Clinical Significance of Marital Status and Changes in Status Extracted from Unstructured Clinical Notes

Dmitry Scherbakov, PhD[1], Paul M. Heider, PhD[1], Leslie Lenert, MD[1]
[1] Medical University of South Carolina, Charleston, South Carolina, USA

Introduction

Stressful family events, such as divorce, can be important predictors of mental and physical health[1,2]. However, this information is often not collected during clinical encounters. In many cases, inferences based on changes in the static structured field "marital status" is the only source of such information. Yet, the reliability of structured marital field is questionable. Arguably, more insight into marital status can be obtained by analyzing unstructured clinical notes. In this project we built an ensemble natural language processing (NLP) model to extract marital status from clinical notes and to test whether marital status and changes in the status are associated with selected health outcomes.

Methods

We use the structured field "marital status" from admissions in Medical Information Mart for Intensive Care dataset (MIMIC-III)[3] and two publicly available NLP models from University of California (UC): one based on bag of words and neural network-based[4] (with social history sections of discharge summaries in MIMIC-III as inputs to these models) – to create an ensemble model which assigns to each admission a marital status label based on the majority vote. To improve precision for the "partnered" group of patients, we used an additional fine-tuned BERT model, which we trained on publicly available MIMIC-III annotations related to living arrangement[5]. This model is used to detect whether a patient is living with their partner or not (all partnered patients who do not live with their partner are coded into a separate group called "dating"). We test the performance of the models using 650 manually annotated social history sections. We then use an ensemble model to bin patients into groups of similar age based on their marital status. From this initial sample, we selected patients with at least two intensive care (ICU) admissions no more than 5 years apart. Using the marital status labels from each admission, we further classified patients into groups based on changes in marital status: "Change in Marital Status" and "Same Marital Status". We downsampled the larger cohorts, so that mean age in all groups is similar. For the static marital status patients, we compared occurrences of depression, multiple ICU admissions, drug abuse, and alcohol abuse, along with mean Elixhauser-van Walraven (E-VW) score using derived MIMIC-III data based on International Classification of Diseases (ICD-9) codes. For the dynamic marital status patients, we compared the same health outcomes (except multiple ICU admissions) and additional stress-related health markers calculated from the derived MIMIC-III data: changes in blood pressure, changes in weight, changes in glycated hemoglobin (HbA1C) levels. Our analysis code is available online[6].

Results

The ensemble model achieved the highest weighted precision – 0.83, while the structured field had the lowest – 0.66. For our purpose of selecting patients from the dataset for outcomes analysis, higher precision is the most important. Precision and recall per class for each model are shown in Figure 1. Table 1 shows the downsampled group sizes for our static and dynamic patient cohorts.

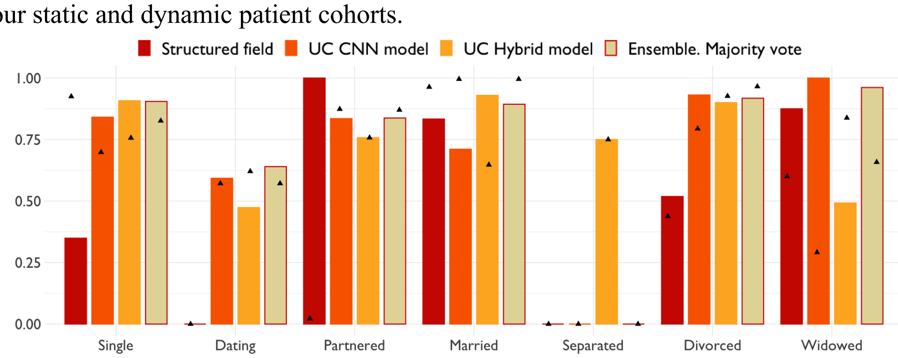

Figure 1 Precision of different models across marital status groups based on 650 notes. Recall displayed as triangle shapes. UC Hybrid is model based on bag of words. CNN = convolutional neural network.

Table 1. Resulting group sizes after sampling from MIMIC-III corpus. Separated group not used. Mean age ~57 years.

Group	Group size	Group	Group size	Group	Group size
Single	262	Married	3774	Same Marital Status	2500
Dating	159	Widowed	62	Change in Marital Status (Got married)	40
Partnered	486	Divorced	592	Change in Marital Status (Other change)	99

As displayed in Figure 2, married patients showed the lowest occurrence of depression, drug abuse, and alcohol abuse. Mean E-VW score was higher for divorced patients when compared to married patients (not displayed), but comparison with other groups didn't reach statistical significance. As shown in Figure 3, patients whose marital status changed showed a higher occurrence of getting a new diagnosis of depression, alcohol abuse, and drug abuse, and a higher occurrence of recovery from these conditions. Similar patterns to Figure 3 were found when assessing separately patients who got married vs. no change group, and when assessing additional health conditions ad-hoc, such as metastatic cancer occurrence and hypertension (not displayed). There was not enough statistical power to detect difference when comparing multiple ICU admissions for static groups; and when comparing changes in blood pressure, weight, HbA1C, E-VW score for "Same Marital Status" vs "Change in Marital Status" groups.

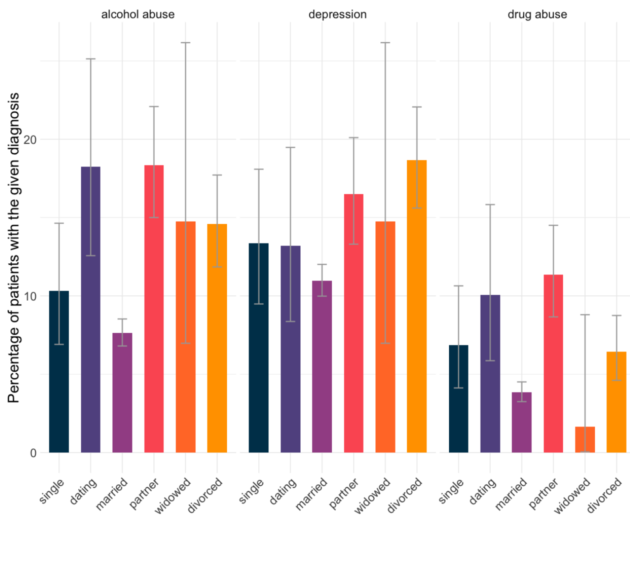

Figure 2. Incidence of depression, alcohol and drug abuse. All figures are displayed with 95% CI.

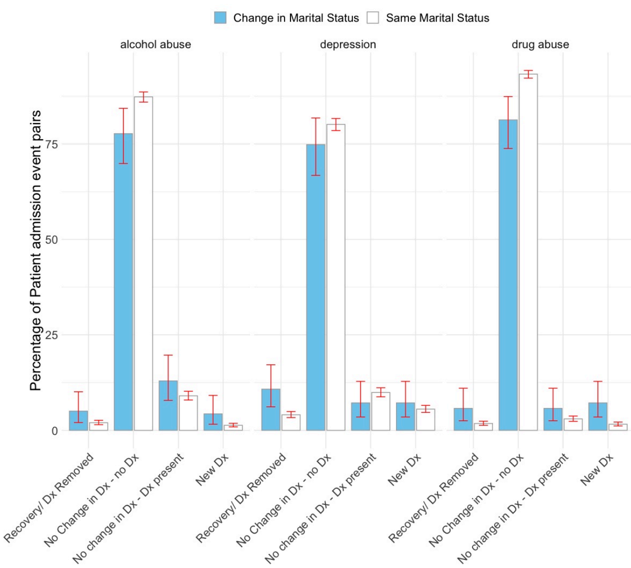

Figure 3. Incidence of depression, alcohol and drug abuse among patients who had a change in marital status vs no change group

Discussion

In agreement with previous studies, we have found that married patients have better health outcomes than other groups. However, when considering change in status, we found that in the sampled MIMIC-III ICU population, which is older and distinct from the general population, experiencing any change in marital status, including getting married, was associated with higher incidences of depression, drug abuse, and alcohol abuse. This finding indicates that marriage and other stressful events can have a significant impact on health. As a next step we plan to validate the results on a larger dataset with newer ICD-10 codes. An essential part of future research is taking steps to protect vulnerable groups, such as people in marriage-like relationships, from enforcing stigma related to their health outcomes.

Conclusion

We have demonstrated how to utilize NLP models to extract marital status and status changes from previously unseen clinical notes and how to use the labeled data in clinical analysis.

References

1. Salleh MR. Life event, stress and illness. The Malaysian journal of medical sciences: MJMS. 2008 Oct;15(4):9.
2. Hobson CJ, Kamen J, Szostek J, Nethercut CM, Tiedmann JW, Wojnarowicz S. Stressful life events: A revision and update of the social readjustment rating scale. International journal of stress management. 1998 Jan;5:1-23.
3. Johnson A, Pollard T, Mark R. MIMIC-III Clinical Database (version 1.4). PhysioNet. 2016. Available from: https://doi.org/10.13026/C2XW26
4. Lituiev DS, Lacar B, Pak S, Abramowitsch PL, De Marchis EH, Peterson TA. Automatic extraction of social determinants of health from medical notes of chronic lower back pain patients. Journal of the American Medical Informatics Association. 2023 Aug 1;30(8):1438-47
5. Lybarger K, Ostendorf M, Yetisgen M. Annotating social determinants of health using active learning, and characterizing determinants using neural event extraction. J Biomed Inform. 2021;113:103631.
6. Scherbakov D. MIMIC-III-MaritalStatus. https://github.com/scherbakovdmitri/MIMIC-III-MaritalStatus. 2024.