Feature Extraction and Visualization of Respiratory Therapist Notes for Pediatric Long-Term Ventilator Dependent Patients

Nathan M. Pajor, MD*
Division of Pulmonary Medicine
Cincinnati Children's Hospital Medical Center
University of Cincinnati College of Medicine

Lindsay Nickels, PhD
Digital Scholarship Center
University of Cincinnati

Ezra Edgerton, BS, BA
Digital Scholarship Center
University of Cincinnati

Danny T.Y. Wu, PhD, MSI
Department of Biomedical Informatics
University of Cincinnati College of Medicine

Dan T. Benscoter, DO
Division of Pulmonary Medicine
Cincinnati Children's Hospital Medical Center
University of Cincinnati College of Medicine

James J. Lee, PhD
Digital Scholarship Center
University of Cincinnati

Abstract

Children with long-term ventilator dependence are a growing population that generate substantial cost to the healthcare system and require very lengthy admissions to initiate support. Respiratory therapy notes contain free-text descriptions of key respiratory events during these admissions but are underutilized. Using a retrospective electronic health record data set from 101 patients, we identified more clinically concerning patients, extracted key features from the free-text notes that differentiated these patients, and displayed those features in a timeline visualization that has implications for clinical decision support.

Keywords: linguistic analysis. information visualization. pediatrics

1 Introduction

Children with long-term ventilator dependence (LTMV) are a growing, complex population that incur substantial expense with aggregate charges in 2006 estimated at $1.5 billion, a 70% increase compared to 2000[1]. Admissions to initiate ventilator support are lengthy, lasting between 140-200 days on average [2]. The prolonged stays and the complexity of the patients generate a massive amount of electronic health record (EHR) data. Among these data, clinical care notes are particularly underutilized. At our tertiary care pediatric center, a random sample of 50 LTMV patients had generated an average of 2,035 lifetime notes, over 2.5 times the number generated by an age-matched cohort of Cardiology patients and over 12 times the number generated by a general pediatric population with at least one hospital admission. Respiratory Therapy notes are brief notes filed during an inpatient stay that contain a free-text summary of important respiratory events or behaviors that are often not captured as discrete data elements elsewhere in the EHR. Timeline visualization of clinical data has long been proposed as a method to interpret EHR data [3,4]. However, fewer visualizations utilize unstructured clinical notes, particularly for prolonged inpatient admissions [5]. We hypothesize that key features from these notes can be identified and used in a timeline visualization to help summarize lengthy hospitalizations and help to differentiate more clinically concerning patients.

*nathan.pajor@cchmc.org

2 Methods

We identified all children less than 2 years old with tracheostomy dependence and long-term ventilator dependence initiated between 2011-2018 at our large, tertiary care children's hospital. In order to establish a subjective rating of a patient's stability during their prolonged hospital stay, seven experienced providers (2 physicians, 2 advanced practice nurses, 2 respiratory therapists and 1 registered nurse) reviewed a list of all included patients and subjectively and independently rated their respiratory status as stable or unstable on a 0-10 continuous scale based on their recollection of that patient's initial or index hospital stay. If a provider did not recognize a patient, they were asked to indicate this on the survey. These scores were averaged by patient to identify the 10 most stable and 10 most unstable patients in order to identify clear differences between the two subgroups. RT notes were extracted from the EHR for the index admission for each patient. The notes were first manually reviewed for content and themes. Then, a keyword analysis was performed using WordSmith tools on the two corpora. The output, in conjunction with clinical expertise, was used to identify terminology salient to each corpus with respect to its counterpart (features). This analysis resulted in the identification of terms of interest that denote specific respiratory behaviors that are relevant to differentiating more clinically concerning patients. Word stems were used to identify variants of these terms in the full note corpus. A web-based timeline visualization of each patient's admission was developed using D3.js based on the results of the above analysis.

3 Results

We included 101 patients (median length of stay: 190 days; range: 63-654 days) and their 22,711 RT notes in our retrospective analysis. The survey had an average of 3.99 (SD 1.56) responses per patient. The mean stability score was 3.90 (SD 2.38). We identified the 10 most unstable and stable patients among the 101 patients. The mean score for the 10 unstable patients was 9.15 and 0.42 for the stable patients. The 10 most unstable and stable patients had 2581 and 2078 RT notes, respectively. By clinical expert review of terms included in the notes, we defined a set of terms that were clinically relevant for describing severe events (e.g. desat, work of breathing, bag) and terms describing stability (e.g. stable, window). The raw frequency and distribution of identified terms of interest in these RT notes are summarized in Table 1 and 2. These terms of interest are listed as their stemmed forms in order to capture the frequency of all variants used to describe their corresponding behaviours. Comparing the terms of interest in both tables, it seems that a finite set of terms can be used to distinguish stable and unstable patients. Moreover, unstable patients are more likely to be associated with **desat** (oxygen desaturation), **WOB** (work of breathing), and **bag** (manual bag ventilation) and less likely with **window** (time off the ventilator).

Unstable Patients			
event	raw freq.	# notes	% notes
desat*	893	796	31%
secretion*	514	454	18%
suction* (+variants)	870	737	29%
WOB (+variant)	212	172	7%
bag*	139	117	4.6%
window	39	28	1%
stable	2035	1503	58%

Table 1. Terms of interest and relative frequency/distribution of the 10 most unstable patients (2581 notes)

Stable Patients			
event	raw freq.	# notes	% notes
desat*	555	506	24%
secretion*	430	375	18%
suction* (+variants)	715	593	29%
WOB (+variant)	82	72	3.5%
bag*	57	50	2.5%
window	231	151	7%
stable	1976	1379	66%

Table 2. Terms of interest and relative frequency/distribution of the 10 most stable patients (2078 notes)
(*word stem, desat = oxygen desaturation, WOB = work of breathing, bag = manual bag ventilation)

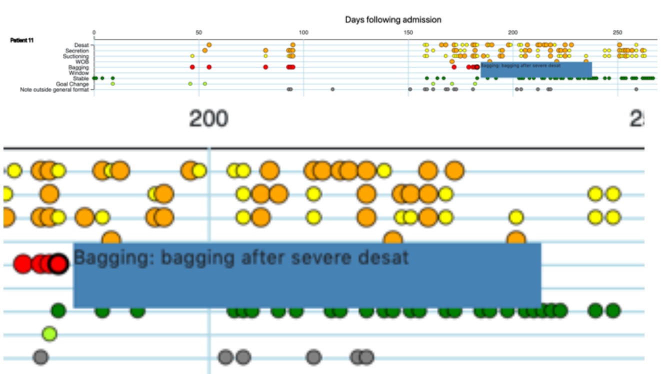

Figure: Top: Timeline visualization for a single unstable patient hospital stay >250 days in duration demonstrating salient terms appearing in free text RT notes Bottom: Expanded view demonstrating contextual information when hovering over a single node.

interest identified in our analysis and described in Table 1 and 2. Each note containing a salient term was represented as a node on the visualization. If multiple salient terms are contained within a single note, then a node appears for each term at the same time point. Nodes were color-coded with yellow/orange/red for clinically concerning terms and green for clinically reassuring terms. Expert providers determined that **bagging** events were the most critical and were represented in red. Yellow was used to identify **desat**, **secretion**, **suction** and **WOB** events that had benign modifiers such as few desats, thin secretions, mild WOB. Orange was used when modifiers such as moderate to severe desats, thick or copious secretions or moderate to severe WOB were used. The clinically reassuring terms were displayed in green (The variable shades of green were used only to more easily identify the different terms and are not associated with variation in degree.) Hovering over any node shows an annotation with the note context in order to provide improved understanding of the event.

4 Discussion & Conclusion

We found differences in RT note content for patients subjectively rated as stable or unstable by experienced providers. Those features which differentiate the stable and unstable note corpora were used to generate an interactive timeline visualization which can summarize otherwise difficult to access RT note information from prolonged hospital stays and has implications for clinical decision support. We are conducting work to pair the features identified in analysis with structured data extracted from the EHR and evaluating correlation with multiple clinical outcomes. For example, we are evaluating the transition from ICU ventilator to portable home ventilator in this patient population and have developed a predictive model using logistic regression and machine learning based on structured EHR data such as patient demographics, vital signs and ventilator settings. Expert providers have identified severe respiratory events as a potential factor in determining success or failure with this transition. By incorporating features identified in unstructured notes such as bagging events that aren't documented elsewhere in the EHR, we anticipate being able to improve our model's predictive performance. Additionally, we are developing methods to automate the extraction of key terms to test this visualization as a clinical tool that providers can use to better understand patterns over prolonged admissions in these patients.

References

[1] Benneyworth BD, Gebremariam A, Clark SJ, Shanley TP, Davis MM. Inpatient Health Care Utilization for Children Dependent on Long-term Mechanical Ventilation. Pediatrics. 2011;127(6):E1533-E41.

[2] Baker CD, Martin S, Thrasher J, Moore HM, Baker J, Abman SH, et al. A Standardized Discharge Process Decreases Length of Stay for Ventilator-Dependent Children. Pediatrics. 2016;137(4):9.

[3] Steve C, Cousins S, Cousins SB, Kahn MG, editors. The Visual Display of Temporal Information. Artificial Intelligence in Medicine; 1991: Citeseer.

[4] Bui AAT, Aberle DR, Kangarloo H. TimeLine: Visualizing Integrated Patient Records. IEEE Transactions on Information Technology in Biomedicine. 2007;11(4):462-73.

[5] Tang C, Plasek JM, Zhang H, Kang M-J, Sheng H, Xiong Y, et al. A temporal visualization of chronic obstructive pulmonary disease progression using deep learning and unstructured clinical notes. BMC Medical Informatics and Decision Making. 2019;19(8):258.