electronic health record

Jeremy L. Warner [a,b,*], Peijin Zhang [c], Jenny Liu [d], Gil Alterovitz [d,e,f]

[a] Department of Medicine, Division of Hematology & Oncology, Vanderbilt University, Nashville, TN, USA
[b] Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA
[c] School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[d] Department of Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[e] Center for Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[f] Children's Hospital Informatics Program at Harvard-MIT Health Sciences & Technology, Boston, MA, USA

ARTICLE INFO

Article history:
Received 4 May 2015
Revised 18 November 2015
Accepted 12 December 2015
Available online 17 December 2015

Keywords:
Hospital acquired complications
Electronic health record
Classification
Temporal
Ranking

ABSTRACT

Hospital acquired complications (HACs) are serious problems affecting modern day healthcare institutions. It is estimated that HACs result in an approximately 10% increase in total inpatient hospital costs across US hospitals. With US hospital spending totaling nearly $900 billion per annum, the damages caused by HACs are no small matter. Early detection and prevention of HACs could greatly reduce strains on the US healthcare system and improve patient morbidity & mortality rates. Here, we describe a machine-learning model for predicting the occurrence of HACs within five distinct categories using temporal clinical data. Using our approach, we find that at least $10 billion of excessive hospital costs could be saved in the US alone, with the institution of effective preventive measures. In addition, we also identify several keystone features that demonstrate high predictive power for HACs over different time periods following patient admission. The classifiers and features analyzed in this study show high promise of being able to be used for accurate prediction of HACs in clinical settings, and furthermore provide novel insights into the contribution of various clinical factors to the risk of developing HACs as a function of healthcare system exposure.

1. Introduction

Hospital acquired complications (HACs) are secondary complications that affect patients following initial hospital admission. Most commonly, these problems are caused by healthcare-associated infections and other issues resulting as side effects from primary treatments such as surgery. Although well known to occur, many HACs are not identified in a timely manner, and often lead to further medical issues causing increased hospitalization time, temporary and permanent morbidities, or even death. It has been estimated that HACs alone cause nearly 99,000 deaths annually in the United States [1] and lead to an approximately 10% increase in total inpatient hospital costs [2]. Early identification of HACs using electronic health records (EHRs) may lead to prevention or proactive treatment, ultimately increasing patient care quality and reducing unnecessary inpatient costs.

The development and deployment of large-scale EHR databases has led to a surge in medical informatics over the past several years [3–6]. In particular, phenotypes across the human "phenome" have been shown to have significant associations with genotype and with lab features [7–9]. Since the manifestation of a HAC is also a patient phenotype, it follows that there may be a way to identify and predict HACs using temporal information that unfolds during a hospitalization. We have previously shown that certain patient phenotypes, many of which are HACs, are associated with prolonged hospitalization [10].

The Multiparameter Intelligent Monitoring in Intensive Care (MIMIC II) database contains extensive structured clinical information on patients admitted to any of the Beth Israel Deaconess Medical Center Intensive Care Units (ICUs) between 2001 and 2007,

Abbreviations: HAC, hospital acquired complication; EHR, electronic health record; MIMIC, Multiparameter Intelligent Monitoring in Intensive Care; ICD-9-CM, International Classification of Diseases, 9th Edition, Clinical Modification; ICU, Intensive Care Unit; TPR, true positive rate; TNR, true negative rate; ROC, receiver operating characteristic curve; AUC, area under the ROC curve.

* Corresponding author at: Medicine and Biomedical Informatics at Vanderbilt University, 2220 Pierce Ave, Preston Research Building 777, Nashville, TN 37232, USA. Tel.: +1 (615) 322 5464; fax: +1 (615) 343 7602.
E-mail address: jeremy.warner@vanderbilt.edu (J.L. Warner).

http://dx.doi.org/10.1016/j.jbi.2015.12.008
1532-0464/

including details of their hospital stay before and after the ICU component [11]. MIMIC II also contains some narrative information: nursing documentation, radiology reports, and Discharge Summaries. Here, we propose to use MIMIC II to develop predictive models to detect HACs early on in hospitalization, possibly in advance of their occurrence. As a secondary objective, we hypothesized that the relative contribution of predictor variables (e.g. patient age, lab results, and medication orders) would change as a function of the timespan used to generate the predictive model. Analysis of this rich medical cohort can lead to new discoveries for the prediction and possible early prevention of problematic HACs.

2. Methods

2.1. Data extraction and modeling

To identify HACs in the MIMIC II cohort, we used a candidate list of International Classification of Diseases, 9th Edition, Clinical Modification (ICD-9-CM) codes that were previously determined to be associated with an increased risk of prolonged hospitalization [10]. These and additional ICD-9-CM codes were manually reviewed and all codes that were indicative of HACs were used to identify patients having HACs. A total of 414 ICD-9-CM codes were selected through this process. We further categorized each identified ICD-9-CM code into five subcategories: (1) infectious complications; (2) bleeding or clotting complications; (3) surgical/procedural complications; (4) medical complications; and (5) all other complications (Table A1). In order to decrease the possibility that a candidate code was representative of a pre-existing condition, we compared our candidate list to the Elixhauser AHRQ-Web ICD-9-CM list of comorbidities [12,13]. There were 2 overlaps (287.41: post-transfusion purpura & 276.61: transfusion with circulatory overload), but neither of these codes was present in the MIMIC II cohort. We also manually reviewed the Discharge Summaries from 50 randomly selected admissions from each subcategory to determine whether the complication was the result of a prior healthcare event (e.g. a preceding outpatient procedure, or a previous hospitalization).

All adult MIMIC II hospitalizations having one or more of these identified diagnosis codes assigned at discharge were considered cases, and an equal number of hospital admissions were randomly sampled from the remaining MIMIC II hospitalizations as controls, excluding those of duration outside the 1st to 99th percentiles of the cases. Patients with more than one hospitalization could be assigned to case or control groups, based on the ICD-9-CM codes obtained for a particular hospital stay.

Structured data was extracted from MIMIC II for both cases and controls, beginning at the time of initial contact, which usually occurs in the outpatient setting (e.g. the emergency room). Initial and last contact of a patient was algorithmically defined based on our prior definition [10]. Clinical information was extracted for each admission based on time intervals of 1, 2, 3, 6, 12, 18, 24, 48, 72, and 96 h following initial contact; patients whose last contact times were outside of the specified interval were censored for that interval. Structured information was extracted by recording the maximum, minimum, and median values for continuous lab results, the mode for categorical lab results, any procedure codes (ever/never), and any medication provider order entry (POE) (ever/never) time-stamped during each interval. In order to increase power, we developed a large library of custom medication aggregations e.g. to combine a medication given by more than one route but with the same predicted pharmacologic effect (Appendix B). Patient demographics (age at admission, gender, and ethnicity) were included as well.

To prevent excessive noise, over-fitting, and unreasonable demands on the prediction model, initial feature selection was carried out as follows: lab features missing in more than 50% and POE/procedure features missing in more than 95% of our cohort were removed prior to classifier construction. Following feature selection, multiple imputation was performed using the aregImpute algorithm to impute missing values (R package Hmisc) [14].

2.2. Feature ranking

To assess the predictive power of individual features, we computed the information gain ratio of each attribute relative to the class feature [15]. Attributes with high information gain ratios with respect to the classification attribute (presence of HAC) are likely

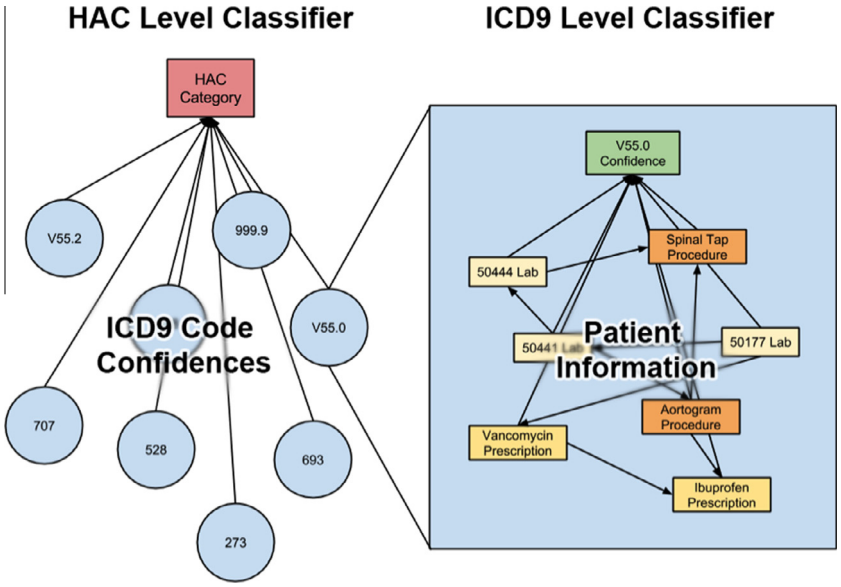

Fig. 1. Illustration of two level classifier model used in this study. The bottom level (ICD-9-CM prediction) model is a Bayesian network that uses patient information to produce a probability of observing the ICD-9-CM code. The top level (HAC prediction) model uses confidences given by the first level Bayesian network to perform binary classification on whether or not a given patient falls under a HAC category.

to be good predictors of HAC in patients for that dataset. The attributes were ranked by their information gain ratios to identify the relative predictive power of each attribute over each time interval. Attributes which are consistently ranked as having the highest information gain or undergo sudden changes in their rankings over time would be of particular interest to clinicians, as closely observing these specific attributes may help lead to earlier detections of HACs.

2.3. Classification

To perform classification, we employed a two level model to predict each HAC category. From the selected case/control cohort for each HAC category, we generated multiple subcohorts for each identified ICD-9-CM code linked to HACs by selecting all patients with the given ICD-9-CM code as cases and randomly selecting an equal number of the remainder as controls. Bayesian network classifiers were then trained on the generated subcohorts using a tree-augmented naïve Bayes (TAN) search algorithm [16,17]. Each of the trained Bayesian networks would then be able to report the posterior probability of each patient having a given ICD-9-CM value using that patient's feature set. We then trained a secondary classifier using AdaBoosted decision stumps on top of these Bayesian network models using the generated posterior probabilities reported by each Bayesian network classifier as features to predict the occurrence of each HAC category. The AdaboostM1 algorithm [18] was used with 100 training iterations with boosting performed by reweighting rather than resampling. The Weka Machine Learning Library [19] implementation of TAN Bayes and AdaboostM1 was used for this study. A graphical illustration of this two level classifier can be seen in Fig. 1.

The structure of the two-layer classification model can be seen as something similar to a neural network consisting of five output nodes (the five HAC categories), a single hidden layer consisting of 414 hidden nodes (each of the identified ICD-9-CM codes), and input nodes being the lab/procedural features. However, instead of learning the entire network simultaneously, we first force each of the hidden nodes to learn the posterior probability of the patient having the given ICD-9-CM code. Then, we have the output nodes learn the whether or not the patient has a HAC from the given category using the predicted probabilities from each of the hidden nodes. We found that while classifiers did not have much success in predicting the general HAC category directly, they were able to predict individual ICD-9-CM codes quite accurately so this two-level model takes advantage of that effect to strengthen our predictive power.

3. Results

The MIMIC II cohort at time of analysis contained a total of 24,580 records for adult hospitalizations. While some patients are admitted directly to the ICU, the majority spend some period of time on an ordinary hospital floor prior to ICU transfer (median time from hospital admission to ICU admission of 19.5 h, interquartile range 12–41 h). Of the entire cohort, we identified 1468 admissions with infectious complications, 1912 admissions with bleeding or clotting complications, 4507 admissions with surgical complications, 484 admissions with medical complications, and 2423 admissions with other complications (numbers fluctuate slightly depending on time interval queried due to some patients not having any lab/procedural/POE entries over smaller intervals and being discharged over larger intervals). In total, there were 8183 unique admissions with one or more candidate codes within our cohort (33% of all adult admissions). Approximately 50% of these had a Discharge Summary available for review, which were

Table 1
Manual review of 50 charts from each subcategory indicate some variation across subcategories, with overall most complications being hospital-acquired.

	Infectious complications		Bleeding/clotting complications		Surgical complications		Medical complications		Other complications		Overall	
Hospital acquired	27	54%	36	72%	37	74%	19	38%	38	76%	157	63%
Pre-existing	11	22%	3	6%	1	2%	8	16%	2	4%	25	10%
Unknown	12	24%	11	22%	12	24%	23	46%	10	20%	68	27%
Sub-total	50	100%	50	100%	50	100%	50	100%	50	100%	250	100%

randomly sampled as described above. Overall, only 10% of complications (95% confidence interval, 7–14%) were determined to be pre-existing – although a significant number could not be determined by manual review of the Discharge Summary (see Table 1).

In the extracted data for patient information obtained during the first hour after initial contact, we found 20 lab features, 45 medication features, and 23 procedural features that passed our feature selection criteria. This number steadily increased to a total of 48 lab features, 135 medication features, and 107 procedural features for the extracted data obtained during the 96 h after initial contact. Thus, we had a range of 91 events per feature at the 1-h mark and 25 events per feature at the 96-h mark.

Full rankings of all extracted features based on information gain ratio with respect to the presence of each HAC category can be found in Appendix B. Tables 2–4 demonstrate selected "interesting" features from the full rankings. Table 2 shows the top three features for each category as measured by the highest mean rank among all the time intervals. Tables 3 and 4 show the top two features for each category as measured by the slope of their change over the time intervals (negative and positive slopes, respectively). Fig. 2 also summarizes this information.

Fig. 3 visualizes the change in AUC for the classifiers over the increasing time intervals for each category. TPR, TNR, and AUC of the classifiers can be found in Table A2 and full information can be found in Appendix B.

4. Discussion

Our analysis of a large-scale critically ill clinical cohort revealed several features demonstrating good predictive power for identifying HACs. Depending on the nature of the feature, careful observation for significant changes (e.g. for lab values) or awareness that an event has occurred (e.g. for procedures) may allow for earlier detection of HACs during hospitalizations, and perhaps a reduction in costs. Of course, due to the fact that ICD-9-CM diagnosis codes are assigned at the end of a hospitalization, we cannot conclude using our current approach the exact time after initial contact that each HAC develops. In fact, our manual review of Discharge

Table 2
Top three features with the highest mean rank over the time intervals for each category. Features were only included if they existed in the data for at least seven of the ten time intervals. Only the top ranking statistic for each lab value is included. NA: feature did not meet initial selection criteria for that time interval so was not included in the classifier.

Attribute name	Hour									
	1	2	3	6	12	18	24	48	72	96
Category 1 – Infectious complications										
Red Blood Cell Distribution Width (RDW) median lab value	14	7	17	2	3	1	3	3	10	38
Parenteral infusion procedure	NA	3	6	38	11	10	25	6	5	9
Vancomycin medication	37	12	1	3	1	2	1	1	2	108
Category 2 – Bleeding or clotting complications										
Serum transfusion procedure	16	3	20	5	27	17	16	14	11	4
Partial Thromboplastin Time (PTT) maximum lab value	49	14	41	2	5	22	10	2	8	7
Hematocrit (HCT) minimum lab value	35	55	34	12	10	4	7	1	4	3
Category 3 – Surgical complications										
Neostigmine medication	NA	NA	42	2	1	1	1	1	4	6
Glycopyrrolate medication	NA	NA	38	1	2	2	2	4	7	9
Ceftriaxone medication	NA	NA	NA	31	10	10	3	18	2	3
Category 4 – Medical complications										
Ethnicity	4	1	1	4	5	2	1	14	14	104
Red Blood Cell Distribution Width (RDW) maximum lab value	3	4	3	1	3	1	2	1	9	125
Atropine sulfate medication	18	15	54	20	51	21	29	7	3	13
Category 5 – Other complications										
Chlorhexidine medication	NA	NA	NA	1	2	2	1	1	1	1
Skin suture procedure	10	7	1	3	15	60	50	NA	NA	NA
Infusion of vasopressor procedure	NA	5	14	33	1	26	16	7	76	10

Table 3
Top two features with the largest increase over the time intervals as measured by the linear regression slope. Features were only included if they existed in the data for at least seven of the 10 time intervals and their max ranking was within the top 10 at some interval. Only the top ranking statistic for each lab value is included. NA: feature did not meet initial selection criteria for that time interval so was not included in the classifier.

Attribute name	Hour										Slope
	1	2	3	6	12	18	24	48	72	96	
Category 1 – Infectious complications											
Single internal mammary-coronary artery bypass procedure	NA	43	44	72	10	13	10	11	6	3	−0.457
Aortocoronary bypass procedure	NA	44	45	73	13	17	9	19	12	11	−0.379
Category 2 – Bleeding or clotting complications											
Corpuscular hemoglobin concentration maximum lab value	87	165	135	25	49	131	109	5	18	18	−1.130
Platelet transfusion procedure	NA	1	19	150	35	9	9	17	10	6	−0.457
Category 3 – Surgical complications											
Neosynephrine medication	99	100	121	25	12	16	16	8	13	8	−0.820
Nitroglycerin medication	89	145	99	12	5	8	5	9	11	16	−0.754
Category 4 – Medical complications											
Nitroprusside medication	90	143	122	14	9	8	13	23	12	5	−0.868
IV line flush	97	129	128	43	55	33	32	71	5	29	−0.809
Category 5 – Other complications											
Lansoprazole medication	NA	NA	NA	90	97	11	95	3	20	3	−0.945
Insertion of intercostal catheter procedure	42	61	158	21	9	5	21	5	34	2	−0.584

Summaries (Table 1) does suggest that a minority of cases are healthcare-associated complications that were present at the time of hospital admission. Unfortunately, it can be very challenging to identify these cases through automated methods. MIMIC II does not have admitting diagnosis as structured data, nor does it have admission H&Ps or clinician progress notes. When the Discharge Summary is available (~50% of the time), the narrative is highly variable and requires extensive temporal reasoning, which has been described as a "very challenging area" for clinical natural language processing [20]. Thus, at this point in the pilot work, we cannot determine with absolute confidence how much earlier our developed classifier may detect HACs when compared to the time of discovery by hospital personnel. Future work will evaluate whether the recent open-source temporal solution described by Lin et al. can assist with the determination of the temporality of the complications, including whether they were present at the time of admission – although the substantial degree of unknown determination even after manual review indicates that this task will remain challenging [21]. Interestingly, medical complications (e.g. constipation from narcotics, renal failure from diuretics) were especially likely to remain undetermined after manual review; one possibility is that these are so widespread that they are considered part of the routine consequences of medical care and are thus not extensively documented.

Furthermore, it is possible that features could reflect the treatment of HACs as opposed to being their causation. It is also possible that what we are classifying as a HAC could have been a pre-existing condition (including the reason for hospitalization in the first place), but we tried to ameliorate this possibility by ensuring that none of the HAC candidate codes overlapped with the Elixhauser comorbidity codes. Although some clinical systems have a "present on admission" (POA) flag to identify pre-existing conditions, MIMIC II does not have this information. Even with these limitations, several interesting patterns emerge in this analysis, as discussed in the following sections:

4.1. Feature rankings

Examination of how the relative contribution of individual lab, procedural, and medication features to the model changes as a function of time revealed several interesting patterns. The vast majority of features identified in our data fell under one of four possible categories: features that were consistently highly ranked across all time intervals (selected features displayed in Table 2), features that started at low ranks but became highly ranked as the time intervals increased (Table 3), features that started at high ranks but steadily decreased as time intervals increased (Table 4), or features that were generally poorly ranked across all time intervals.

Early in a hospitalization, features that measure patient physiology, such as hematologic and chemistry parameters, clearly drive the model. As the time horizons lengthen, features that represent procedural and medical interventions begin to dominate the model. While causality cannot be determined from this analysis (e.g. procedures such as platelet transfusion may cause complications, or could be used to treat complications), there is certainly the suggestion that the interventionist approach ubiquitous in the United States healthcare system is a strong driver of HAC phenotypes. These findings indicate that the patient features clinicians should pay close attention to for early HAC detection evolve as a function of a patient's time from admission.

Looking at the top increasing and decreasing features for the various categories (Fig. 2), a common trend is that the initial incline and decline in feature rankings are observed at the same hour. For example, in infectious complications single internal mammary-coronary artery bypass procedure and aortocoronary bypass procedure both start to rise in rank at hour 12 while red blood cell count (RBC) and hematocrit (HCT) also start to decrease in rank at hour 12. Looking across all categories, this suggests that hours 6–18 are crucial for observing initial incline and decline of features to predict which category of HAC will occur, which agrees with previous literature [22–30]. Another interesting commonality is that bleeding or clotting complications (Fig. 2b), medical complications (Fig. 2d), and other complications (Fig. 2e) share the feature that measures hemoglobin (HGB) maximum lab value as one of the top 2 largest decreasing features. This suggests that hemoglobin (HGB) maximum lab value could be a feature to help predict HAC within the first 18 h [31–35]. This data supports the notion that the first 24 h of a critical illness admission are crucial and that excessive handoffs during this period are likely to be problematic [36].

One of the consistently highly ranked features found in this study is RDW, which agrees with previous literature that has found RDW to be a good independent predictor of morbidity [37], mortality [38], and hospital readmission [39]. In addition, while lab features tend to dominate the model during early intervals and

Table 4
Top two features with the largest decrease over the time intervals as measured by the linear regression slope. Features were only included if they existed in the data for at least seven of the 10 time intervals and their max ranking was within the top 10 at some interval. Only the top statistic for each lab value is included. NA: feature did not meet initial selection criteria for that time interval so was not included in the classifier.

Attribute name	Hour										Slope
	1	2	3	6	12	18	24	48	72	96	
Category 1 – Infectious complications											
Red Blood Cell Count (RBC) maximum lab value	39	14	12	9	20	24	62	168	148	189	+2.010
Hematocrit (HCT) maximum lab value	29	11	7	14	26	21	45	139	147	170	+1.862
Category 2 – Bleeding or clotting complications											
Basophils maximum lab value	NA	4	5	34	206	227	176	256	286	370	+3.390
Hemoglobin (HGB) maximum lab value	39	53	10	67	58	118	133	150	180	254	+2.152
Category 3 – Surgical Complications											
Basophils median lab value	NA	7	3	43	196	243	231	309	366	364	+3.731
Monocytes minimum lab value	NA	97	1	114	210	212	245	302	288	309	+2.482
Category 4 – Medical complications											
Hemoglobin (HGB) maximum lab value	9	18	22	16	8	35	21	54	149	245	+2.235
Insertion of endotracheal tube procedure	115	32	4	151	252	212	78	178	274	295	+2.187
Category 5 – Other complications											
Red Blood Cell Distribution Width (RDW) maximum lab value	6	1	5	5	11	21	5	115	230	239	+2.815
Hemoglobin (HGB) maximum lab value	22	13	9	13	38	24	105	140	246	249	+2.811

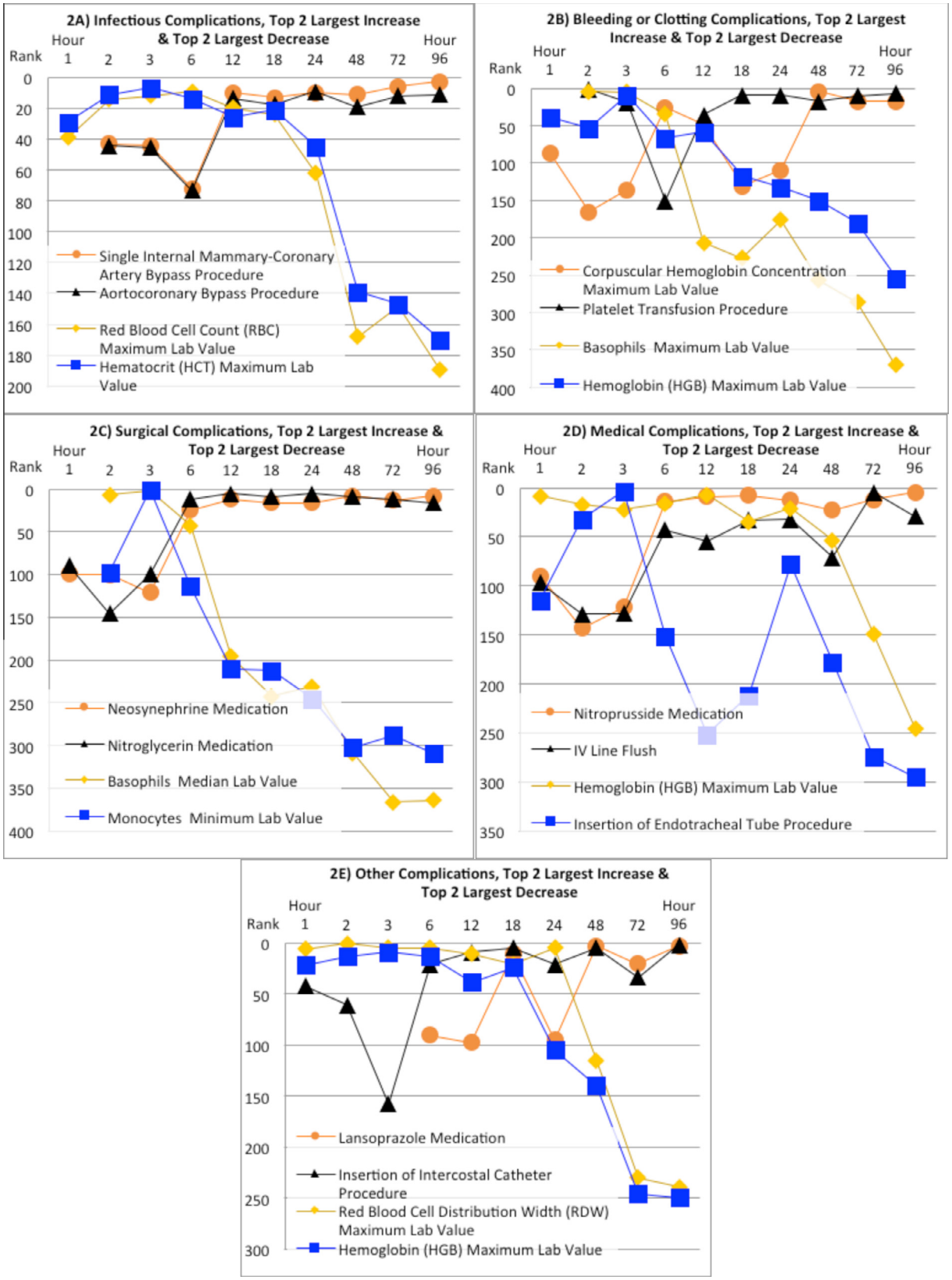

Fig. 2. Top two features with the largest decrease and top two features with the largest increase over the time intervals as measured by the linear regression slope for each category. Features were only included if they existed in the data for at least seven of the 10 time intervals and their max ranking was within the top 10 at some interval. Only the top statistic for each lab value is included.

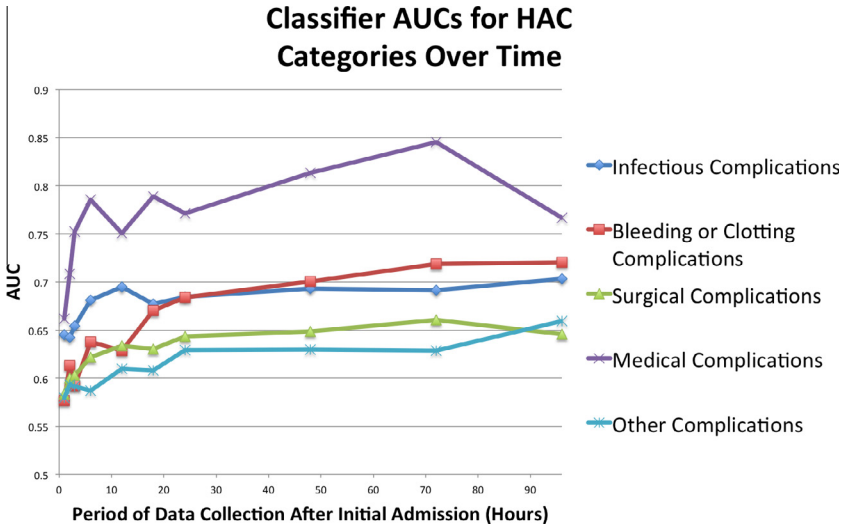

Fig. 3. AUC of HAC predicting classifiers for each category over the data collection periods.

procedural/medication features for later intervals, there are still a good number of features that are highly ranked across all time intervals, and the distribution of these features seems to be roughly equal between labs, procedures, and medications. This indicates that while labs may be more useful for predicting HACs immediately after admission and procedural/medication information may be more useful for predicting HACs several days after admission, all three of these feature types are needed to have accurate prediction of HACs.

If we look within each of the categories, there are commonalities between the top two increasing features and top two decreasing features. For example, the infectious complications (Fig. 2) top two increasing features of single internal mammary-coronary artery bypass procedure and aortocoronary bypass overlay one another very closely. The same occurs for red blood cell count (RBC) maximum lab value and hematocrit (HCT) maximum lab value. This suggests that bypass-related procedures [22,23,26] and red blood cell related lab values are predictors for infectious complications. Similarly, in surgical complications (Fig. 2c) neosynephrine medication and nitroglycerin medication overlay one another very closely. The same occurs for the features that measure basophils median lab value and monocytes minimum lab value. This suggests that vasodilator and vasopressor medications and white blood cell lab values could be features to help predict surgical complications [40–46]. In other complications, we also observed that the feature that measures chlorhexidine medication is the top median ranked feature, the top mean ranked feature, and also the number one ranked feature at hour 96. This suggests that chlorhexidine may be useful in predicting other complications [47–51].

One of the drawbacks of using information gain ratio to score the predictive power of features is that this measure tends to inaccurately assign high values to nominal features with many possible categories [52]. However, aside from ethnicity of which there are 34 distinct values, the feature space consists of continuous and binary features aside from a few categorical lab exceptions with a small amount of nominal values. For example, urine color can be described as straw, light amber, red, yellow or dark amber. Likewise, urine appearance can be described as clear, cloudy, slightly cloudy, or slightly hazy. Thus, this disadvantage of using information gain ratio is mostly mitigated in this study, although it is likely the reason why ethnicity is ranked so high as a predictor for medical complications (Table 2).

Data extraction in this study used several summative statistics (min/max/median) for continuous features. Such measures do not encode information about the history of these values over time for each patient. For example, different patients having measurements of (1,2,3), (3,2,1), and (1,1,3,3) over time for a given lab test would end up with identical lab features from our extraction pipeline. Future work could go into developing methods for extracting such temporal information into features for use in the downstream models, or adopting methods such as those described by Lasko et al. [53].

4.2. Classification

The generated two-level classifiers were able to achieve relatively good performance, peaking at an AUC of 0.845 for HAC category 4 (medical complications). The evolution of classifier AUC over the time intervals shown in Fig. 3 reveals that the classifier for medical complications demonstrates a fair amount of nonlinearity, whereas the other classifiers perform with a relatively constant AUC across all the tested time intervals. With the exception of somewhat inferior performance at the shortest time intervals (mostly due to missing data), the mostly linear performance is reassuring for the generalizability of the method.

One-to-one case versus control sampling was used to ensure a relatively high sensitivity in the resulting classifiers. Since some types of HACs are relatively uncommon within the clinical cohort, there is an issue of class imbalance. For example, the medical complications subcategory has a class imbalance of nearly 1:50 within our cohort. Failing to account for the class imbalance would likely result in classifiers with high specificity but low sensitivity, which would not be very useful in the clinical setting for detecting HACs.

Bayesian network models were employed for the bottom level classifiers in order to model the many potential causal relationships between clinical features. There is a significant amount of dependency between features in our feature space, especially between the three features that are derived from serial measurements of the same lab tests (min/median/max). In addition, due to our control down-sampling, our data contained a relatively high ratio of features to training data, ranging from 25 to 91 events per feature depending on the time period included. While "traditional" classifiers such as logistic regression need as few as 20 events per feature for reproducibility, it is possible that our classifier may suffer from overfitting, which would limit generalizability [54]. Efforts

are underway to expand the MIMIC II cohort to upwards of 150,000 patients across multiple institutions, at which point we could pursue re-training to establish generalizability.

The reasoning behind utilizing the two level model is that we found our machine learning classifiers to generally have good accuracy in predicting individual ICD-9-CM codes, but failed to have good accuracy in predicting overall HAC categories. This two level model takes advantage of the specificity of ICD-9-CM codes to give good classifier accuracy but still allows for the classifier to output HAC categories rather than individual ICD-9-CM predictions. Empirically, we found that the two-level classifier performed at levels either comparable to the single-level models or up to 6% more accurate than the single-level models, depending on the HAC category or time interval tested.

5. Conclusion

There are many potential applications of our preliminary findings, which have demonstrated internal validity but would also benefit from validation across other cohorts. The generated classifiers could act as a clinical decision support tool to assist doctors in the early detection of HACs in clinical settings, although any such system would have to be prospectively evaluated. The individual identified features could serve as important markers for clinicians to keep an eye out for as hospitalization progresses, in order to catch early signs of HACs. It is clear, based on the fact that all attributes undergo dynamic rank change as a function of time, that the process of hospitalization for critically ill patients is highly dynamic and likely progresses through distinct epochs. Future work will focus on whether these epochs can be determined mechanistically, such that the complicated process of critical illness can be more thoroughly understood.

If early prediction of HACs can be operationalized into prevention of HACs, the potential savings could be enormous. Based on previous literature of medical claims, we hypothesize that at least $10 billion of excessive hospital costs due to HACs could be saved annually (Table A3). In addition to hospital costs, preventing HACs would translate into the saving of more than 4.6 million days of excessive hospital stays [55–59]. In conclusion, HACs are a prominent problem in modern day hospitals and provide a significant drain on the healthcare system. Machine learning classifiers such as the ones developed in this study could prove to be valuable tools to assist clinicians in the detection, mitigation, and ultimately prevention of HACs.

Conflict of interest

The authors have no conflicts of interest to disclose.

Acknowledgment

This work was supported in part by grant NLM 4R00LM009826-03.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2015.12.008.

References

[1] R.M. Klevens et al., Estimating health care-associated infections and deaths in U.S. hospitals, 2002, Public Health Rep. 122 (2007) 160–166.
[2] R.L. Fuller, E.C. McCullough, M.Z. Bao, R.F. Averill, Estimating the costs of potentially preventable hospital acquired complications, Health Care Financ. Rev. 30 (2009) 17–32.
[3] D. Houle, D.R. Govindaraju, S. Omholt, Phenomics: the next challenge, Nat. Rev. Genet. 11 (2010) 855–866, http://dx.doi.org/10.1038/nrg2897.
[4] I.S. Kohane, Using electronic health records to drive discovery in disease genomics, Nat. Rev. Genet. 12 (2011) 417–428, http://dx.doi.org/10.1038/nrg2999.
[5] S. Murphy et al., Instrumenting the health care enterprise for discovery research in the genomic era, Genome Res. 19 (2009) 1675–1681, http://dx.doi.org/10.1101/gr.094615.109.
[6] S. Lyalina et al., Identifying phenotypic signatures of neuropsychiatric disorders from electronic medical records, J. Am. Med. Inform. Assoc. 20 (2013) e297–e305, http://dx.doi.org/10.1136/amiajnl-2013-001933.
[7] J.L. Warner, G. Alterovitz, Phenome-based analysis as a means for discovering context-dependent clinical reference ranges, in: AMIA Annual Symposium Proceedings, 2012, pp. 1441–1449.
[8] J.C. Denny et al., Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data, Nat. Biotechnol. 31 (2013) 1102–1110, http://dx.doi.org/10.1038/nbt.2749.
[9] S.J. Hebbring, The challenges, advantages and future of phenome-wide association studies, Immunology 141 (2014) 157–165, http://dx.doi.org/10.1111/imm.12195.
[10] J.L. Warner et al., Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications, J. Am. Med. Inform. Assoc. 20 (2013) e281–e287, http://dx.doi.org/10.1136/amiajnl-2013-001861.
[11] M. Saeed et al., Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database, Crit. Care Med. 39 (2011) 952–960, http://dx.doi.org/10.1097/CCM.0b013e31820a92c6.
[12] H. Quan et al., Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data, Med. Care 43 (2005) 1130–1139.
[13] C. van Walraven, P.C. Austin, A. Jennings, H. Quan, A.J. Forster, A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data, Med. Care 47 (2009) 626–633, http://dx.doi.org/10.1097/MLR.0b013e31819432e5.
[14] F.E. Harrell Jr., M.F.E. Harrell Jr., Package 'Hmisc', 2015.
[15] J.T. Kent, Information gain and a general measure of correlation, Biometrika 70 (1983) 163–173, http://dx.doi.org/10.1093/biomet/70.1.163.
[16] G.F. Cooper, E. Herskovits, A Bayesian method for the induction of probabilistic networks from data, Mach. Learn. 9 (1992) 309–347, http://dx.doi.org/10.1007/bf00994110.
[17] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Mach. Learn. 29 (1997) 131–163, http://dx.doi.org/10.1023/a:1007465528199.
[18] Y. Freund, R.E. Schapire, in; ICML, 1996, pp. 148–156.
[19] M. Hall et al., The WEKA data mining software: an update, ACM SIGKDD Explor. Newslett. 11 (2009) 10–18.
[20] L. Zhou, G. Hripcsak, Temporal reasoning with medical data – a review with emphasis on medical natural language processing, J. Biomed. Inform. 40 (2007) 183–202, http://dx.doi.org/10.1016/j.jbi.2006.12.009.
[21] C. Lin, D. Dligach, T.A. Miller, S. Bethard, G.K. Savova, Multilayered temporal modeling for the clinical domain, J. Am. Med. Inform. Assoc. (2015), http://dx.doi.org/10.1093/jamia/ocv113.
[22] S.G. Raja, M. Rochon, J.W. Jarman, Brompton Harefield Infection Score (BHIS): development and validation of a stratification tool for predicting risk of surgical site infection after coronary artery bypass grafting, Int. J. Surg. 16 (2015) 69–73, http://dx.doi.org/10.1016/j.ijsu.2015.02.008.
[23] M.P. Robich et al., Prolonged effect of postoperative infectious complications on survival after cardiac surgery, Ann. Thorac. Surg. 99 (2015) 1591–1599, http://dx.doi.org/10.1016/j.athoracsur.2014.12.037.
[24] J.H. Lin et al., Prognostic factors and complication rates for double-filtration plasmapheresis in patients with Guillain-Barre syndrome, Transfus. Apheresis Sci.: Off. J. World Apheresis Assoc.: Off. J. Eur. Soc. Haemapheresis 52 (2015) 78–83, http://dx.doi.org/10.1016/j.transci.2014.12.005.
[25] A. Hart et al., Blood transfusion in primary total hip and knee arthroplasty. Incidence, risk factors, and thirty-day complication rates, J. Bone Joint Surg. American vol. 96 (2014) 1945–1951, http://dx.doi.org/10.2106/JBJS.N.00077.
[26] G. Yumun et al., Deep sternal wound infection after coronary artery bypass surgery: management and risk factor analysis for mortality, Heart Surg. Forum 17 (2014) E212–E216, http://dx.doi.org/10.1532/HSF98.2014346.
[27] A. Kunac et al., Bacteremia and ventilator-associated pneumonia: a marker for contemporaneous extra-pulmonic infection, Surg. Infect. 15 (2014) 77–83, http://dx.doi.org/10.1089/sur.2012.030.
[28] S. Ozkardesler et al., Effects of blood products on nosocomial infections in liver transplant recipients, Exp. Clin. Transplan.: Off. J. Middle East Soc. Organ Transplan. 11 (2013) 530–536, http://dx.doi.org/10.6002/ect.2012.0286.
[29] G.S. Chahal et al., A comparison of complications requiring return to theatre in hip and knee arthroplasty patients taking enoxaparin versus rivaroxaban for thromboprophylaxis, Ortopedia, traumatologia, rehabilitacja 15 (2013) 125–129, http://dx.doi.org/10.5604/15093492.1045953.
[30] N.I. Omoke, C.G. Nwigwe, An analysis of risk factors associated with traumatic extremity amputation stump wound infection in a Nigerian setting, Int. Orthopaed. 36 (2012) 2327–2332, http://dx.doi.org/10.1007/s00264-012-1641-3.
[31] B.A. Ramo et al., Surgical site infections after posterior spinal fusion for neuromuscular scoliosis: a thirty-year experience at a single institution, J. Bone Joint Surg. American vol. 96 (2014) 2038–2048, http://dx.doi.org/10.2106/JBJS.N.00277.

[32] E.F. Mostafa, A.N. Mohammad, Incidence and predictors of rebleeding after band ligation of oesophageal varices, Arab J. Gastroenterol.: Off. Pub. Pan-Arab Assoc. Gastroenterol. 15 (2014) 135–141, http://dx.doi.org/10.1016/j.ajg.2014.10.002.
[33] K. Nie et al., Risk factors of intra-abdominal bacterial infection after liver transplantation in patients with hepatocellular carcinoma, Chin. J. Cancer Res. = Chung-kuo yen cheng yen chiu 26 (2014) 309–314, http://dx.doi.org/10.3978/j.issn.1000-9604.2014.06.10.
[34] J. Humphers, N. Shibuya, B.L. Fluhman, D. Jupiter, The impact of glycosylated hemoglobin and diabetes mellitus on postoperative wound healing complications and infection following foot and ankle surgery, J. Am. Podiat. Med. Assoc. (2014), http://dx.doi.org/10.7547/13-026.1.
[35] C.Y. Chien, C.H. Lin, J.W. Chen, R.B. Hsu, Blood stream infection in patients undergoing systematic off-pump coronary artery bypass: incidence, risk factors, outcome, and associated pathogens, Surg. Infect. 15 (2014) 613–618, http://dx.doi.org/10.1089/sur.2012.213.
[36] S.A. Collins et al., In search of common ground in handoff documentation in an Intensive Care Unit, J. Biomed. Inform. 45 (2012) 307–315, http://dx.doi.org/10.1016/j.jbi.2011.11.007.
[37] A.C. Salisbury et al., Red blood cell indices and development of hospital-acquired anemia during acute myocardial infarction, Am. J. Cardiol. 109 (2012) 1104–1110, http://dx.doi.org/10.1016/j.amjcard.2011.11.045.
[38] M. Gul et al., The relationship between red blood cell distribution width and the clinical outcomes in non-ST elevation myocardial infarction and unstable angina pectoris: a 3-year follow-up, Coronary Artery Dis. 23 (2012) 330–336, http://dx.doi.org/10.1097/MCA.0b013e3283564986.
[39] G. Ephrem, Red blood cell distribution width is a predictor of readmission in cardiac patients, Clin. Cardiol. 36 (2013) 293–299, http://dx.doi.org/10.1002/clc.22116.
[40] C. Bleriot et al., Liver-resident macrophage necroptosis orchestrates type 1 microbicidal inflammation and type-2-mediated tissue repair during bacterial infection, Immunity 42 (2015) 145–158, http://dx.doi.org/10.1016/j.immuni.2014.12.020.
[41] R. Agrawal et al., Infection with human rhinovirus 16 promotes enhanced IgE responsiveness in basophils of atopic asthmatics, Clin. Exp. Allergy: J. Br. Soc. Allergy Clin. Immunol. 44 (2014) 1266–1273, http://dx.doi.org/10.1111/cea.12390.
[42] U. Dapunt, S. Maurer, T. Giese, M.M. Gaida, G.M. Hansch, The macrophage inflammatory proteins MIP1alpha (CCL3) and MIP2alpha (CXCL2) in implant-associated osteomyelitis: linking inflammation to bone degradation, Mediators Inflamm. 2014 (2014) 728619, http://dx.doi.org/10.1155/2014/728619.
[43] M. Jog, R. Sachidananda, K. Saeed, Risk of contamination of lidocaine hydrochloride and phenylephrine hydrochloride topical solution: in vivo and in vitro analyses, J. Laryngol. Otol. 127 (2013) 799–801, http://dx.doi.org/10.1017/S0022215113001564.
[44] R. Mariappan, P. Manninen, E.M. Massicotte, A. Bhatia, Circulatory collapse after topical application of vancomycin powder during spine surgery, J. Neurosurg. Spine 19 (2013) 381–383, http://dx.doi.org/10.3171/2013.6.SPINE1311.
[45] A. Ramadan et al., Activation of basophils by the double-stranded RNA poly(A:U) exacerbates allergic inflammation, Allergy 68 (2013) 732–738, http://dx.doi.org/10.1111/all.12151.
[46] E. Hamaguchi, H. Kawano, A case of severe hypotension with catecholamine-resistant septic shock in the perioperative period, Masui. Jpn. J. Anesthesiol. 61 (2012) 400–403.
[47] H.C. Tsai et al., Central venous catheter-associated bloodstream infections in pediatric hematology-oncology patients and effectiveness of antimicrobial lock therapy, J. Microbiol., Immunol., Infect. = Wei mian yu gan ran za zhi (2014), http://dx.doi.org/10.1016/j.jmii.2014.07.008.
[48] C.Y. Chien, C.H. Lin, R.B. Hsu, Care bundle to prevent methicillin-resistant Staphylococcus aureus sternal wound infection after off-pump coronary artery bypass, Am. J. Infect. Control 42 (2014) 562–564, http://dx.doi.org/10.1016/j.ajic.2014.01.016.
[49] K. Falk-Brynhildsen, B. Soderquist, O. Friberg, U.G. Nilsson, Bacterial recolonization of the skin and wound contamination during cardiac surgery: a randomized controlled trial of the use of plastic adhesive drape compared with bare skin, J. Hosp. Infect. 84 (2013) 151–158, http://dx.doi.org/10.1016/j.jhin.2013.02.011.
[50] K. Carroll, M. Dowsey, P. Choong, T. Peel, Risk factors for superficial wound complications in hip and knee arthroplasty, Clin. Microbiol. Infect.: Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis. 20 (2014) 130–135, http://dx.doi.org/10.1111/1469-0691.12209.
[51] P. Thompson, S. Houston, Decreasing methicillin-resistant Staphylococcus aureus surgical site infections with chlorhexidine and mupirocin, Am. J. Infect. Control 41 (2013) 629–633, http://dx.doi.org/10.1016/j.ajic.2012.09.003.
[52] A. White, W. Liu, Technical note: bias in information-based measures in decision tree induction, Mach. Learn. 15 (1994) 321–329, http://dx.doi.org/10.1023/A:1022694010754.
[53] T.A. Lasko, J.C. Denny, M.A. Levy, Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data, PLoS One 8 (2013) e66341, http://dx.doi.org/10.1371/journal.pone.0066341.
[54] T. van der Ploeg, P.C. Austin, E.W. Steyerberg, Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints, BMC Med. Res. Methodol. 14 (2014) 137, http://dx.doi.org/10.1186/1471-2288-14-137.
[55] S.S. Magill et al., Multistate point-prevalence survey of health care-associated infections, N. Engl. J. Med. 370 (2014) 1198–1208, http://dx.doi.org/10.1056/NEJMoa1306801.
[56] C. Zhan, M.R. Miller, Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization, Jama 290 (2003) 1868–1874, http://dx.doi.org/10.1001/jama.290.14.1868.
[57] J. Shreve et al., The Economic Measurement of Medical Errors Sponsored by Society of Actuaries' Health Section, Milliman Inc, 2010.
[58] National Hospital Ambulatory Medical Care Survey, Emergency Department Summary Tables, 2011. <http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2011_ed_web_tables.pdf>.
[59] Interim Update on 2013 Annual Hospital-Acquired Condition Rate and Estimates of Cost Savings and Deaths Averted From 2010 to 2013. <http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2013.pdf>.