Peer-reviewed version available at *Informatics* **2019**, *6*, 4; doi:10.3390/informatics6010004

Unstructured text in EMR improves prediction of death after surgery in children

Oguz Akbilgic [1,2,*], Ramin Homayouni [3,4], Kevin Heinrich [4], Max Raymond Langham, Jr [5], Robert Lowell Davis [1]

[1] UTSHC-ORNL Center for Biomedical Informatics, Department of Pediatrics, Memphis, TN, USA
[2] Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA
[3] Department of Biology, Bioinformatics Program, University of Memphis, Memphis, TN, USA
[4] Quire Inc, Memphis, TN, USA
[5] Department of Surgery, University of Tennessee Health Science Center, Memphis, TN, USA

* Correspondence: oz@uthsc.edu, +1-901-287-5841

Abstract: Text fields in electronic medical records (EMR) contain information on important factors that influence health outcomes, however, they are underutilized in clinical decision making due to their unstructured nature. We analyzed 6,497 inpatient surgical cases with 719,308 free text notes from Le Bonheur Children's Hospital EMR. We used a text mining approach on preoperative notes to obtain the text-based risk score algorithm as predictive of death within 30 days of surgery. We studied the additional performance obtained by including text-based risk score as a predictor of death along with other structured data based clinical risk factors. The C-statistic of a logistic regression model with 5-fold cross-validation significantly improved from 0.76 to 0.92 when text-based risk scores were included in addition to structured data. We conclude that preoperative free text notes in EMR include significant information that can predict adverse surgery outcomes.

Keywords: Post-operative death, unstructured data, logistic regression, text mining, surgery outcome

---

1. Introduction

The introduction should briefly place the study in a broad context and highlight why it is important. It The Health and Medicine Division of the National Academies of Science, Engineering and Medicine have documented medical error as a leading cause of death in the United States and suggested that health information technology has the potential to improve care and health outcomes[1]. To make this vision a reality requires the capacity to predict uncommon events that are potentially preventable. Death after surgery in children is an infrequent occurrence, with an incidence rate of <1.0%[2], [3]. Other adverse outcomes such as unplanned return to the operating room, reintubation after surgery, need for blood transfusions, and unplanned readmission are more common, with incidence rates ranging from 0.2% to 4.4%[4]–[6]. Since over 5 million operations are performed on children each year in the United States[7], even a low rate of postoperative mortality represents thousands of lives lost prematurely. The best published models predicting these events rely on structured data such as that contained in the National Surgical Quality Improvement Program-Pediatric (NSQIP-Ped)[3], [8], [9]. Such models are useful in quality improvement work, risk-based payment methods, in improving surgical decision-making, and in providing accurate informed consent discussions with parents [10]–[13].

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

The value of EMR data mining for research applications and clinical care is beginning to be realized[14], [15]. However, analysis of the unstructured text data found in clinical narratives such as admission and discharge summaries, observation notes, and a variety of reports in the EMR has been relatively underutilized. Several Natural Language Processing (NLP) methods have been specifically developed for mining EMR data, including MedLEE[16], HiTex[17], cTAKES[18], and TEPAPA[19]. A recent systematic review showed that the majority of NLP approaches have focused on information extraction tasks such as case-detection[20]. Meta-analysis of 19 case-detection studies showed that the median precision and C-statistic significantly increased when unstructured text was used in addition to structured data from the EMR[20]. Examples of case-detection studies include identification of cancer patient trajectories[14], surveillance of coronary stents in a time- and cost-effective manner[21], and improved accuracy in diagnosis of prostate cancer[22]. There are already some studies that have utilized unstructured text in the EMR to predict health outcomes. Frost et al.[23], using over 11,000 words from text fields in the EMR of over 43,000 patients in a logistic regression model, predicted risk of frequent emergency department visits and high system costs with a C-statistic of 0.71 and 0.76, respectively. Weissman et al. [24] showed that inclusion of unstructured text along with structured data improved prediction of death in the ICU by using four different predictive modeling approaches.

Text-mining of clinical notes has been successfully used to identify postoperative complications in a population of veterans[25]. However, to the best of our knowledge, free text notes in EMR have not been utilized for predicting postoperative surgery outcomes in children. In this study, we examined the use of free text notes in the EMR for preoperative prediction of death after surgery in children. We hypothesized that 1) a risk score can be created through mining unstructured free text notes in EMRs and 2) this text-based risk score will improve the performance of models that only use structured data such as those defined by NSQIP-Ped[26].

2. Materials and Methods

We analyzed a convenience sample of children undergoing inpatient surgical procedures at Le Bonheur Children's Hospital, Memphis, TN, on or before their 19th birthday, whose medical records included preoperative free text notes and whose operation occurred between January 1, 2014 and May 31, 2017. LeBonheur Children's Hospital participates in the National Surgical Quality Improvement Project-Pediatric, and a surgical case reviewer abstracts clinical data for a nonrandom sample of children undergoing operative procedures. We term this the NSQIP cohort; children not included in the abstracted sample are termed the non-NSQIP cohort. We separately extracted and analyzed free text from records of patients in the non-NSQIP cohort to develop the text-based risk score algorithm and implemented this algorithm on the NSQIP (test) cohort as described below. Backward elimination stepwise logistic regression was used to determine if the addition of the text-based risk scores improved performance of risk models derived from structured variables (See Figure 1 for details).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

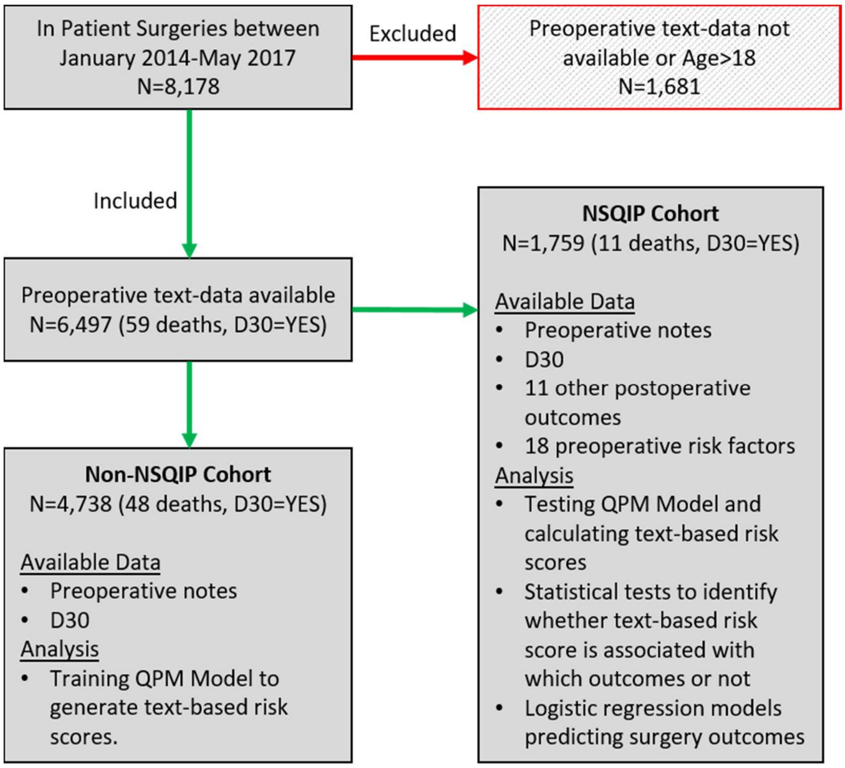


Figure 1 Summary of cohort determination and methods

2.1. Text Mining and Development of Text-Based Risk Score: Unstructured text fields from all inpatient, outpatient and ambulatory settings prior to the date of surgery were extracted for individuals who had surgeries from January 2014 to May 2017. Records were extracted using HL7 Clinical Document Architecture (HL7/CCA) standard from the Cerner EMR system by Methodist/LeBonheur Hospital, then converted to Continuity of Care Document (CCD) in XML, and submitted to Quire Inc. (Memphis, TN) for downstream processing. Each XML document represents a single patient encounter, which may potentially have multiple notes spanning multiple days at multiple locations. In an example where a patient arrived at ER then transferred to surgery and had a later follow-up visit, each of these three related interactions had an independent provider note but all notes were linked to one encounter ID (FIN). The unstructured text was UTF8 HTML encoded and extracted from the "text" element in the "Assessment and Plan" section of the XML document. The HTML tags were removed and minimal formatting such as tabs and line breaks were kept in the final plain text document. The top 10 document types and corresponding counts were as follows: Clinical Document (311,960), Depart Clinical Summary (15,496), Ambulatory Visit Summary Depart (13,905), Office Visit Note (11,251), ED Clinical Summary (11,232), Pediatric Surgery (10,915), ENT (10,125), pediatric surgery (6,647), Teacher Note (6,462), and Progress Note (5,582). A more extensive list of document types is provided in Appendix B. For each patient, all corresponding unstructured text fields were concatenated into one patient document. Patient documents were then pre-processed using a set of Python scripts to remove: 1) Form letters; 2) Tabulated numeric lab data; 3) Vital signs; and 4) Negation phrases. Negation rules for text processing were developed by Quire and were modified iteratively to achieve high precision for this

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

specific collection. Finally, only the most recent history and physical examination was included in each patient-document. All text processing steps described above were performed on the entire patient cohort (including the non-NSQIP training and NSQIP test sets described below).

Semantic analysis and text-based risk prediction was performed using a proprietary software developed by Quire, which uses a vector space modeling approach called Latent Semantic Indexing[27]. Here, patients were represented as a vector of weighted terms extracted from their medical records. A log-entropy term weighting scheme was used as described by Berry and Browne[27]. Once the term-by-patient matrix was constructed, singular value decomposition (SVD) was performed to reduce the dimensionality of this matrix into lower rank approximation (concept space). We used a rank of 500 in this study based on evaluation of this and other collections. Patient similarities were calculated using the cosine of the vector angles[27].

The Quire Predictive Modeling (QPM) algorithm ranks all patients in a collection based on the reduced-rank vector cosine values to a set of sentinel patients who exhibit the target outcome. A risk score is calculated for each patient based on the percentage of sentinels who have cosine similarities above a preset threshold. In this study, QPM was trained on the non-NSQIP cohort that included 4,738 patients, among whom 48 patients died within 30 days of surgery. A different cohort (NSQIP) of 1,759 patients was used for testing and evaluation. A pseudo-document for each of the NSQIP patients was compared to the 48 non-NSQIP D30 patients to rank and generate risk scores. The risk scores of the non-NSQIP cohort were included in the regression model as described below.

2.2. NSQIP-Ped Cohort: The methodology for NSQIP-Ped sampling of all pediatric surgery cases has been previously published[26]. Over 300 perioperative-standardized variables were collected. Death within 30 days of surgery (D30) was chosen as the main outcome variable. Based on previous work by our group[3], [8], [9], fifteen preoperative variables were identified as risk factors of D30. Dichotomous risk factors included ventilator dependency, oxygen support, previous cardiac intervention, cerebral palsy, open wound with or without infections, neuromuscular disorder, bleeding disorder, hematologic disorder, inotropic support, blood transfusion, malignancy, do-not-resuscitate order, and neonatal status. Case type and sepsis were the two risk factors with more than two categories and were converted to multiple dichotomous risk factors. Finally, a total of 18 dichotomous preoperative risk factors were included in our analysis and were coded as Present (1) Absent (0), or missing. We excluded the American Society of Anesthesiology (ASA) class score since it has been shown to be covariate with most of the included variables, and is itself a risk score. NSQIP-Ped definitions for risk factors and outcomes were used throughout [26].

2.3. Hypothesis Testing and Prediction: We used the Kolmogorov-Smirnov test to check for the normality of data and the Mann-Whitney U test to check whether the distribution of text-based risk variable was the same for different categories of the outcome variables. We implemented stepwise logistic regression analysis with backward elimination in predicting outcomes of cases in the NSQIP-Ped cohort by A) using only text-based risk variables from unstructured data, B) using only 18 NSQIP-Ped risk factors, and C) using all factors in A and B as predictors. C-statistic calculation was used for model prediction performance, and the DeLong test[28] was used to compare the models. In addition to D30, the model was used to predict 11 separate secondary surgery outcomes including death within 90 days of surgery, unplanned readmission, unplanned readmission to operating room, unplanned repeat

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at *Informatics* **2019**, *6*, 4; doi:10.3390/informatics6010004

surgery related to the principle surgery, unplanned second surgery, blood transfusion within 72 hours of surgery start time, postoperative unplanned intubations, postoperative systemic sepsis, septic shock, postoperative superficial incisional surgical site infections (SSI), and postoperative organ/space SSI. We have also implemented 5-fold cross-validation for each logistic regression model to avoid and detect any possible overfitting.

3. Results

This We analyzed inpatient records for 6,497 operative cases performed at Le Bonheur Children's Hospital, Memphis, TN between January 1, 2014 and May 31, 2017 on patients aged 18 years or younger. A total of 719,308 free text notes were available for these cases. On 1,739 of these cases, we also had structured preoperative, operative, and postoperative data from the American College of Surgeons (ACS) National Surgery Quality Improvement Program (NSQIP)[26]. This group of cases makes up the NSQIP-Ped Cohort and used as a testing data set for test-based risk score algorithm. The remaining 4,738 operations in the non-NSQIP-Ped Cohort had text but no accompanying structured data. Therefore, the larger non-NSQIP-Ped group was used as the training data set to develop a text-based risk score and NSQIP cohort was used to evaluate the text-based risk score algorithm and for assessing whether text based risk score improve performance of models to predict surgery outcome when used with other structured data based preoperative risk factors.

Mortality in the NSQIP-Ped and non-NSQIP-Ped cohorts were 11/1759 (0.63%) and 48/4738 (1.01%, $p<0.01$), respectively. Age at operation was younger for the NSQIP-Ped cohort (mean ± standard deviation of 6.4±6.0 years vs 7.1±5.3 years for non-NSQIP-Ped). Gender (55% male for both) and race (47% white vs. 45% white, 40% black vs. 43% black for NSQIP-Ped vs. non-NSQIP-Ped, respectively) were similar in the two cohorts.

3.1. Association between free text-based risk score and death after surgery: The Quire Predictive Model (QPM) approach utilizes a variant of vector-space modeling to represent patients in a cohort as a vector of weighted terms in a reduced rank matrix. Given a set of patients with the target outcome, in this case death within 30 days of surgery (D30), all patients in the cohort can be ranked and scored based on their cosine similarities to the D30=Yes cases. We used the Mann-Whitney U test to compare text-based risk scores since the text-based risk scores were non-normally distributed (Kolmogorov-Smirnov test $p>0.1$). The text-based risk scores were significantly higher ($p<0.001$, Mann-Whitney U Test) for D30 cases compared with those who survived beyond 30 days in both non-NSQIP training set and NSQIP test set (Table 1). D30=Yes cases were concentrated at higher risk scores, both in the training data set (non-NSQIP-Ped, data not shown) and in the test set (NSQIP-Ped) (see Figure 2).

Table 1 Comparison of text-based risk scores for categories of D30

Data Sets	Categories	Count	Mean text-based risk; 95% CI	Mann-Whitney U Test p value
Non-NSQIP-Ped	D30=No	4690	0.35; 0.34-0.36	<0.001
	D30=Yes	48	0.64; 0.59-0.72	
NSQIP-Ped	D30=No	1748	0.44; 0.42-0.46	<0.001

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

	D30=Yes	11	0.84; 0.78-0.90	

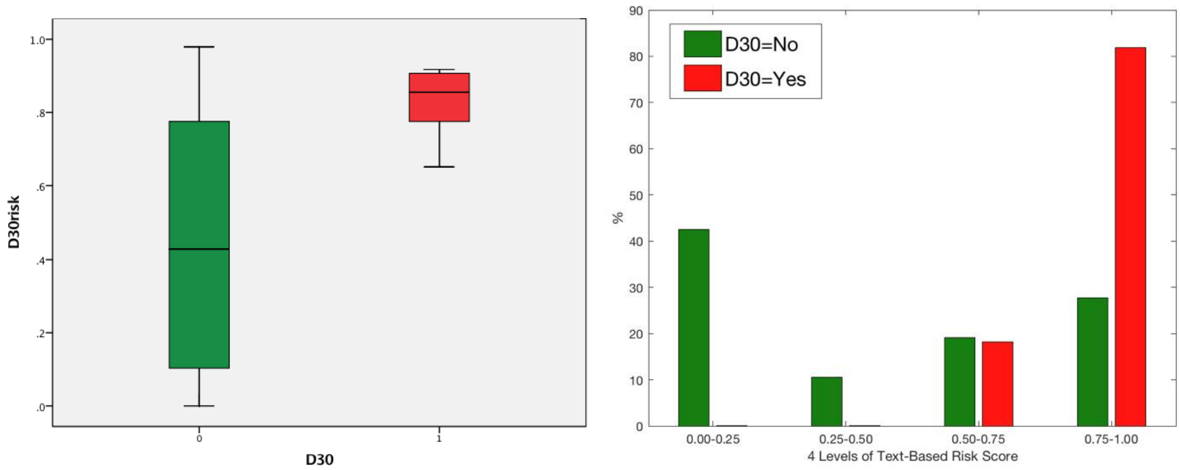


Figure 2 (Left) Text-based risk scores for NSQIP cohort for D30: 0 (No, Green) and 1 (Yes, Red). **(Right)** observed D30 (percentage) for four different levels of text-based risk scores

3.2. Sensitivity Analysis for Text Based Risk Scores: The results above were based on risk scores for NSQIP cohort that were generated by a QPM model developed on non-NSQIP cohort that included 48 D30=Yes cases, with a C-statistic of 0.83 (95% confidence interval (CI) 0.77-0.89). We performed a sensitivity analysis with 5 randomly selected sets of 10, 20, 30, and 40 D30=Yes cases that were projected onto the NSQIP-Ped cohort to calculate mortality risk. These sets had average C-statistic of 0.80, 0.83, 0.83, 0.83 in the NSQIP-Ped cohort, respectively. Figure 3 summarizes the C-statistics obtained for NSQIP cohort at each run of sensitivity analysis. These results suggest that as few as 20 sentinel D30 patients in the training cohort can be effectively used to calculate mortality risk in the test cohort.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at *Informatics* **2019**, *6*, 4; doi:10.3390/informatics6010004

Figure 3 Sensitivity analysis for QPM model

3.3. Prediction of postsurgical mortality in the NSQIP-Ped cohort: We developed three logistic regression models predicting risk of death within 30 days of surgery; Model A using text-based risk score as a single predictor, Model B using 18 risk factors from structured data fields, and Model C using all variables from Model A and B. Model A, a logistic regression model with the text-based risk variable as a single predictor of D30 yielded a C-statistic of 0.83 (0.77-0.89, 95% CI) (Appendix A.1) without cross-validation and 0.81 (0.74-0.88, 95% CI) with 5-fold cross-validation. Model B, a stepwise logistic regression model, identified ventilator status, bleeding disorder, and inotropic support as significant risk factors, yielding a C-statistic of 0.86 (0.69-1.00, 95% CI) (Appendix A.2) without cross-validation and 0.76 (0.54-0.99) with 5-fold cross-validation. However, the difference between C-statistics of models A and B was found non-significant (p>0.1) by DeLong test.

Finally, in Model C, we examined whether a stepwise logistic regression using 18 risk factors from structured data fields and one text-based risk score improved the prediction of D30. The final logistic regression model selected the text-based risk score, ventilator status, bleeding disorder, current receipt of inotropic support, and emergent case as significant risk factors, resulting in a C-statistic of 0.96 (0.92-1.00, 95% CI) (Appendix A.3) without cross-validation and 0.92 (0.84-0.99) with 5-fold cross-validation using the same five selected variables. The values of regression coefficients and odd ratios of these five variables obtained for each run of cross-validation were found to be within the 95% confidence intervals (Appendix A.4). We have also implemented another cross-validation by implementing a stepwise logistic regression model at each run instead of using only the variables selected when a stepwise regression applied on the entire NSQIP data. Out of five runs of 5-fold cross validation process, text-based risk score, ventilator status, bleeding disorder, current receipt of inotropic support, and emergent case were selected in the stepwise logistic regression model 3, 5, 4, 3, 3 times, respectively. However, blood transfusion up to 72 hours prior to surgery, neonate status and septic shock status each were also selected in the logistic regression models one time each (Appendix A.5). We further implemented the DeLong test on C-statistics to confirm that simultaneous use of the text-based variable and NSQIP-Ped variables improved the logistic regression model's performance. The performance of the final logistic regression model including both text-based risk scores and structured data was significantly better than the performance of models

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

using only text-based risk score (p=0.036) and the model using structured data-based risk factors (p=0.055) in terms of C-statistics.

The combined logistic regression model for D30, using a cutoff value of 0.2667, correctly classified 7 of 11 deaths (sensitivity of 63.6%) but also produced only 12 false positives (specificity of 99.3%), yielding positive predictive value (precision) of 36.8% and negative predictive value of ~100.0%. The cutoff value of 0.2667 was selected based on the maximum F1 score, defined as $2 \times Precision \times Recall/(Precision + Recall)$, obtained from the predicted probabilities from cross validation process and applied on the final non-cross validated predicted probabilities. We compared these high risk patients to the preoperative risk score (ranging from 1 to 5) developed by the American Society of Anesthesiology (ASA) class[29]. Eight of these false positives had ASA-class 4 and four had ASA-class 3 assignments. Moreover, on case resulted death within 30-90 days of surgery, one case had received unplanned postoperative intubation, four cases received unplanned blood transfusion within 72 hours of the surgery start time, once case had deep wound disruption, one case had surgery related readmission within 30 days of surgery, and two cases had repeat surgeries related to the primary surgery. These results confirm that the text-based prediction identifies high-risk individuals.

3.4. Association between free text-based risk score and other adverse surgery outcomes: The text-based risk scores for the NSQIP-Ped cohort were also significantly (p<0.05, Mann-Whitney U test) associated with other outcomes such as death within 90 days of surgery, intra- or post-operative blood transfusion within 72 hours of surgery, unplanned readmission within 30 days of surgery, postoperative unplanned intubation, and first unplanned return to operating room (Table 2). In contrast, the text-based risk scores were not significantly associated with post-operative deep organ space surgical site infection.

Table 2 Distribution of text-based risk scores over categories of binary outcomes for the NSQIP-Ped cohort

Outcome		**Count**	**Mean text-based risk value with 95%CI**	**p value**
Death within 30 days of surgery	No	1748	0.44; 0.42-0.45	<0.001
	Yes	11	0.84; 0.78-0.90	
Death within 90 days of surgery	No	1738	0.44; 0.42-0.45	<0.001
	Yes	21	0.82; 0.77-0.87	
Postoperative superficial (incisional) surgical site infection	No	1736	0.44; 0.42-0.45	0.015
	Yes	23	0.62; 0.52-0.74	
Intra- or post-operative blood transfusion within 72 hours of surgery start time	No	1625	0.45; 0.43-0.47	<0.001
	Yes	134	0.31; 0.25-0.37	
Unplanned readmission within 30 days of surgery	No	1621	0.43; 0.41-0.45	<0.001
	Yes	138	0.57; 0.51-0.62	
Postoperative Unplanned Intubation	No	1735	0.43; 0.42-0.45	0.001
	Yes	24	0.71; .061-0.81	
First Unplanned Return to Operating Room	No	1690	0.44; 0.42-0.45	0.039
	Yes	69	0.52; 0.43-0.60	

3.5. The role of free text-based risk score in predicting other adverse surgery outcomes: The text-based risk score (derived for predicting D30) was significantly predictive of death between 30-90 days after surgery (C-statistic 0.96, 0.92-0.99 95% CI), along with additional outcomes such as postoperative superficial incisional surgical site infection, intra- or post-operative blood transfusion within 72 hours of surgery start time, and unplanned readmission within 30 days of surgery (Table 3). Table 3 also includes the 5-fold cross-validation results for each logistic regression model. More details about the logistic regression models can be found in the Appendix A.6-A.10.

***Table* 3** Logistic regression of outcome including text-based risk score as a predictor

Outcome	c-statistics with 95%CI		Selected Preoperative Risk Factors
Death within 30 days of surgery	No CV	0.96; 0.92-1.00	Text-based risk score, ventilator dependency, bleeding disorder, inotropic support, Emergent Case
	5-fold CV	0.92; 0.84-0.99	
Death within 90 days of surgery	No CV	0.95; 0.92-0.99	Text-based risk score, ventilator dependency, neonate, bleeding disorder, Emergent case
	5-fold CV	0.94; 0.89-0.99	
Postoperative superficial incisional surgical site infection	No CV	0.72; 0.61-0.83	Text-based risk score, neonate
	5-fold CV	0.67; 0.55-0.79	
Intra- or post-operative blood transfusion within 72 hours of surgery start time	No CV	0.76; 0.71-0.80	Text-based risk score, oxygen support, neuromuscular disorder, hematologic disorder, inotropic support, malignancy, urgent case
	5-fold CV	0.73; 0.69-0.78	
Unplanned readmission within 30 days of surgery	No CV	0.67; 0.62-0.72	Text-based risk score, neonate, SIRS, Sepsis
	5-fold CV	0.66; 0.61-0.70	

4. Discussion

Our study suggests that unstructured preoperative text available in EMRs contains critical information predictive of postoperative death in children undergoing surgical procedures. Further, these data suggest that information contained in unstructured text notes can be useful even when distilled to a single risk variable developed via a text modeling approach. Finally, we found that the use of text-based risk scores combined with structured data improves the prediction accuracy of death within 30 days of surgery when compared with models using either unstructured or structured data from the NSQIP-Ped database alone.

Data from the unstructured notes is currently used in creating clinically useful risk assessments for surgical procedures. An example is the ASA class that is included as a key variable in the Pediatric Risk Calculator by NSQIP-Ped[30] developed by American Colleague of Surgeons. Automated systems

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

utilizing algorithms such as the one used in this study have the potential to decrease bias introduced by human retrieval and interpretation of such data and may save time for clinicians.

The clinical utility of any risk assessment depends on its accuracy. US health expenditures are higher than other technically advanced countries reporting better objective health outcomes due in part to the provision of expensive care that is unlikely to provide meaningful benefit [31]. Sharing accurate risk estimates with the patient and family is a key component of informed consent and shared decision-making. This allows providers and consumers of surgical care to better weigh alternative treatments that may be less expensive and have equivalent benefit, or to forego treatment in settings where the probability of death after surgery approaches certainty. Formal studies of the impact of clinical decision support tools for surgery are limited, but are needed to determine their impact on practice and patient outcomes.

The integrative prediction model presented here was created to predict the risk of death within 30 days of surgery. The finding that the novel text-based risk score also contributed to accurate prediction of other major surgery outcomes such as death within 90 days of surgery, postoperative surgical site infection, and unplanned blood transfusion and readmission within 30 days of surgery suggest that postoperative adverse events are interrelated. Logistic regression models for each of these outcomes performed almost equally well in 5-fold cross-validation, suggesting that the logistic regression models built on the text-based risk score are robust and generalizable to a broad variety of adverse events despite the challenges of a relatively small sample size and low event rate.

Limitations: Our study has some limitations. The training set included all types of operations while the test set, NSQIP-Ped cohort, systematically excluded some operations[26]. Since the text-based risk scores are calculated based on vectorized combinations of thousands of terms extracted from patient records, the precise words that contribute to postsurgical mortality risk is difficult to deduce. The next step in our work will be to investigate specific keywords that are associated with modifiable risk factors to guide clinicians in developing interventions designed to reduce the risk of severe surgery outcomes.

5. Conclusions

We conclude that text data in EMRs can improve the ability of structured data tools to predict serious patient outcomes after surgery.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Appendix A and B.

Author Contributions: OA and RLD designed and conceptualized the research. OA performed the statistical analysis and drafted the manuscript. RH and KH carried out text data acquisition and text mining. ML provided cleaned structured data and clinical outcomes. All authors contributed to reviewing and writing of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to extend our appreciation to Mark McMath, Anuj Mathur, Deepthi Kurup, and Brian Marshall for making the EMR data available for this study. We also thank Robert Beck for abstracting NSQIP data, Ari Walter and Ruchi Ruchi for gathering and validating data and Shane Morrell for processing the text data and running the Quire applications. We would also like to thank Amada Preston from the LeBonheur Scientific Writing Office and Rich Redfearn and Patti Smith for editing the manuscript.

Conflicts of Interest: Authors KH and RH are affiliated with Quire Inc. (Memphis, TN). Quire personnel were involved in data extraction from the EMR system at Methodist Hospital (Memphis, TN) under a Business Associates Agreement. The other authors have no conflict of interest.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at *Informatics* **2019**, *6*, 4; doi:10.3390/informatics6010004

Appendix

- Appendix A: Supplementary Analysis Results
- Appendix B: Logistic Regression Results for D30 using structured data based risk factors.

References

[1] B. S. Systems, *Health IT and Patient Safety : Building Safer Systems for Better Care*, no. November. 2011.

[2] B. F. Van Der Griend *et al.*, "Postoperative mortality in children after 101,885 anesthetics at a tertiary pediatric hospital," *Anesth. Analg.*, vol. 112, no. 6, pp. 1440–1447, 2011.

[3] O. Akbilgic, M. R. Langham, A. I. Walter, T. L. Jones, E. Y. Huang, and R. L. Davis, "A novel risk classification system for 30-day mortality in children undergoing surgery," *PLoS One*, vol. 13, no. 1, 2018.

[4] E. C. Cheon *et al.*, "Unplanned, Postoperative Intubation in Pediatric Surgical Patients: Development and Validation of a Multivariable Prediction Model," *Anesthesiology*, vol. 125, no. 5, pp. 914–928, 2016.

[5] A. M. Stey *et al.*, "Variation in intraoperative and postoperative red blood cell transfusion in pediatric surgery," *Transfusion*, vol. 56, no. 3, pp. 666–672, 2016.

[6] E. G. Brown, J. E. Anderson, D. Burgess, R. J. Bold, and D. L. Farmer, "Pediatric surgical readmissions: Are they truly preventable?," *J. Pediatr. Surg.*, vol. 52, no. 1, pp. 161–165, 2017.

[7] *Optimal Resources fpr Children's Surgical Care*, vol. v.1. 2015.

[8] M. R. Langham, A. Walter, T. C. Boswell, R. Beck, and T. L. Jones, "Identifying children at risk of death within 30 days of surgery at an NSQIP pediatric hospital.," *Surgery*, vol. 158, no. 6, pp. 1481–1491, 2015.

[9] O. Akbilgic, M. R. Langham, and R. L. Davis, "Race, Preoperative Risk Factors, and Death after Surgery," *Pediatrics*, vol. 141, no. 2, 2018.

[10] A. H. S. Harris, "Path from predictive analytics to improved patient outcomes," *Annals of Surgery*, vol. 265, no. 3. pp. 461–463, 2017.

[11] K. Y. Bilimoria *et al.*, "Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons," *J. Am. Coll. Surg.*, vol. 217, no. 5, 2013.

[12] R. B. Parikh, M. Kakad, and D. W. Bates, "Integrating Predictive Analytics Into High-Value Care," *JAMA*, vol. 315, no. 7, p. 651, 2016.

[13] R. Amarasingham *et al.*, "Consensus Statement on Electronic Health Predictive Analytics: A Guiding Framework to Address Challenges," *eGEMs (Generating Evid. Methods to Improv. patient outcomes)*, vol. 4, no. 1, 2016.

[14] K. Jensen *et al.*, "Analysis of free text in electronic health records for identification of cancer patient trajectories," *Sci. Rep.*, vol. 7, p. 46226, 2017.

[15] J. Lin, T. Jiao, J. E. Biskupiak, and C. McAdam-Marx, "Application of electronic medical record data for health outcomes research: a review of recent literature.," *Expert Rev. Pharmacoecon. Outcomes Res.*, vol. 13, no. 2, pp. 191–200, 2013.

[16] C. Friedman, L. Shagina, Y. Lussier, and G. Hripcsak, "Automated encoding of clinical documents based on natural language processing," *J. Am. Med. Informatics Assoc.*, vol. 11, no. 5, pp. 392–402, 2004.

[17] Q. T. Zeng, S. Goryachev, S. Weiss, M. Sordo, S. N. Murphy, and R. Lazarus, "Extracting principal diagnosis, co-morbidity and smoking status for asthma research: Evaluation of a natural language processing system," *BMC Med. Inform. Decis. Mak.*, vol. 6, 2006.

[18] G. K. Savova *et al.*, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications," *J. Am. Med. Informatics Assoc.*, vol. 17, no.

5, pp. 507–513, 2010.

[19] F. P. Y. Lin, A. Pokorny, C. Teng, and R. J. Epstein, "TEPAPA: A novel in silico feature learning pipeline for mining prognostic and associative factors from text-based electronic medical records," *Sci. Rep.*, vol. 7, no. 1, 2017.

[20] E. Ford, J. A. Carroll, H. E. Smith, D. Scott, and J. A. Cassell, "Extracting information from the text of electronic medical records to improve case detection: A systematic review," *J. Am. Med. Informatics Assoc.*, vol. 23, no. 5, pp. 1007–1015, 2016.

[21] Y. S. Kim *et al.*, "Extracting information from free-text electronic patient records to identify practice-based evidence of the performance of coronary stents," *PLoS One*, vol. 12, no. 8, 2017.

[22] A. A. Thomas *et al.*, "Extracting data from electronic medical records: Validation of a natural language processing program to assess prostate biopsy results," *World J. Urol.*, vol. 32, no. 1, pp. 99–103, 2014.

[23] D. W. Frost, S. Vembu, J. Wang, K. Tu, Q. Morris, and H. B. Abrams, "Using the Electronic Medical Record to Identify Patients at High Risk for Frequent Emergency Department Visits and High System Costs," *Am. J. Med.*, vol. 130, no. 5, p. 601.e17-601.e22, 2017.

[24] G. E. Weissman *et al.*, "Inclusion of Unstructured Clinical Text Improves Early Prediction of Death or Prolonged ICU Stay," *Crit. Care Med.*, p. 1, 2018.

[25] H. J. Murff *et al.*, "Automated identification of postoperative complications within an electronic medical record using natural language processing.," *JAMA*, vol. 306, no. 8, pp. 848–855, 2011.

[26] A. C. of S. N. S. Q. I. Program-Pediatrics, "User Guide for the ACS NSQIP Pediatric Participant Use File," 2013.

[27] M. Berry and M. Browne, *Understanding Search Engines: Mathematical Modeling and Text Retrieval*, 1st ed. Philadelphia: SIAM, 1999.

[28] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics*, vol. 44, no. 3, p. 837, 1988.

[29] S. Merrick and M. Shaker, "ASA Relative Value Guide (RVG): A Defining Moment in Fair Pricing of Medical Services," *ASA Monit.*, vol. 78, pp. 26–27, 2014.

[30] K. Kraemer *et al.*, "Development and Evaluation of the American College of Surgeons NSQIP Pediatric Surgical Risk Calculator," *J. Am. Coll. Surg.*, vol. 223, no. 5, pp. 685–693, 2016.

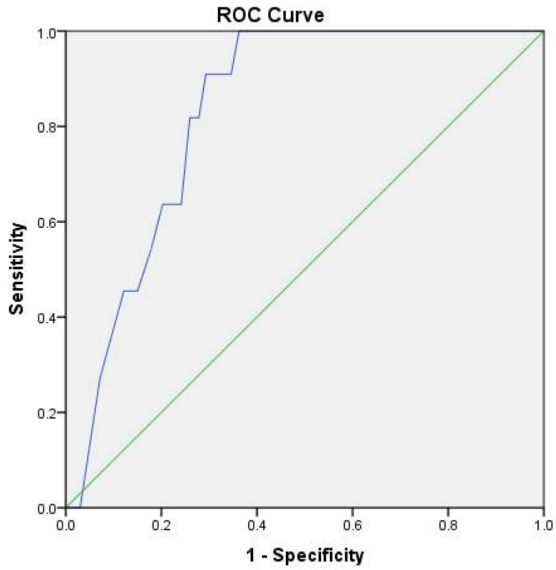
[31] H. Moses, D. H. M. Matheson, E. R. Dorsey, B. P. George, D. Sadoff, and S. Yoshimura, "The anatomy of health care in the United States.," *JAMA*, vol. 310, no. 18, pp. 1947–63, 2013.

Appendix A: Supplementary Analysis Results

1. Logistic Regression Results for D30 using text-based risk score as a single predictor

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B) with 95% CI
Text-based risk	6.22	2.29	7.36	1	0.01	502.32 [5.61-44966.12
Constant	-9.40	1.94	23.52	1	0.00	



Diagonal segments are produced by ties.

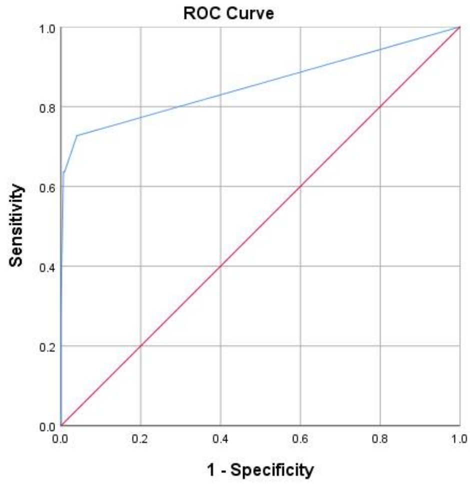
Area Under the Curve

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.83	0.03	0.00	0.77	0.89

2. Logistic Regression Results for D30 using structured data based risk factors

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B) with 95% CI
Ventilator	2.85	.98	8.51	1	0.00	17.21 [2.55-116-41]
Bleeding Disorder	4.18	1.27	10.93	1	0.00	65.50 [55.-781-39]
Inotropic Support	2.66	.99	7.34	1	0.01	14.29 [2.09-97.87]
Constant	-6.50	.61	114.26	1	0.00	0.00



Diagonal segments are produced by ties.

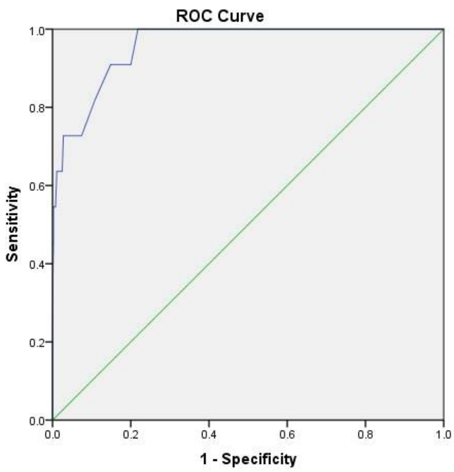
Area Under the Curve

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
0.86	0.08	0.00	0.70	1.00

3. Logistic Regression Results for D30 using both unstructured and structured data

Variables in the Equation

	B with 95% CI	Sig.	Exp(B) with 95% CI
Text-based risk	6.28 (0.42, 12.98)	0.066	533.40 [1.94-146823.50]
Ventilator	2.07 (0.15, 3.99)	0.035	7.96 [1.58-40.00]
Bleeding Disorder	3.56 (1.35, 5.81)	0.002	35.30 [5.33-233.67]
Inotropic Support	2.29 (0.19, 4.39)	0.032	9.84 [1.70-57-11]
Emergent Case	1.75 (0.03, 3.48)	0.046	5.76 [1.34-24.47]
Constant	-11.12 (-16.98, -5.26)	.0000	



Diagonal segments are produced by ties.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

Area Under the Curve

Area	Std. Error[a]	Asymptotic Sig	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
0.96	0.02	0.00	0.92	1.00

4. 5-Fold Cross Validation using the five predictors selected in the Model C

	AUC:Train-Test	Regression Coefficients					
		Constant	Text-Based Risk	Ventilator	Bleeding Disorder	Inotropic Support	Emergent Case
Run 1	**0.98 - 0.84**	**-15.028**	**9.320**	**3.684**	**4.015**	**2.599**	**2.395**
Run 2	**0.95 - 1.00**	**-10.307**	**5.772**	**1.626**	**2.281**	**2.398**	**1.697**
Run 3	**0.95 - 0.99**	**-10.486**	**5.657**	**1.105**	**4.079**	**3.325**	**2.044**
Run 4	**0.95 - 1.00**	**-11.294**	**7.013**	**1.687**	**3.623**	**1.897**	**1.359**
Run 5	**0.96 - 0.87**	**-11.576**	**6.504**	**2.505**	**3.854**	**1.678**	**1.500**

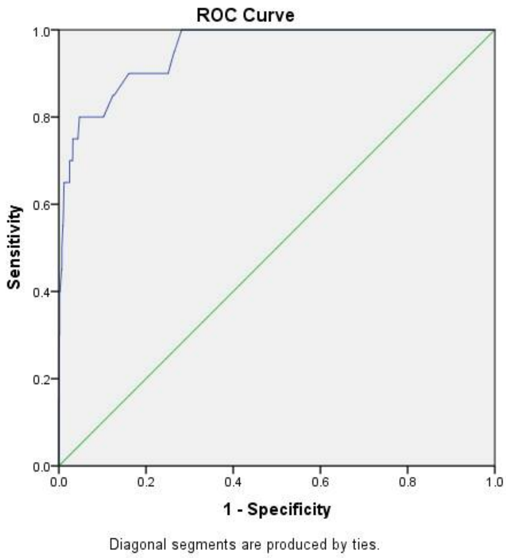
5. 5-Fold Cross Validation by implementing Backward Elimination at each run

	AUC: Train-Test	Selected Variables
Run 1	**0.99 - 0.34**	**Ventilator, Bleeding Disorder, Emergent Case, Neonate**
Run 2	**0.96 - 0.91**	**Text-based risk, Inotropic Support, Emergent Case, Neonate**
Run 3	**0.96 - 0.90**	**Test-based risk, Bleeding Disorder, Inotropic Support, Emergent Case**
Run 4	**0.94 - 1.00**	**Text-based risk, bleeding disorder, inotropic support, blood transfusion**
Run 5	**0.87 - 0.84**	**Ventilator, Bleeding Disorder, Septic Shock**

6. Logistic Regression Results for D90 using both unstructured and structured data

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B) with 95% CI
Ventilator	2.98	0.61	23.75	1	.00	19.64 [5.93-65.04]
Bleeding Disorder	2.91	0.98	8.80	1	.00	18.27 [2.68-124-52]
Neonate	1.41	0.58	5.85	1	.02	4.09 [1.31-12.80]
Emergent Case	1.41	0.67	4.48	1	.03	4.11 [1.11-15.25]
Text-based risk	4.42	1.96	5.09	1	.02	83.31 [1.79-3886.07]
Constant	-9.23	1.68	30.25	1	.00	.00



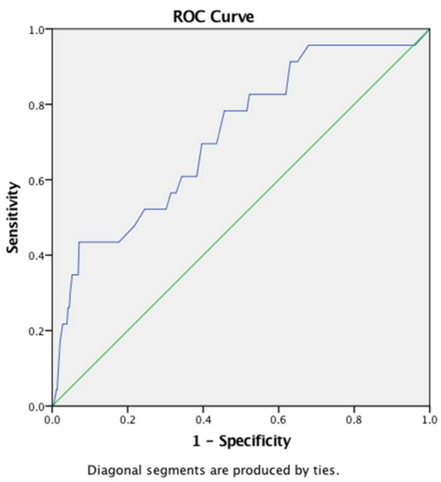
Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.95	0.02	0.00	0.92	0.99

7. Logistic Regression Results for Postoperative Superficial Incisional Surgical Site Infection using both unstructured and structured data

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B) with 95% CI
Text-based risk	1.42	0.74	3.66	1	0.05	4.13 [0.97-17.66]
Neonate	1.65	0.44	14.22	1	0.00	5.19 [2.21-12.21]
Constant	-5.49	0.51	115.29	1	0.00	



Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

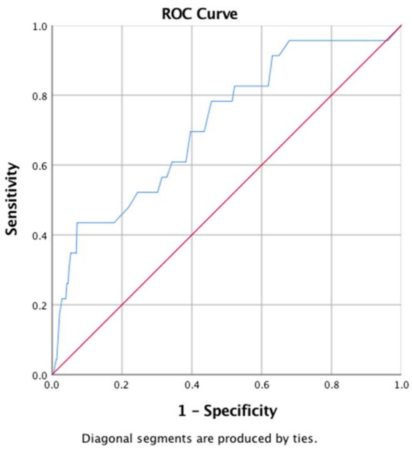
Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004

			Lower Bound	Upper Bound
0.72	0.06	0.00	0.61	0.83

8. Logistic Regression Results for Postoperative Superficial Incisional Surgical Site Infection using both unstructured and structured data

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B) with 95% CI
Text-based risk	1.42	0.74	3.66	1	0.05	4.13 [0.97-17.66]
Neonate	1.65	0.44	14.22	1	0.00	5.19 [2.21-12.21]
Constant	-5.49	0.51	115.29	1	0.00	0.00



Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval Lower Bound	Upper Bound
0.72	0.06	0.00	0.61	0.83

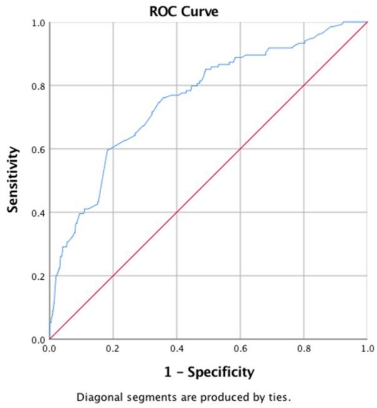
9. Logistic Regression Results for Intra or Postoperative Blood Transfusion within 72 hours of Surgery Start Time using both unstructured and structured data

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B) with 95% CI
Oxygen Support	1.06	0.40	7.092	1	0.01	2.87 [1.32-6.25]
Neuromuscular Disorder	1.31	0.24	30.42	1	0.00	3.72 [2.33-5.94]
Hematologic Disorders	1.20	0.33	12.96	1	0.00	3.32 [1.73-6.38]
Inotropic Support	2.53	0.62	16.76	1	0.00	12.52 [3.73-41.98]
Childhood Malignancy	0.52	0.25	4.116	1	0.04	1.67 [1.02-2.75]
Urgent Case	-1.95	0.74	7.014	1	0.01	.14 [0.03-0.60]
D30risk	-2.24	0.35	40.35	1	0.00	0.106
Constant	-2.11	0.15	198.08	1	0.00	0.12

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at Informatics 2019, 6, 4; doi:10.3390/informatics6010004



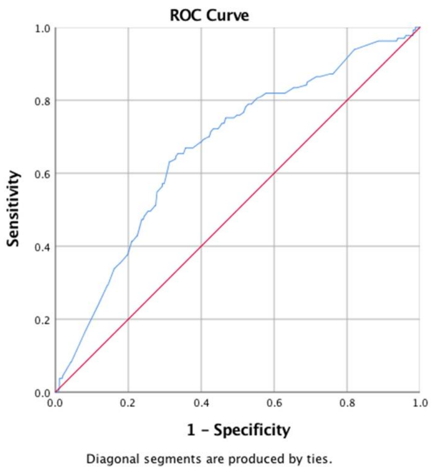
Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
0.76	0.02	0.00	0.71	0.80

10. Logistic Regression Results for Unplanned Readmission within 30 days of Surgery using both unstructured and structured data

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B) with 95% CI
SIRS	0.89	0.43	4.32	1	0.04	2.43 [1.05-5.63]
Sepsis	0.82	0.33	6.31	1	0.01	2.27 [1.20-4.49]
Neonate	-1.20	0.43	7.82	1	0.01	0.30 [0.13-0.70]
Text-based risk	1.54	0.29	28.14	1	0.00	4.67 [2.64-8.25]
Constant	-3.28	0.20	272.16	1	0.00	0.04



Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
0.67	0.02	0.00	0.62	0.72

 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at *Informatics* **2019**, *6*, 4; doi:10.3390/informatics6010004

Appendix B: Top 100 Document types with text which were extracted from the EMR.

Doc Type	TOTAL
Clinical Document	311,960
Depart Clinical Summary	15,496
Ambulatory Visit Summary Depart	13,905
Office Visit Note	11,251
ED Clinical Summary	11,232
Pediatric Surgery	10,915
ENT	10,125
pediatric surgery	6,647
Teacher Note	6,462
Progress Note	5,582
Surgery	5,292
Parenteral Nutrition	5,138
General Message	5,119
Oto-HNS	4,903
NICU Progress Note	4,724
Parenteral Nutrition Service (269-0112)	4,603
PICU Progress Note	4,481
LeBonheur Pathology Final Report	3,968
NSR	3,896
Progress	3,619
IMCU	3,348
progress	3,239
Ortho	3,171
attempt	3,090
IMCU Daily Progress Note	2,917
Team D Progress Note	2,907
AMB Visit Summary Depart	2,891
Palliative Care Progress Note	2,841
ortho	2,800
Team A Progress Note	2,750
follow up	2,674
PT note	2,592
Parenteral Nutrition (269-0112)	2,573
Discharge Summary	2,417
OtoHNS	2,396
Team B Progress Note	2,386
IMCU PN	23,17
IMCU daily	2,268
general surgery	2,223
Parenteral Nutrition 269-0112	2,172
Neurology	2,081
Team C Progress Note	1,962
IP-D	1,916
Discharge	1,884
surgery	1,845
MNT Follow Up	1,827
EEG	1,693
discharge	1,692
Inpatient ST Progress Note	1,676
Discharge Planning	1,647
Attempt	1,606
Pulmonology	1,574

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0678.v1

Peer-reviewed version available at *Informatics* **2019**, *6*, 4; doi:10.3390/informatics6010004

Urology	1,560
DS	1,470
Pulmonology Attending Note	1,469
Teacher note	1,458
Trauma NP	1,450
PRS	1,418
PICU Daily Progress Note	1,380
up date	1,375
IMCU Daily PN	1,321
PICU Daily Note	1,244
Pet Therapy	1,241
Peds Surgery	1,221
Peds Surg	1,203
Contact	1,198
NSx	1,150
Daily Progress Note	1,148
NICU	1,134
neurosurgery staff	1,133
Assessment	1,117
Ped Nephrology	1,094
Cardiology	1,066
FEFH	1,066
CVICU Progress Note	1,035
tx note	1,033
peds surgery	1,024
PICU Accept Note	1,023
Neurosurgery	978
Abdominal pain	962
Team D PN	937
Teaching-Supervisory Addendum-Expand *ED	928
GI Progress Note	924
Fever	904
Plastic Surgery	897
IMCU note	887
Consult	876
ID Progress Note	869
Neurosurgery staff	855
Neurosurgery NP note	844
GI Daily Note	837
Vomiting	823
nsr	821
NICU Daily Progress Note	820
DC Planning	813
Pediatric Nephrology	798
Cardiology Office Visit Note	790
GI	782
Parenteral Nutrition Service	769
IP-D PT Note	759

Note that this study utilizes only preoperative text notes. Any post-operative document in the list above were belong the previous surgeries.