Original Paper

Kim Huat Goh[1#], Le Wang[1], Adrian Yong Kwang Yeow[2], Ding Yew Yoong[3], Lydia Au Shu Yi[4], Hermione Poh[5], Ke Li[5], Joannas Jie Lin Yeow[5], Gamaliel Yu Heng Tan[4,5]

[1]Division of Information Technology and Operations Management, Nanyang Technological University, Singapore

[2]School of Business, Singapore University of Social Sciences, Singapore

[3]Tan Tock Seng Hospital, Singapore

[4]Ng Teng Fong General Hospital, Singapore

[5]Medical Informatics, National University Health System, Singapore

[#]*Corresponding Author:*

Associate Professor Kim Huat Goh

S3-B2A-34, 50 Nanyang Avenue, Singapore 639798

akhgoh@ntu.edu.sg

Mining of Clinical Notes for Readmission Prediction in Geriatric Patients

Abstract

Background

Prior literature suggests that psychosocial factors adversely impact health and healthcare utilization outcomes. However, psychosocial factors are typically not captured by the structured data in electronic medical records (EMR) but recorded as free text in different types of clinical notes.

Objective

Propose a text-mining approach to analyze electronic medical records to identify older adults with key psychosocial factors that predict adverse health care utilization outcomes – measured by 30-day readmission. The psychological factors are appended to the LACE Index for Readmission to improve the prediction of readmission risk.

Materials and Methods

We conduct a retrospective analysis using EMR notes of 43,216 hospitalization encounters in a hospital from 01 Jan 2017 to 28 Feb 2019. Descriptives (mean; standard deviation) of the entire cohort include age (67.51; 15.87), length of stay (5.57; 10.41); ICU stay (5%; 22%). We employed text-mining techniques to extract psychosocial topics which are representative of these patients and test the utility of these topics in predicting 30-day hospital readmission beyond the predictive value of the LACE Index for Readmission.

Results

We observed that the added text-mined factors improved the AUROC of the readmission prediction by 8.46% for geriatric patients, 6.99% for the general hospital population, and 6.64% for frequent admitters. Medical social workers and case managers capture more psychosocial text topics compared to physicians.

Discussion

The results show the feasibility of extracting psychosocial factors from EMR clinical notes and the value of these notes in improving our readmission risk prediction.

Conclusion

Psychosocial profiles of patients can be curated and quantified from text-mining clinical notes and these profiles can be successfully applied to AI models to improve readmission risk prediction.

Keywords: readmission risk, geriatrics, electronic medical record, psychosocial factors, text-mining

INTRODUCTION

Hospital readmission of older adults is a significant challenge for the individual, caregivers, and health system. For individuals, readmissions can be distressing, may compromise quality of care, and increase the risk of adverse health outcomes. For caregivers, it is often burdensome and increases their healthcare spending. As for health systems, readmissions often cause resource demands and financial costs to escalate [1]. The 30-day readmission rate among patients aged 65 years or older in Singapore has been reported to be 19%, [2] which is comparable to the readmission rate of Medicare patients in the United States, most of whom are older adults [3]. Significant risk factors for hospital readmission in adults aged 65 years and older include a) socio-demographic factors such as higher age, male gender, some specific ethnicity, and poor living conditions, b) health-related factors such as poor overall condition, comorbidity, functional disability, and recent hospital admissions, and c) organizational factors such as prolonged length of stay in the index hospitalization and discharge destination [4, 5]. These risk factors have been used extensively in predictive models for hospital readmission by health services researchers across the world [6-10]. Recently, other predictors such as those in the psychosocial domain have begun to receive more attention.

Psychosocial factors can be defined as "the combination and interplay of psychological and social factors that potentially influence health, injury, illness, and disease" [11]. A review of various medical literature suggests that different medical specialties have slightly different definitions of psychological factors [11-17]. Based on the various factors identified in earlier studies, we observed that psychosocial factors can be divided into three relevant dimensions: 1) individual psychological well-being, 2) social structures, and 3) resources. Individual psychological well-being include psychological conditions such as mood [11, 18], attitude [11, 19], coping mechanism [11, 17], depression [15, 16, 20], perceived control [13, 19], and psychological distress [16, 17, 21]. Social structures represent the conditions of the environment in which the individual lives, and they include support structures [11, 14, 16, 17], social relationships [14, 18], social norms [19], and family life [22]. Finally, resources represent the means available to the individual, such as financial means, accessibility to healthcare [13, 14] and the health service system [19].

Research has found that these factors—depressive symptoms [23], poor social support, and financial stress—contribute to hospital readmission for specific patient subgroups such as those with chronic obstructive lung disease, chronic kidney disease, and heart failure [24-26]. In general, psychosocial factors could play a significant role in hospital readmission of older adults and account for a significant proportion of readmission risk. At the same time, psychosocial factors are indicators of a patient's complex needs that are amenable to tailored care interventions. Such interventions can improve the patient's clinical outcomes and reduce utilization of healthcare resources.

Objective

This study proposes a text-mining approach to identify older adults with key psychosocial factors from the clinical notes to help predict adverse health and health care utilization outcomes. To validate the efficacy of including psychological factors in the predictive model, we append these psychosocial factors to the commonly used LACE Index for Readmission [27] to improve readmission risk prediction accuracy on an independent, hold-out sample of patients.

Literature Review

There are two conceptual models in extant literature that link psychosocial factors to hospital readmissions for older adults. The first is Anderson's [28] Behavioral Model of Health Services Use that posits an individual's use of health services as a function of predisposing, enabling, and need factors. Psychosocial factors i.e., individual-level and structural-level variables can be categorized as the model's predisposing and enabling factors respectively. The other is Adler and Stewart's [29] model of Pathways Linking Socioeconomic Status and Health. It suggests that environmental resources and constraints, as well as, psychological influences are mechanisms in the pathways model that lead to health outcomes such as hospital readmission. Individual-level and structural-level psychosocial factors map to the model's psychological and environmental variables.

Whereas many clinically-related risk factors are stored as structured data in electronic medical records (EMR), most psychosocial factors are recorded as free text in the patient's clinical notes such as the initial and progress clinical notes of physicians, allied health professionals, case managers, and social workers. Thus, such unstructured, textual data in the EMR represents a potentially rich and untapped source of data related to patients' psychosocial factors. The manual extraction of psychosocial keywords from unstructured data is challenging and impractical given the copious and ever-increasing amount of clinical notes recorded in a typical EMR system. As such, there have been systematic efforts by clincians to capture social and behavioural data, including psychosocial information, as structured data in EMR systems [30]. Yet the effectiveness of these efforts in different healthcare contexts remains to be seen. At the same time, other researchers have begun to apply text mining techniques to efficiently extract and analyze unstructured text data in EMR clinical notes to identify these psychosocial factors.

Text mining techniques are a broad range of technologies for analyzing and processing semi-structured and unstructured text data to construct structured data. By using powerful algorithms applied to large textual documents such as those typically found in EMR systems, text mining can "turn text into numbers" to be used for further analysis. Topic modeling, which is a specific domain area in text mining that examines individual words to identify common topics and concepts, holds significant promise for extracting psychosocial factors from EMR clinical notes.

To date, only a few text mining studies have set out to identify individuals with psychosocial factors using EMR data [31]. As such, we have limited evidence on how effective extraction of psychosocial information from EMR can be used for the purpose of secondary healthcare research or routine clinical care.

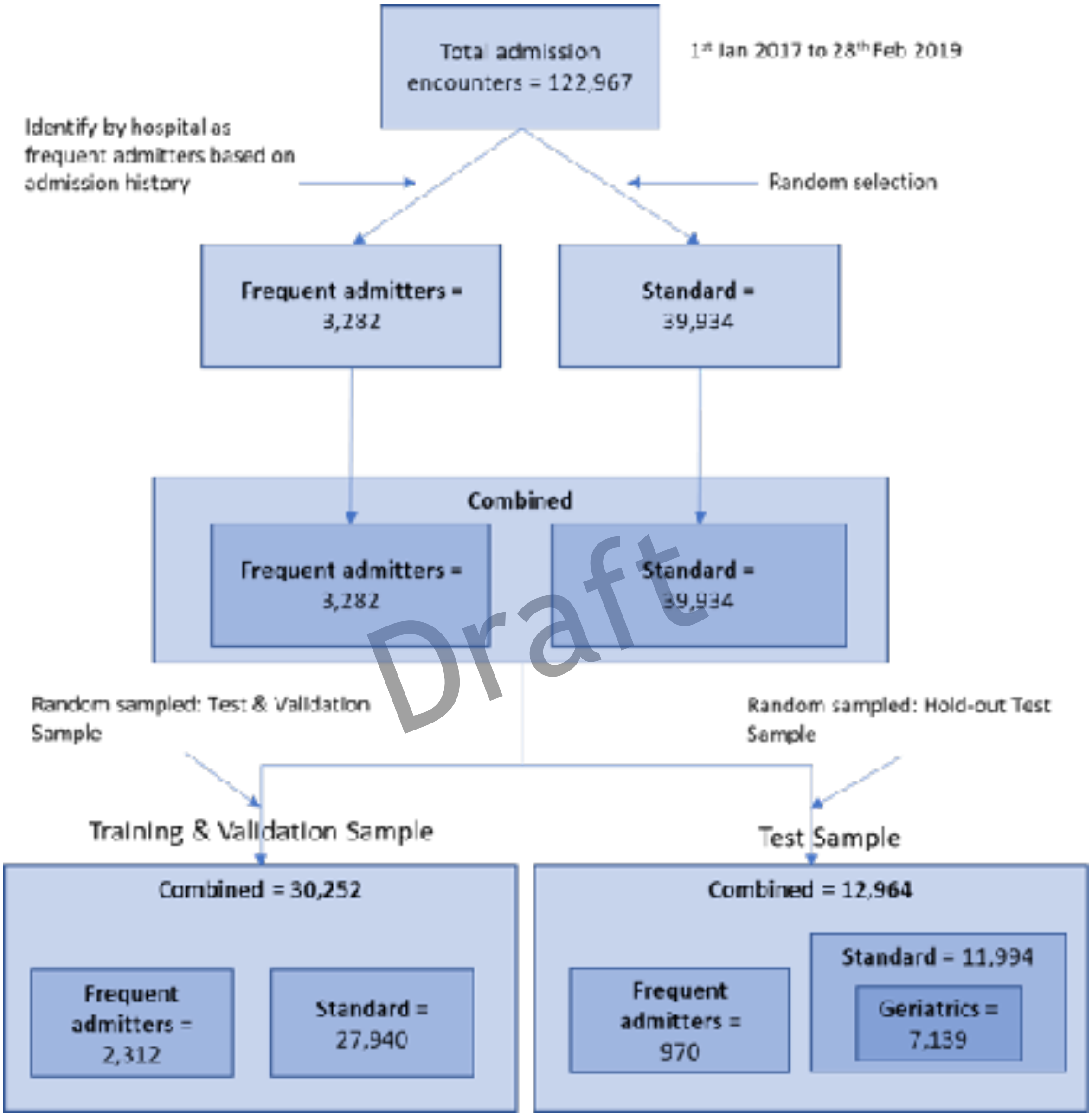
MATERIALS AND METHODS

The study is a retrospective analysis of electronic medical records captured by the EPIC® EMR system over a 26-month period from 01 Jan 2017 to 28 Feb 2019. Ethics approval was provided by the Domain-Specific Review Board (DSRB) of the National Healthcare Group, Singapore (NHG DSRB Ref: 2018/01072).

Settings and Data Context

The sample consists of 9,393 patients with 43,216 admission encounters in a 26-month period from all wards in Ng Teng Fong General Hospital, Singapore. Each clinical record is classified by the role of the note's author. In this sample, clinical records were authored by physicians, medical social workers, or case managers. Specifically, medical social workers and case managers are assigned to some patients who may require additional social support upon hospital admission. The study consists of two cohorts of patients— the first cohort consists of 892 patients (3,282 admission encounters) identified by the hospital as frequently readmitted patients (cohort: “Frequent Admitters”). This cohort consists of patients who 1) are frequently admitted to the hospital despite having their acute medical needs met, or 2) have medical conditions that require multi-disciplinary care, or 3) show signs of caregiver stress, or 4) encounter frequent falls (more than two falls in last 12 months) and require functional management at home, or 5) face medication management issues (e.g., non-compliance to their medication regime). The second cohort consists of 9,377 randomly selected patients (39,934 admission encounters) admitted to the hospital (cohort: “Standard”). The “Standard” cohort consists of patients admitted to the hospital’s inpatient wards during the sampled time period. The purpose of having the “frequent admitters” cohort is to oversample the frequent readmission cases and facilitate the training of the text-mining algorithm to extract psychosocial topics often associated with readmission risks. Such oversampling method is commonly applied in healthcare research to train machine learning algorithms [32, 33]. The combined cohorts are randomly split into a training dataset and a hold-out/test dataset to ensure that both the training and test dataset are of similar distribution containing both “frequent admitters” and “standard” cohorts. The training dataset consists of 30,252 admission encounters, and the hold-out/test dataset consists of 12,964 admission encounters. The unit of analysis is each admission encounter.

We use the 10-fold cross-validation method to train and validate the model with the training dataset. The validated model is subsequently tested against the hold-out/test dataset in four ways. First, we use the “Combined” test dataset - the dataset with a similar distribution as the training dataset, which contains proportionally more “frequent admitters” to ensure that we test the model using a similar distribution of patients as the one used to train the model. Second, we use the "Standard" test dataset - the sample, which is a random draw of patients from the hospital. This sample represents a typical patient that a hospital encounters. We use this to test for the generalizability of the model and to rule out over-fitting. Third, we use a "Frequent" test dataset: this sample consists of "frequent admitters" – mainly to test the fit of the model in predicting frequent admitters; a key concern for many hospitals. Finally, we use a "Geriatrics" test dataset – this sample consists only of geriatric patients (patients ≥ 65 years old from the "standard" cohort test dataset) – to test if the model works for the geriatric specialty where this model is likely to be deployed. Additional details and procedures of the cohort selection are shown in Figure 1.



Note: All values represent the admission encounters. For "Frequent admitters" cohort (3,282 encounters), there are 892 unique patients and for "standard" cohort (39,934 encounters), there are 9,377 unique patients. "Geriatrics" cohort represents a subsample of patient encounters within the "Standard" cohort where the age of the patient at time of encounter is greater or equal to 65.

Figure 1. Sampling Methodology

Data Processing and Algorithm Development

For ***each*** admission encounter, we combined the clinical notes written by authors of similar roles, e.g., all notes written by physicians were combined as physician's notes. The notes are combined based on the author's role (physician, medical social worker, and case manager) because each role would potentially document similar issues. Hence, it is more efficient to mine each role's unstructured clinical notes to identify common or similar topics. We combine the notes for each admission encounter instead of analyzing each note entry as a unit of analysis because a patient's psychosocial conditions are less likely to vary for each admission encounter.

We then apply natural language processing (NLP) text-mining to the clinical notes in the training dataset. We use the Latent Dirichlet Allocation (LDA) topic modeling algorithm to extract the common topics present in the clinical notes and then numerically weigh each topic's intensity (loadings) in the clinical notes. A vector of lexicographically related words represents each **topic** due to these words' frequent occurrence in proximity across different notes. A high **loading** represents the presence of the topic in the clinical note. This routine was performed separately for the physician, medical social worker, and case manager notes. 100 topics were extracted from each set of notes based on the clinician's role (i.e., physician, medical social worker, case manager).

Two geriatric specialists reviewed and classified these 100 topics into broader themes, specifically dividing them into psychosocial issues or non-psychosocial-related issues see Table 2. Additionally, we conducted four interviews with a group of medical social worker and case managers to triangulate if this classification is appropriate. It is important to note that this added classification into broader themes by clinicians is solely to facilitate reporting and interpretation of results. These broader themes are ***not used*** in the subsequent development of the readmission risk model, and only the LDA classification loadings are used in the training of the readmission risk model.

We combine the topic's intensity (loadings) for each set of notes with structured predictors of readmission established in the LACE Index for Readmission as predictors for estimating readmission risk. As readmission risk is a function of various factors beyond psychosocial factors, we incorporated the LACE Index to take into account some of the factors found in the literature. The LACE Index is a score commonly used to predict a patient's 30-day hospital readmission risk [27]. The index consists of the following variables: 1) the length of stay (L), 2) the acuity of the current or previous admission (A), 3) comorbidities of the patient as measured by the Charlson Comorbidity Index score (C), and 4) the number of visits to the emergency department in the preceding six months (E).

The readmission risk model was fitted using gradient boosting trees algorithm to predict the outcome of readmission within the next 30-days from the discharge date of the current admission. Gradient boosted trees (GBT) use an ensemble of multiple trees to generate more accurate prediction models for classification and regression. The algorithm's premise is to build a series of trees, where each tree is trained with the objective to correct the misclassification errors of the previous tree in the series.

We test the model's predictive accuracy by evaluating it using four different hold-out test samples described earlier. To assess the clinical notes' predictive value, we fitted a LACE baseline readmission model ***without*** using the topics from the notes. We then compared this baseline model against models that include the physician notes and social notes (i.e., medical social worker notes and case manager notes) jointly and separately.

RESULTS

Evaluating the Predictive Value of Psychosocial Information

As expected, we observed that physicians record fewer psychosocial issues than medical social workers and case managers (refer to Table 1). For conciseness, the distribution of the specific topics extracted can be obtained from the appendix.

Table 1. Distribution of Psychosocial Topics

Role of Author	% of Psychosocial Topics	% of Non-Psychosocial Topics
Physician	25%	75%
Medical Social Worker	100%	0%
Case Manager	88%	12%

The descriptive statistics of the variables used in the readmission risk model can be seen in Table 2. Here we report the descriptive statistics for each of the test cohorts.

Table 2: Descriptive Statistics of Variables in LACE Readmission Model (Patient Encounter Level)

Variable Name	Cohort[1]	Mean	Standard Deviation
Age	Frequent	72.94	13.24
	Standard	67.07	15.98
	Geriatrics	77.62	8.09
	Combined	67.51	15.87
Gender (1: Male, 0: Female)	Frequent	0.50	0.50
	Standard	0.55	0.50
	Geriatrics	0.50	0.50
	Combined	0.54	0.50
Length of Stay (in Days)	Frequent	6.73	13.18
	Standard	5.47	10.15
	Geriatrics	6.65	10.81
	Combined	5.57	10.41
Charlson Comorbidity Index (CCI)	Frequent	0.47	1.39
	Standard	0.41	1.18
	Geriatrics	0.49	1.38
	Combined	0.42	1.20
Emergency Department admission (1: Yes, 0: No)	Frequent	0.50	0.50
	Standard	0.57	0.50
	Geriatrics	0.58	0.49
	Combined	0.56	0.50
ICU stay (1: Yes, 0: No)	Frequent	0.03	0.17
	Standard	0.05	0.22
	Geriatrics	0.04	0.21
	Combined	0.05	0.22

Variable Name	Cohort[1]	Mean	Standard Deviation
Emergency Department visit in last 6 months (Count of visits)	Frequent Standard Geriatrics Combined	2.86 1.39 1.47 1.50	3.07 2.77 2.44 2.82

Note: [1] Cohort represents the patient sample. "Frequent" represent the patients identified by the hospital as frequent readmission patients. "Standard" represents the sample of a typical hospital patient. "Geriatrics" represents the sub-set of patients in the "Standard" sample who are 65 years or older. "Combined" represents the combination of both "Frequent" and "Standard" samples.

The area under the ROC (AUROC) of the LACE baseline predictive model ranges from 0.8288 to 0.8397 for the four different test cohorts (Frequent, Standard, Geriatrics, and Combined). The baseline model only considers common factors identified in prior literature associated with readmission risks and does not include psychosocial factors extracted from the clinical notes. The ROC is a plot representing the diagnostic ability of a binary classifier as one varies the discriminatory threshold (i.e., the cut-off value to re-classify one state to the other). With varying different discriminatory threshold values, the different sets of true positive rate (sensitivity) with the corresponding false positive rates (1-specificity) are plotted on the axes. Thus, the area under the ROC is a representation of the overall performance of the classifier.

Adding the text-mined notes from the medical social worker and case manager increases the model's AUROC to between 0.8573 and 0.8707. Further appending the clinical notes from physicians increase the AUROC to between 0.8952 and 0.9100.

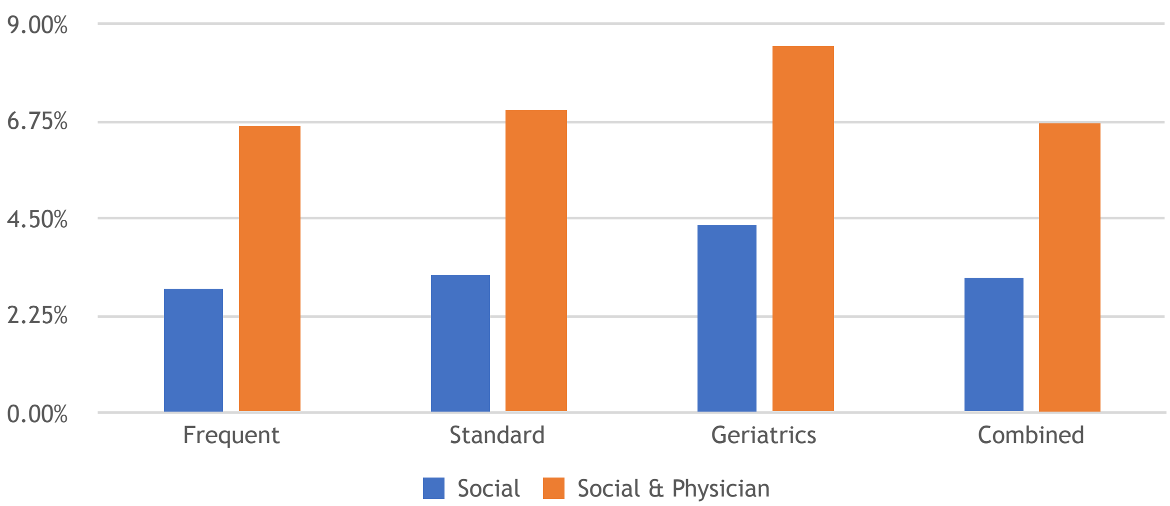
Table 3: Results of Readmissions Prediction Model

Model	Cohort[1]	AUROC	Sensitivity	Specificity	PPV	NPV
LACE Baseline	Frequent	0.8288	0.7021	0.7840	0.7466	0.7438
	Standard	0.8302	0.7341	0.7606	0.6649	0.8156
	Geriatrics	0.8254	0.7479	0.7328	0.6713	0.7994
	Combined	0.8397	0.7303	0.7757	0.6696	0.8221
LACE Baseline + Social[2]	Frequent	0.8573	0.7598	0.7573	0.7394	0.7767
	Standard	0.8621	0.7661	0.7796	0.6922	0.8375
	Geriatrics	0.8686	0.7749	0.7832	0.7228	0.8267
	Combined	0.8707	0.7763	0.7825	0.6896	0.8490
LACE Baseline + Physician[3] + Social	Frequent	0.8952	0.8232	0.8136	0.8001	0.8354
	Standard	0.9001	0.8224	0.8235	0.7509	0.8776
	Geriatrics	0.9100	0.8318	0.8331	0.7843	0.8716
	Combined	0.9069	0.8254	0.8318	0.7534	0.8845

Note: [1] Cohort represents the sample on which the model is tested on. "Frequent" represent the hold-out sample of patients identified by the hospital as frequent readmission patients. "Standard" represents the hold-out sample of a typical hospital patient. "Geriatrics" represents the sub-set of patients in the "Standard" sample who are 65 years or older. "Combined" represents the combination of both "Frequent" and "Standard" hold-out samples. [2]"Social" represents the text-mined notes that medical social workers and case managers provided. [3] "Physician" represents the text-mined notes provided by physicians.

Comparing Across Different Patients Profiles

The addition of textual information improves the AUROC of the readmission model. This improvement is particularly more significant for geriatric patients than other cohorts of patients (Figure 2). For geriatric patients, notes from the medical social workers and case managers improve the AUROC by 4.32%. Combining these notes with physician notes further improves the AUROC by 8.46% compared to the baseline LACE readmission model.



Note: "Social" represents the readmission model with clinical notes from the medical social worker and case manager. "Physician" presents the readmission model with clinical notes from the medical social workers and case managers. The models are tested on "frequent", "standard", "geriatrics," and "combined" test samples.

Figure 2. Improvements of AUROC over baseline LACE Model for Different Test Cohorts

DISCUSSION

The AUROC of our readmission risk model is higher than the typical accuracy of readmission predictive models, ranging from 0.66 to 0.83, as reported in an earlier review of 30 studies [34]. The results also suggest that the readmission predictive algorithm's performance for all four cohorts – frequent admitters, standard, geriatrics, and the combination of both – are relatively similar. Thus, this model can be applied to geriatric patients – the typical pool of patients who require additional management for readmission risks. Further, we observe that when we take into account the psychosocial information captured by non-physicians, such as medical social workers and case managers, by adding social topics, prediction accuracy improved by 0.0285 to 0.0432. When we add in the physicians' additional textual clinical notes, the AUROC further increases by between 0.0362 and 0.0414 in different cohorts.

Overall, the results show that with the addition of text-mined clinical notes from physicians and other clinicians, the AUROC of readmission prediction improves by 0.0664 to 0.0842, suggesting the added benefits of extracting psychosocial information from the textual clinical notes in predicting readmission risk.

This research showed that clinicians could leverage natural language processing to gain more information from the EMR system beyond traditional structured data commonly used to predict readmission risk. Specifically, this study establishes poof-of-concept for the use of text-mining techniques with EMR unstructured free text to identify psychosocial predictors of hospital readmission, particularly among geriatric patients. In doing so, our findings support the viability of the psychosocial approach in potentially reducing readmission rates. Thus our study represents a T2 translational stage (to patients) of research, paving the way towards T3 translational stage (to practice). In terms of development along the translational pathway, the next phase will focus on proof-of-value of embedding text-mining techniques in prediction models used to identify the risk of early readmission among hospitalized patients. The purpose of this phase is to conduct comprehensive geriatric assessment for those high risk patients with the goal of offering tailored care management. By managing their specific physical and psychosocial needs, we should observe improvement in their quality of care and a reduction in unnecessary health care utilization. In this way, precious heathcare resources can be optimally allocated to those patients who may benefit from it most. This strategy is particularly relevant for older hospitalized patients, who are more likely to have unmet psychosocial needs and among whom our augmented risk prediction model performs best. To achieve proof-of-value, future research could use quasi-experimental designs to compare the feasibility and effectiveness of a product that combines text-mined psychosocial factors with state-of-the-art prediction model with a product that only has a prediction model.

Beyond the application of text-mining techniques to the prediction of hospital readmission, this study also presents a broader and extended possibility of using the same technical approach developed for the EMR to identify a set of underdiagnosed clinical conditions in older adults that have an important influence on their health and healthcare utilization outcomes.

CONCLUSION

Psychosocial profiles of patients can be curated and quantified from text-mining clinical notes, and these profiles can be successfully applied to AI models to predict readmission risks. The use of text-mining improves the accuracy of predicting readmission, and this improved predictive accuracy is higher for geriatric patients than other patient cohorts.

References

1. Felix, H.C., et al., *Why do patients keep coming back? Results of a Readmitted Patient Survey.* Soc Work Health Care, 2015. **54**(1): p. 1-15.
2. Lim, E., et al., *Using hospital readmission rates to track the quality of care in public hospitals in Singapore.* BMC Health Services Research, 2011. **11(Suppl 1)**: p. A16.
3. Jencks, S.F., M.V. Williams, and E.A. Coleman, *Rehospitalizations among patients in the Medicare fee-for-service program.* N Engl J Med, 2009. **360**(14): p. 1418-28.
4. Pedersen, M.K., G. Meyer, and L. Uhrenfeldt, *Risk factors for acute care hospital readmission in older persons in Western countries: a systematic review.* JBI Database System Rev Implement Rep, 2017. **15**(2): p. 454-485.
5. Garcia-Perez, L., et al., *Risk factors for hospital readmissions in elderly patients: a systematic review.* Qjm, 2011. **104**(8): p. 639-51.
6. Kansagara, D., et al., *Risk Prediction Models for Hospital Readmission: A Systematic Review.* Jama, 2011. **306**(15): p. 1688-98.
7. Futoma, J., J. Morris, and J. Lucas, *A comparison of models for predicting early hospital readmissions.* J Biomed Inform, 2015. **56**: p. 229-38.
8. Zhou, H., et al., *Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review.* BMJ Open, 2016. **6**(6).
9. Jamei, M., et al., *Predicting all-cause risk of 30-day hospital readmission using artificial neural networks.* PLoS One, 2017. **12**(7): p. e0181173.
10. Low, L.L., et al., *Predicting 30-Day Readmissions in an Asian Population: Building a Predictive Model by Incorporating Markers of Hospitalization Severity.* PLoS One, 2016. **11**(12): p. e0167413.
11. Rosenberger, P.H., P. Jokl, and J. Ickovics, *Psychosocial factors and surgical outcomes: an evidence-based literature review.* JAAOS-Journal of the American Academy of Orthopaedic Surgeons, 2006. **14**(7): p. 397-405.
12. Bonde, J.P.E., *Psychosocial factors at work and risk of depression: a systematic review of the epidemiological evidence.* Occupational and environmental medicine, 2008. **65**(7): p. 438-445.
13. Kimmel, P.L., *Psychosocial factors in dialysis patients.* Kidney international, 2001. **59**(4): p. 1599-1613.
14. Kronborg, H. and M. Vaeth, *The influence of psychosocial factors on the duration of breastfeeding.* Scandinavian journal of public health, 2004. **32**(3): p. 210-216.
15. Singh-Manoux, A., *Psychosocial factors and public health.* Journal of Epidemiology & Community Health, 2003. **57**(8): p. 553-556.
16. Strike, P.C. and A. Steptoe, *Psychosocial factors in the development of coronary artery disease.* Progress in cardiovascular diseases, 2004. **46**(4): p. 337-347.
17. Zaza, C. and N. Baine, *Cancer pain and psychosocial factors: a critical review of the literature.* Journal of pain and symptom management, 2002. **24**(5): p. 526-542.
18. Lutgendorf, S.K. and E.S. Costanzo, *Psychoneuroimmunology and health psychology: An integrative model.* Brain, Behavior, and Immunity, 2003. **17**(4): p. 225-232.
19. Bradley, E.H., et al., *Expanding the Andersen model: The role of psychosocial factors in long−term care use.* Health services research, 2002. **37**(5): p. 1221-1242.
20. Bigos, S.J., et al., *A prospective study of work perceptions and psychosocial factors affecting the report of back injury.* Spine, 1991. **16**(1): p. 1-6.

21. Paarlberg, K.M., et al., *Psychosocial factors and pregnancy outcome: a review with emphasis on methodological issues.* Journal of psychosomatic research, 1995. **39**(5): p. 563-595.
22. Welin, C., G. Lappas, and L. Wilhelmsen, *Independent importance of psychosocial factors for prognosis after myocardial infarction.* Journal of internal medicine, 2000. **247**(6): p. 629-639.
23. Kartha, A., et al., *Depression is a risk factor for rehospitalization in medical inpatients.* Prim Care Companion J Clin Psychiatry, 2007. **9**(4): p. 256-62.
24. Coventry, P.A., I. Gemmell, and C.J. Todd, *Psychosocial risk factors for hospital readmission in COPD patients on early discharge services: a cohort study.* BMC Pulm Med, 2011. **11**: p. 49.
25. Flythe, J.E., et al., *Psychosocial Factors and 30-Day Hospital Readmission among Individuals Receiving Maintenance Dialysis: A Prospective Study.* Am J Nephrol, 2017. **45**(5): p. 400-408.
26. Retrum, J.H., et al., *Patient-identified factors related to heart failure readmissions.* Circ Cardiovasc Qual Outcomes, 2013. **6**(2): p. 171-7.
27. van Walraven, C., et al., *Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community.* Cmaj, 2010. **182**(6): p. 551-557.
28. Andersen, R.M., *Revisiting the behavioral model and access to medical care: does it matter?* J Health Soc Behav, 1995. **36**(1): p. 1-10.
29. Adler, N.E. and J. Stewart, *Health disparities across the lifespan: meaning, methods, and mechanisms.* Ann N Y Acad Sci, 2010. **1186**: p. 5-23.
30. Medicine, I.o., *Capturing social and behavioral domains in electronic health records: Phase 1*. 2014: Washington, DC.
31. Ohno-Machado, L., *Realizing the full potential of electronic health records: the role of natural language processing.* J Am Med Inform Assoc, 2011. **18**(5): p. 539.
32. Carnielli, C.M., et al., *Combining discovery and targeted proteomics reveals a prognostic signature in oral cancer.* Nature Communications, 2018. **9**(1): p. 3598.
33. Xia, B., et al., *Machine learning uncovers cell identity regulator by histone code.* Nature Communications, 2020. **11**(1): p. 2696.
34. Kansagara, D., et al., *Risk prediction models for hospital readmission: a systematic review.* Jama, 2011. **306**(15): p. 1688-1698.