Investigating Resuscitation Code Assignment in the Intensive Care Unit using Structured and Unstructured Data

Sharon L. Lojun, MD, SM[1], Christina J. Sauper, SM[2], Mitchell Medow, MD, PhD[1], William J. Long, PhD[2], Roger G. Mark, MD, PhD[2], Regina Barzilay, PhD[2]
[1]Boston University Medical Center, Boston, MA, [2]Massachusetts Institute of Technology, Cambridge, MA

Abstract

This study investigates the feasibility of using structured data (age, gender, and medical condition), and unstructured medical notes on classification accuracy for resuscitation code status. Data was extracted from the MIMICII database. Natural language processing (NLP) was used to evaluate the social section of the nurses' progress notes. BoosTexter was used to predict the code-status using notes, age, gender, and Simplified Acute Physiology Score (SAPS). The relative impact of features was analyzed by feature ablation. Unstructured notes were the greatest single indicator of code status. The addition of text to medical condition features increased classification accuracy significantly (p<0.001.) N-gram frequency was analyzed. Gender differences were noted across all code-statuses, with women always more frequent (e.g. wife>husband.) Logistic regression on structured features was used determine gender bias or ageism. Evidence of bias was found; both females (OR=1.45) and patients over age 70 (OR=3.72) were more likely to be Do-Not-Resuscitate (DNR).

Introduction

Most patients come to ICU care after a sudden change in health, rather than by a foreseen medical episode. Resuscitation code status is generally assigned as Full Code (FC), until it is possible to sort out the likely prognosis and obtain information about a patient's wishes. In many cases, assignment of less aggressive code statuses (e.g., Do-Not-Resuscitate (DNR) or Comfort-Measures-Only (CMO)) is not easy because the outcome of treatment is unknown. Additional difficulties arise when patients are incapacitated; in these cases, it falls to the family to make a decision, informed by the physician's recommendations and expectations

The concept of advanced directives (AD), or living wills, has sought to help in making a patient's wishes known and followed; however they are sometimes vague and unable to predict all possible clinical scenarios. Additionally, AD have not been very successful in the United States, in comparison to, e.g., Japan, where they have been more successful.[1] Cultural differences and many other reasons are cited as the causes. In the U.S., the likelihood to have an AD is more likely with advanced age and increased income.[2] However, some elderly are interested only in discussing Cardio-Pulmonary-Resuscitation (CPR), but do not necessarily want their wishes committed on paper.[3] Joos found, in a self-administered questionnaire of general medical patients, 72% had knowledge of AD, 53% discussed with family, and only 14% had discussed with their physician.[4] Half of the patients felt the terminology should be simplified.[4] So, it may be possible in many instances that the family has the best understanding of the patient's wishes. Of note, the majority of geriatricians do not establish AD.[5]

Even in the case of AD, it is often difficult to predict whether stabilization will occur with brief critical care interventions; and therefore difficult for the physician to interpret the AD in all situations. As a result, less aggressive code assignment usually does not occur until after entry into the ICU. In many cases, the patient is unable to communicate due to treatment or illness. The assignment of code status would then be made by the closest family relative in conjunction with the medical staff.

In this study, a large ICU database was used to determine the relationship of medical, familial, and social factors in the assignment of resuscitation code status. Three analyses were conducted: first, investigation of the feasibility of incorporating unstructured data (medical notes) in combination with structured medical data as a predictor of code status assignment; second, performance of an n-gram frequency analysis on the unstructured data to seek potentially influential factors; third, creation of a model with logistic regression, using the structured data only, to determine the most relevant medical features. In addition, throughout all three studies, the evaluation of the possibility of ageism and gender bias is considered.

Data

Database MIMIC II Database (an ongoing NIH-sponsored Bioengineering Research Partnership (NIBIB BRP 5RO1EB001659) including investigators at MIT, Philips Medical Systems, and Boston's Beth Israel Deaconess Medical Center.) was used. The database is a repository of information from multiple critical care units. It includes ICU information (observations, measurements, interventions, and ICU daily notes from all services except physicians), and hospital medical information (laboratory results, medications prescribed, and hospital discharge summaries.) The data are de-identified, and reformatted. The database contains information from over 30,000 patient admissions (from over 26,000 unique patients.)

Dataset Data extraction included adults (age greater than 15years) from all critical care areas, and was stratified according to code status. For patients who transitioned from FC status, to DNR, or to CMO, the last recorded code status was used. Do-Not-Intubate (DNI) patients were grouped with DNR. No cases were identified where a less aggressive code status transitioned to a more aggressive one, e.g. CMO to FC. Total number of ICU admissions included 17,548 (FC); 2060 (DNR); 784 (CMO). The remainder of the dataset consisted of patients from the Neonatal Intensive Care Unit (NICU), which were excluded from the study.

Structured Data Structured data includes the code status, age, and gender. In addition, Simplified Acute Physiology Score I (SAPSI) was calculated from the dataset.[6] It is by definition calculated on the first day of ICU admission, SAPSI day 1 (1). SAPSI is calculated from 14 biologic or physiologic parameters, such as age, heart rate, systolic blood pressure, temperature; and includes the Glasgow coma scale, which itself is comprised of several physiologic parameters and waas calculated from the dataset. In order to augment the medical measures, SAPSI was calculated for day two (2), and the difference between the two SAPS scores was calculated as the delta (D). These three measures were used to quantify the patient's overall medical condition.

Unstructured Data The medical notes were derived from daily free text input from all services, except physicians (which were not available in the dataset.) The notes were limited to the social sections (only) of the nursing progress notes. By convention, this section catalogues family visits; meetings with physicians; and overall description of understanding by the family of the patient's condition. Some typical excerpts include: "very supportive family has been in to visit today, wife and children," "family all in agreement that they want him to be extubated and not to be re-intubated, comfort will be main goal if he fails," "family meeting planned," "no family contact this shift." Text entries from all social sections of a single admission were tied to the respective code status and demographic information, so that unstructured and structured data were evaluated concurrently.

Methods

Institutional Review Board (IRB) approval was obtained.

Preprocessing The text entries of the daily medical notes were preprocessed in the following way: First, stop words ("and", "the", etc.), rare words (those appearing fewer than 5 times in the entire corpus), and words directly indicating code status ("DNR", "full code", "comfort measures", etc.) were removed. The Porter stemming algorithm was used to stem the remaining words (converting "sons" to "son", etc.). Extraction of key phrases was considered; however, it was not obvious what types of phrases would be most relevant. Therefore, the entire text was used in order to avoid biasing the classifier.

Data was randomized into a training set (80%) and a test set (20%), ensuring a consistent representation of each code status in both sets.

BoosTexter Classification The first study is an analysis of classification accuracy using BoosTexter[7], a freely available classification package. BoosTexter uses a boosting algorithm to classify text and feature attributes. In particular, this study attempts to visualize the relative contribution of medical notes to improve prediction accuracy among different subsets of patients. It is also a study in relative contribution to code prediction; that between the feature of unstructured medical notes and that of the medical condition.

To accomplish this, a feature ablation study was performed on the data. For each combination of medical features (SAPS scores and delta), BoosTexter was run with and without the social medical notes as a feature.

Statistical significance was calculated using McNemar's test on differences in classification error for each combination of features with and without medical notes. E.g., significance was tested between

SAPSI(1) and delta without text and SAPSI(1) and delta with text; and likewise, for all other combinations of medical metrics.

N-Gram Analysis The second study deals with n-grams in the medical notes. Specifically, the ratios of male and female visitors were analyzed to determine whether patterns existed. The count of each unigram for the male and female variations of relatives (spouse, child, sibling, parent, grandparent, grandchild, aunt/uncle, and niece/nephew) was calculated for each code status relative to the total number of unigrams for that code status.

Logistic Regression The third study is a logistic regression calculation which used the structured attributes to predict divisions in code status (FC vs. DNR and FC vs. DNR+CMO) using the R statistical framework. Each attribute was tested individually and with any confounding attribute. The odds ratios, confidence intervals and p-values were calculated.

Results

BoosTexter Classification Figure 1 represents code status classification error as computed by BoosTexter for all ablations of medical metrics with and without notes as a feature. Statistical significance at all medical metrics was demonstrated with $p < 0.001$.

In each case, the notes had a profound effect on classification accuracy. This may imply that the text contains more information relevant to determining code status than medical condition does. Since the text primarily consists of a record of social visits and meetings of the physician with the family, this could possibly represent an effect of family or physician sentiment.

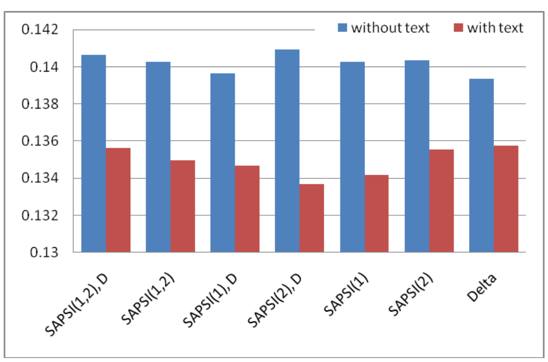

Figure 1: BoosTexter error rate over the 3-way code status assignment with varying combinations of medical metrics; impact of text demonstrated. Difference at each combination is significant with $p < 0.001$.

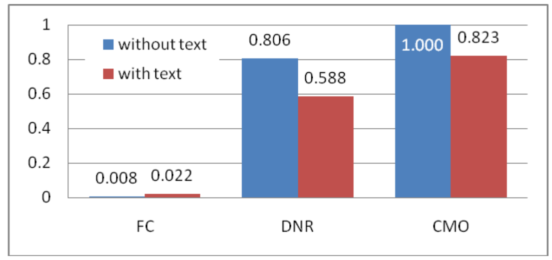

Figure 2: BoosTexter error rate stratified by code status, using medical metrics SAPSI(2) and Delta. There is significant error reduction in both DNR and CMO codes.

Figure 2 demonstrates the data point from figure 1 (SAPSI(2), Delta), stratified by each of the three resuscitation code statuses. The contribution of the medical notes significantly reduces classification error ($p<.001$) for codes DNR and CMO. In the case of FC, there is a small increase in error; this is due to the classification algorithm taking more risks and attempting to distinguish other code statuses, rather than to classify nearly everything as FC.

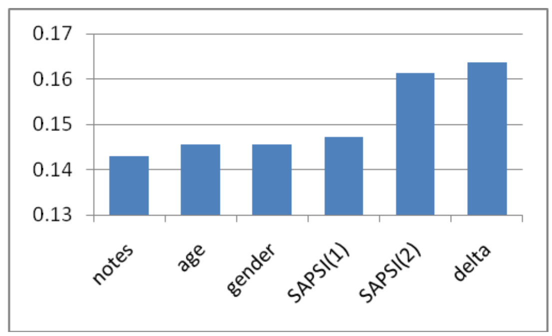

Figure 3: BoosTexter error rate with single features; full data set. The medical notes feature yields the lowest error rate among all single features.

Figure 3 demonstrates the classification errors for each feature individually. The lowest error rate is found using the medical notes. Surprisingly, social text, gender, and age are all more informative for classification than the medical metrics provided by the SAPS scores.

N-Gram Analysis Figure 4 demonstrates the frequency of gender-specific words for spouse, parent, and child in the social text as a ratio of word count / total words in the corpus. Comparisons were made within code groups and between groups. There was a marked gender difference in each case across all code statuses. This may suggest that there is more daily support from female relatives while in the ICU.

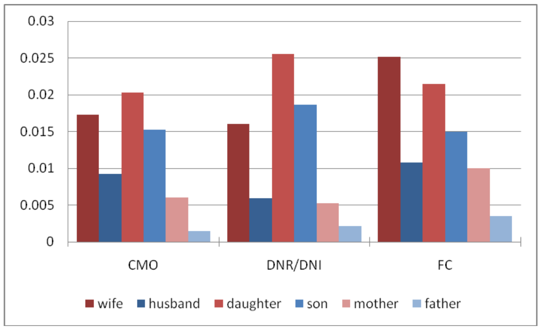

Figure 4: Most frequent visitors by percentage of total corpus, stratified by gender of each relative type (spouse, child, parent) for each code status. The female relative of each pair is more frequently mentioned across all codes.

	Odds Ratio	*Conf. Intervals*	*p-value*
Gender = Female			
FC vs. DNR	1.45	1.31-1.60	< 0.001
FC vs. DNR+CMO	1.35	1.24-1.47	< 0.001
Age ≥ 70			
FC vs. DNR	3.72	3.35-4.13	< 0.001
FC vs. DNR+CMO	2.91	2.66-3.18	< 0.001
Age ≥ 70 and Gender = Female			
FC vs. DNR	4.19	3.62-4.87	< 0.001
FC vs. DNR+CMO	3.36	2.95-3.82	< 0.001
Age ≥ 70 and Gender = Male			
FC vs. DNR	3.08	2.65-3.58	< 0.001
FC vs. DNR+CMO	2.40	2.12-2.72	< 0.001

Table 1: Odds ratios of code status of female and older patients based on logistic regressions, computed separately for FC vs. DNR (CMO ignored) and FC vs. both DNR+CMO. All numbers control for medical status as indicated by SAPSI(1) score. E.g., 1.45 odds ratio for Gender = Female indicates that women are more likely to be DNR than men, even when controlling for medical condition. Note that there is a clear age and gender bias.

Logistic Regression Table 1 shows the odds ratios of gender and age bias in code status. The gender evaluation was controlled for age and medical condition (SAPS scores), and the age evaluation was controlled for medical condition. Less bias appears in CMO status, but the greatest bias is found when comparing FC to DNR statuses. For the age evaluation, categorical sets were calculated for age < 70 and age ≥ 70. The age bias was greatest amongst women, which is not surprising given the existing gender bias; however, there was marked bias amongst older males as well.

Discussion

This is the first study, known to the authors, to evaluate prediction of resuscitation code status using nursing notes. The nursing notes (unstructured data) alone proved to be a better indicator of code status than the available medical statistics. Age and gender were also highly predictive. When combined with medical features, unstructured medical notes significantly improve classification accuracy.

There has been some previous work comparing the outcomes of structured and unstructured data. Lai et al. described a study of influenza-like illness and gastrointestinal disease in ambulatory electronic health records (EHR).[8] The authors of that work found that it was feasible to use unstructured and structured EHR data to perform syndromic surveillance. The structured data was more reliable than the unstructured data, and there were no studies demonstrating the combination of both types of data. Turchin et al used both structured and unstructured (EHR) data to evaluate anti-hypertensive medication intensification.[9] They found that use of both data types was complimentary in evaluation of anti-hypertension. These studies differ from the present study in which there was clear classification superiority using the narrative data. Unlike these previous results, this is the first finding, known to the authors, which favors narrative text over structured data. In addition, the current work shows further benefit when unstructured text is combined with the structured features.

The specific differences in gender involvement are interesting. It cannot readily be concluded that the female relatives (wife, daughter, sister) have a more decisive role than males (husband, son, brother); but it is reasonable to conclude that there is more daily involvement, via visits or phone calls. This relationship is consistent across all code states.

Using logistic regression, a distinct bias in both age and gender was demonstrated, even while controlling for medical condition as represented by SAPSI. These results are consistent with those of Eachempati, et al, who found a gender bias in DNR assignment for elderly patients undergoing emergency surgery.[10] Bardach et al found that women and Hispanics were more likely to have DNR order, and when adjusted for, hospital mortality rates reversed the advantage to Hispanics.[11] It is possible that women themselves more frequently chose DNR status. Covinsky et al, as part of the SUPPORT project, found women less likely to want CPR.[12] However, Perkins et al showed that women are more open to invasive treatments.[13] The gender difference is especially concerning, given the study by Zettel-Watson et al.[14] In this study, wives were found to be more accurate compared with husbands regarding their spouse's wishes. It seems equally plausible, therefore, that the gender difference may be a reflection of the cultural devaluation of women.

Conclusion

These findings highlight four main points. First, combining unstructured and structured data as input to machine learning can improve classification accuracy compared with using either type of data individually. Second, unstructured medical notes are more predictive of code status than any combination of gender, age, or SAPS medical metric. Third, there is decidedly more daily involvement from family members of female gender for patients of any code status; however, the significance of this is unclear. Finally, female and older patients are disproportionately likely to have less aggressive code statuses, even when controlling for medical condition. This warrants in-depth further study.

References

1. Matsui M. Perspectives of elderly people on advance directives in Japan. Journal of Nursing Scholarship. 2007; 2:172-6.
2. Rosnick CB, Reynolds SL. Thinking ahead: factors associated with executing advance directives. Journal of Aging &Health. 2003; 2:409-29.
3. Watson DR, Wilkinson TJ, Sainsbury R, et al. The effect of hospital admission on the opinions and knowledge of elderly patients regarding cardiopulmonary resuscitation. Age & Ageing. 2007; 6:429-34.
4. Joos SK, Reuler JB, Powell JL, Hickam DH. Outpatients' attitudes and understanding regarding living wills. Journal of General Internal Medicine. 1993; 5:259-63.
5. Lester PE, Sykora A, Wolf-Klein GP, et al. Do Geriatricians Practice What they Preach?: Geriatricians' personal establishment of advance directives. Gerontology & Geriatrics Education. 2009; 1:61-74.
6. Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. Crit Care Med. 1984; 12: 975-977.
7. Schapire RE, et al. BoosTexter: A boosting-based system for text categorization. Machine Learning. 2000;39:135-168.
8. Lai AM, Friedman C, Calman NS, et al. Syndromic surveillance using ambulatory electronic health record data. JAMIA. 2009;16:354-361.
9. Turchin A, Shubina M, Breydo E, et al. Comparison of information content of structured and narrative text data sources on the example of medication intensification. JAMIA. 2009;16:362-370.
10. Eachempati SR, Hydo L, Shou J, et al. Sex differences in creation of do-not-resuscitate orders for critically ill elderly patients following emergency surgery. Journal of Trauma-Injury Infection & Critical Care. 2006; 1:193-7.
11. Bardach N, Zhao S, Pantilat S, et al. Adjustment for do-not-resuscitate orders reverses the apparent in-hospital mortality advantage for minorities. American Journal of Medicine. 2005; 4:400-8.
12. Covinsky KE, Fuller JD, Yaffe K, et al. Communication and decision-making in seriously ill patients: findings of the SUPPORT project. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. Journal of the American Geriatrics Society. 2000; (5 Suppl):S187-93
13. Perkins HS, Cortez JD, Hazuda HP. Advance care planning: does patient gender make a difference? American Journal of the Medical Sciences. 2004; 1:25-32.
14. Zettel-Watson, Ditto PH, Danks JH, et al. Actual and perceived gender differences in the accuracy of surrogate decisions about life-sustaining medical treatment among older spouses. Death Studies. 2008; 3:273-90