Clinical notes: An untapped opportunity for improving risk prediction for hospitalization and emergency department visit during home health care

Jiyoun Song[a,b,*], **Mollie Hobensack**[a], **Kathryn H. Bowles**[b,c], **Margaret V. McDonald**[b], **Kenrick Cato**[a,d], **Sarah Collins Rossetti**[a,e], **Sena Chae**[f], **Erin Kennedy**[c], **Yolanda Barrón**[b], **Sridevi Sridharan**[b], **Maxim Topaz**[a,b,g]

[a]Columbia University School of Nursing, New York City, NY, USA

[b]Center for Home Care Policy & Research, Visiting Nurse Service of New York, New York City, NY, USA

[c]University of Pennsylvania School of Nursing, Department of Biobehavioral Health Sciences, Philadelphia, PA, USA

[d]Emergency Medicine, Columbia University Irving Medical Center, New York, NY, USA

[e]Columbia University, Department of Biomedical Informatics, New York City, NY, USA

[f]College of Nursing, University of Iowa, Iowa City, IA, USA

[g]Data Science Institute, Columbia University, New York City, NY, USA

Abstract

Background/Objective: Between 10 and 25% patients are hospitalized or visit emergency department (ED) during home healthcare (HHC). Given that up to 40% of these negative clinical outcomes are preventable, early and accurate prediction of hospitalization risk can be one strategy to prevent them. In recent years, machine learning-based predictive modeling has become widely used for building risk models. This study aimed to compare the predictive performance of four risk models built with various data sources for hospitalization and ED visits in HHC.

Methods: Four risk models were built using different variables from two data sources: structured data (i.e., Outcome and Assessment Information Set (OASIS) and other assessment items from the electronic health record (EHR)) and unstructured narrative-free text clinical notes for patients who received HHC services from the largest non-profit HHC organization in New York between 2015 and 2017. Then, five machine learning algorithms (logistic regression, Random Forest, Bayesian

[*]Corresponding author at: Columbia University School of Nursing, 560 West 168th Street, New York, NY 10032, USA. js4753@cumc.columbia.edu (J. Song).

network, support vector machine (SVM), and Naïve Bayes) were used on each risk model. Risk model performance was evaluated using the F-score and Precision-Recall Curve (PRC) area metrics.

Results: During the study period, 8373/86,823 (9.6%) HHC episodes resulted in hospitalization or ED visits. Among five machine learning algorithms on each model, the SVM showed the highest F-score (0.82), while the Random Forest showed the highest PRC area (0.864). Adding information extracted from clinical notes significantly improved the risk prediction ability by up to 16.6% in F-score and 17.8% in PRC.

Conclusion: All models showed relatively good hospitalization or ED visit risk predictive performance in HHC. Information from clinical notes integrated with the structured data improved the ability to identify patients at risk for these emergent care events.

Keywords

Home health care; Predictive modeling; Natural language processing; Risk assessment; Clinical deterioration; Nursing informatics

1. Introduction

Home health care (HHC) is one of the fastest-growing healthcare sectors in the United States.[1] The number of adults with complex healthcare needs is rising, and they rely heavily on HHC after being discharged from an acute care setting.[2,3] Through home visits, HHC clinicians (e.g., registered nurses, physical therapists, social workers) deliver comprehensive care including skilled nursing, rehabilitation, case management, and social work services. Approximately 3.4 million adults currently receive HHC across the United States,[1] and the demand for HHC services is expected to increase over the next few years. [4,5]

The increase in HHC has prompted discussions of patient safety.[1] During the last decade, several national and local quality improvement initiatives have focused on preventing negative outcomes in HHC, such as hospitalizations and emergency department (ED) visits. Despite these efforts, up to one in five patients still have a hospitalization or ED visits during HHC services;[1] these numbers have not improved over the last several years.[6] Recent estimates show that up to 30% of hospitalizations and ED visits are due to preventable causes,[7] ongoing efforts to early identify patient risk which can assist clinicians to make better decisions will be able to help prevent these negative outcomes.[8]

To reduce avoidable negative outcomes, patients at risk must be accurately identified so that they are able to receive timely risk mitigation interventions. A predictive analytical approach can assist with the early identification of risk. In the recent years, machine learning-based risk prediction models have become widely used for these goals. An emerging body of evidence indicates that early patient risk detection and clinician notification can reduce risk for death or rehospitalization in the hospital setting.[9] However, fewer clinical risk prediction models have been developed in the HHC setting. In one study, Tao and colleagues examined items of routine clinical assessment data (i.e., Outcome and Assessment Information Set (OASIS)) to identify contributing factors that can be used

to predict rehospitalization in HHC.[10] Furthermore, Shang and colleagues developed predictive models that use routine clinical assessment data (i.e., OASIS) to predict a patient's risk of infection-related hospitalization or ED visit.[11] Routine clinical assessment data (such as OASIS) was also used by Lo and colleagues to develop a machine learning-based predictive models to predict fall risk. [12] These studies achieved only modest predictive performance.

To improve risk prediction in HHC, it is possible to use an additional important and understudied resource- clinical narrative notes. The utilization of clinical notes in predicting clinical outcomes is becoming more widespread because narrative documentation often includes rich information on patients' medical and socio-behavioral risk indicators. [13] For example, one study found that documented nurse concerns and documentation patterns predict mortality and cardiac arrest for hospitalized patients.[14] In the HHC setting, several studies demonstrated that nurses' language in free-text clinical notes were associated with negative outcomes such as hospitalization or ED visit.[15–17] These results indicate the importance of utilizing clinical notes as sources of risk factors that are potentially associated with negative outcomes. However, previous studies in HHC only focused on specific risk factors (e.g., symptoms[18,19]), and did not combine a broader range of insights from standardized datasets and clinical notes to improve risk prediction. Given that HHC is a community-based program, personal and environmental-related information should also be explored to improve the identification of hospitalization or ED visit risk in HHC.

To address these knowledge gaps, this study aimed to compare and determine optimal risk predictive performance of four risk models and varied data sources for predicting hospitalizations and ED visits in HHC. Our hypothesis was that combining standardized data with information extracted from clinical notes will improve the performance of risk prediction models.

2. Methods

This study was a retrospective cohort study using information obtained from two data sources including structured data (i.e., OASIS and other assessment items from the electronic health record (EHR)) and unstructured data (i.e., clinical notes). Fig. 1 provides a general overview of the study methods. The study was approved by the Institutional Review Boards of the participating institutions.

2.1. Study dataset and population

This study included patients who received HHC services between 1/1/2015 and 12/31/2017 from the largest non-for-profit HHC organization in the Northeastern United States. A HHC "episode" is defined as any service a patient received between HHC admission and discharge. Patients could have more than one HHC episode during the study period. This study included 86,866 HHC episodes for 66,317 unique patients.

(1)   Structured datasets: OASIS and EHR—OASIS is a standardized assessment tool for HHC. Mandated by the Center for Medicare and Medicaid Services, OASIS assessments are performed upon admission and the end of a HHC episode. It comprehensively

assessed over 100 patient characteristics in the domains of socio-demographics, physiologic conditions, comorbidities, medication/equipment management, neuro-cognitive/behavioral status, functional status (Activities of Daily Living (ADLs)/Instrumental Activities of Daily Living (IADLs)) and health service utilization during the HHC episode.[20,21] This study used OASIS-C1 released in 2015 and OASIS-C2 released in 2017, which had a crossover table of corresponding data elements.

The EHR dataset included additional factors beyond OASIS such as socio-economic factors (e.g., insurance, residence county), medical conditions, length of HHC episodes, and medications.

(2) Unstructured dataset: Clinical notes—Approximately 2.3 million HHC clinical notes were extracted for this patient cohort. Most of the notes were documented by nurses, but the dataset also included notes documented by physical/occupational therapists and social workers. These clinical notes included (1) visit notes that include documentation of the patient's status and the care provided during the HHC visit (total n = 1,029,535), and (2) care coordination notes include documentation of communication between healthcare clinicians and other administrative care-related activities (total n = 1,292,442).

2.2. Study outcome

Indications of hospitalizations or ED visits were extracted from OASIS. Specifically, this information was extracted from OASIS item M0100: "reason for assessment (i.e., transfer to an inpatient facility – patient not discharged, or transfer to an inpatient facility – patient discharged) and M2300 "emergent care". All analyses were conducted at the HHC episode level.

2.3. Potential risk factors extracted from clinical notes

Two previously developed and validated natural language processing (NLP) approaches were used to identify and extract potential risk information from HHC clinical notes. These methods briefly summarized below were previously developed by our team and are described in detail elsewhere.[22]

(1) NLP approach #1: Identifying "concerning" clinical notes—The first approach utilized machine learning-based NLP methods – specifically, Convolutional Neural Networks (CNN) – to classify each clinical note as either "*concerning*" or "*not concerning*". A "*Concerning*" note was defined as a note that contains one or more risk factors associated with a HHC patient's risk of hospitalizations or ED visits. Previously, our team assembled a dataset that included 4000 HHC clinical notes. These notes were annotated by HHC clinical experts and each note was labeled as either "*concerning*" (~20% of notes) or "*not concerning*" (~80% of notes). We developed several machine learning algorithms and tested classification accuracy among them; the results indicated that CNN achieved better performance for the binary classification task (F-score = 0.66). We applied this previously developed CNN model to classify all clinical notes included in this study as either "*concerning*" or "*not concerning*". Appendix 1 provided more details about the NLP approach #1. Next, we calculated the proportion of "*concerning*" notes out of all

clinical notes generated during each HHC episode (e.g., 33% "*concerning*" notes per HHC episode). This aggregated variable allowed us to overcome issues associated with simply counting "*concerning*" notes, since patients with more clinical notes would likely have more "*concerning*" notes.

(2) NLP approach #2: Identifying hospitalization and ED visit risk factors
—The second approach utilized rule-based NLP system, based on the Omaha System - a standardized nursing terminology. We chose the Omaha System because it fits the domain of HHC, as shown by previous research and provided a comprehensive dataset of risk factors to consider as study variables.[23] The Omaha System covers environmental, psychosocial, physiological and health related behavior domains.[24] In a previous study, [22] we assembled a group of HHC experts to identify factors associated with HHC patients risk for hospitalization and ED visit. The detailed methods to identify hospitalization and ED visit risk factors using the Omaha System were described in Appendix 2. Specifically, out of the total 42 Omaha System problems, the experts selected and agreed on a subset of 31 Omaha System problems (e.g., "Circulation", "Bowel function", "Abuse") that were identified as risk factors for hospitalizations or ED visits in HHC (see Appendix 3 for full list of risk factors). Then, a rule-based NLP system was developed and tested to automatically identify these risk factors in HHC clinical notes. The NLP system showed good performance in identifying the risk factors with an average F-score = 0.84. For this study, this NLP system was applied, and an indicator was generated for whether the 31 risk factors (Omaha System problems) were documented or not during the HHC episode (i.e., yes/no for each risk factor). In addition, a total number of risk factors were counted per HHC episode.

2.4. Dataset preparation for analysis

The initial full dataset had 522 variables generated from the structured data (e.g., socio-demographic, physiologic conditions, functional status, etc.) and clinical notes. Initially, redundant variables between OASIS and EHR and variables with more than 20% missing data were removed from the analysis. The remaining missing values (i.e., less than 2% of the total dataset) were replaced with the mean in continuous variables and the most frequent value in categorical variables.[25] Then, we used univariate analysis (i.e., student t-test or Fisher exact test) to select variables that were statistically significant to include in further analysis ($p < 0.05$). Through the univariate analysis, seven risk factors (Omaha System problems) obtained from clinical notes by rule-based NLP were excluded: "Community resources," "Hearing," "Neighborhood workplace safety," "Residency," "Sanitation," "Sexuality" and "Sleep and rest patterns." ($p < 0.05$). In addition, redundant variables that showed strong correlations with each other were excluded (Pearson correlation coefficient above 0.5 or below −0.5) to avoid linear dependency issues. As a result, the following potential risk factors were included for the further analysis: 74 variables from the structured dataset; one variable (the proportion of "*concerning notes*") obtained from clinical notes by machine learning-based NLP (NLP #1); and 25 risk factors, including 24 risk factors (Omaha System problems) and one total number of risk factors that were obtained from clinical notes by rule-based NLP (NLP #2).

Lastly, least absolute shrinkage and selection operator (LASSO) was used for each risk model to select the best subset of variables.[26] The final selected potential risk factors for building risk models were as follows (see Appendix 4): (1) ***Risk Model 1*** utilized the structured datasets (i.e., OASIS and EHR) and included 58 potential risk factors; (2) ***Risk Model 2*** utilized both the structured dataset and the clinical notes processed by machine learning-based NLP approach #1(described above) and included 59 potential risk factors. They were 58 factors from structured datasets and 1 factor from clinical notes; (3) ***Risk Model 3*** that utilized both the structured dataset and the clinical notes processed by rule-based NLP approach #2 and included 74 potential risk factors. They were 56 factors from structured datasets and 18 factors from clinical notes; (4) ***Risk Model 4*** utilized both the structured dataset and the clinical notes processed by machine learning-based NLP approach #1 and rule-based NLP approaches #2, and included 75 potential risk factors. They were 57 factors from structured datasets, 1 factor from NLP approach #1 and 17 factors from NLP approach #2. The dataset preparation was implemented using R software version 4.1.0 (Foundation of Statistical Computing, Vienna).

2.5. Building predictive models

The dataset was randomly split into two subsets—90% for training and 10% for testing. Given the low prevalence of the study outcome (i.e., hospitalization or ED visits), Synthetic Minority Over-sampling Technique (SMOTE) was used to overcome the limitation of data imbalance. SMOTE is an oversampling technique that randomly increases minority class examples by replicating them to balance class distribution.[27,28] To minimize bias and variance in the risk model-building process and to avoid overfitting,[29] a 10-fold-cross-validation was performed. Then, five machine learning algorithms that are frequently used in classification tasks (logistic regression, Random Forest, Bayesian network, support vector machine (SVM), and Naïve Bayes) were used on each four risk models.[30,31] Risk models were conducted at the HHC episode level using an open-source tool WEKA (version 3.8.5). Hyperparameters were optimized for each algorithm on four risk models.

(1) Logistic regression—A logistic regression algorithm is a widely used method to quantify the relationship between a binary outcome and independent variables (i. e., predictors).[32] Multivariable logistic regression was used in this study. The relationship was presented with an adjusted odds ratio at a significance level < 0.05.

(2) Random Forest—A Random Forest is an ensemble classifier that generates decision trees.[33] Since the Random Forest is a modification of bagging that builds and averages many trees, it can reduce the variance and generate a robust prediction model.

(3) Bayesian network—Bayesian networks are probabilistic models that support both direct probability and inverse probability. It enables interrelated information about causalities to be integrated and utilized in probabilistic inference. [34]

(4) Support vector machine (SVM)—SVM algorithm is the function to find the hyperplane which classifies the data. SVM maps each data item into an n-dimensional space where n is the number of features, then the hyperplane separates the data items

into two classes while maximizing the marginal distance for both classes and minimizing classification errors.[35]

(5)   Naïve Bayes—The Naïve Bayes is a classification method based on Bayes' theorem, which calculates the posterior probability of an event based on the prior probabilities of conditions related to that event. It assumes that the predictors for the events are independent, but the predictors could have interdependence among themselves in real data.[36]

2.6.   Model evaluation

To evaluate the model's risk predictive ability, the following estimates were calculated: (1) **sensitivity** (i.e., **precision**) referring to the proportion of episodes when the hospitalization or ED visit was correctly identified; (2) **positive predictive value** (i.e., recall) referring to the probability that the episodes identified as a positive case truly admitted to hospital or ED, (3) **F-score** referring the weighted harmonic mean of the precision and recall; and (4) **precision-recall curve (PRC)** referring the area under the curve of the precision and recall.

3.   Results

During the study period, 8373/86,823 (9.6%) of HHC episodes had hospitalization or ED visits. Specifically, 7666/8373 (91.6%) were hospitalized and 6505/8373 (77.7%) had ED visits.

3.1.   Cohort demographics and clinical characteristics

The average patient's age was 78.8 years and 64% were female. The most common diagnoses were hypertension, diabetes, and arthritis (65%, 30%, and 24%, respectively). Approximately 80% of patients took 5 or more medications, and it was more common in the patients with hospitalization or ED visits than those without (84% vs. 79%). About 25% of patients had multiple hospitalizations in the past 6 months before HHC, and it was predominant in the patients with hospitalization or ED visits than those without (40% vs. 23%). See Table 1 for further detail.

3.2.   Information extracted from clinical notes through NLP approaches

In each HHC episode, an average of 28 clinical nursing notes were documented. Patients with a hospitalization or ED visit had more notes that those without (37.8 vs. 25.4 notes). NLP approach #1: Over 90% of the episodes included at least one 'concerning' note based on the machine learning-based NLP approach #1. Approximately 25% of the notes were identified as 'concerning notes' per HHC episode and the proportion of 'concerning notes' was higher among patients with a hospitalization or ED visit than those without (31% vs. 25%). NLP approach #2: An average of 3.3 Omaha System risk factors were documented in the clinical notes per HHC episode. More risk factors were documented for patients with a hospitalization or ED visit than those without (4.5 vs 3.2 problems). The most frequent documented risk factors were "Pain", followed by "Neuro-musculo-skeletal function", and "Circulation" issues (48%, 46% and 35%, respectively). These risk factors were more frequently documented among patients with a hospitalization or ED visit than those without (56% vs. 48%, 50% vs. 46%, and 46% vs. 34%, respectively). See Table 1 for details.

3.3.   Risk prediction models performance evaluation

Table 2 provides a comparison of the risk prediction ability among the risk models. Overall, most models showed an F-score over 0.7, which suggests relatively good risk prediction ability with low false positives and low false negatives. Among the risk models, ***Risk Model 4*** utilizing both the structured dataset and clinical notes processed using both NLP approaches had the highest risk prediction ability with the F-score = 79.6% in SVM algorithms, and the highest PRC area of 86.4% in Random Forest algorithms. Compared with ***Risk Model 1*** that utilized only structured data, the predictive ability of the model utilizing both structured data and clinical notes improved markedly, with largest improvement in Bayes Network algorithms. Compared to ***Risk Model 1***, the F-score increased by 7% in ***Risk Model 2***, 12.4% in ***Risk Model 3***, and 16.6% in ***Risk Model 4***. The PRC area also increased by 6.6% in ***Risk Model 2***, 12% in ***Risk Model 3***, and 17.8% in ***Risk Model 4*** (all p-values < 0.05). Except for Naive Bayes algorithms, the F-score and PRC area increased incrementally between ***Risk Model 1, 2, 3, and 4***. In addition, compared with ***Risk Model 2*** (utilizing NLP approach #1), the predictive ability in ***Risk Model 3*** (utilizing NLP approach #2) improved by 5.2% of F-score in SVM, and 5% of PRC area in Bayes Network.

4.   Discussion

This study explored the development of risk models that can help to identify HHC patients at risk for hospitalizations or ED visits. Early risk identification during HHC can support clinicians' clinical management decisions, thus preventing negative outcomes such as hospitalization or ED visits.[37–40] In previous studies,[22] our team developed and validated two NLP approaches to identify factors that are hypothetically associated with an increased risk for hospitalization or ED visits in HHC. In this study, we compared prediction models with and without the NLP generated variables and confirmed that both NLP approaches generated valid indicators associated with the increased risk for hospitalization or ED visits.

This study extends the use of the Omaha System in NLP methods as previous studies that used the Omaha System to estimate patients' risk of hospitalization did not use NLP of clinical notes to extract potential risk factors.[41,42] In this study, NLP was applied to extract HHC domain-related unique risk factors from clinical notes. For example, those NLP-extracted risk factors included 'health care supervision' problems, indicating that patients fail to obtain routine/preventive care or have inconsistent source of health care, and 'medication regimen' problem indicating that patients are not following the recommended dosage/schedule or are unable to take medications without help. These HHC-specific risk factors go beyond risk factors used in previous studies; for example, the Omaha System problem of 'health care supervision' was not used in previous NLP or predictive modeling research.[43] Our results suggest that HHC-specific risk factors which take into account the unique environment of community health can be identified in clinical notes, and they should be included comprehensively in risk indicators for hospitalization or emergency department visits. In future research, automated and data-driven methods using advanced NLP techniques such as topic modeling[44] should be warranted to extract risk factors.

Most risk models showed relatively good risk predictive performance (average F-score 0.75 [range 0.68–0.82]; average PRC area 0.77 [range 0.68–0.86]). While adding information extracted from clinical notes with NLP, risk prediction ability improved significantly by up to 17.8% compared to the baseline model that used data from standardized assessment and EHR only. These results confirm the hypothesis that risk predictions improved when insights from NLP were incorporated into the risk models. This highlights the importance of narrative clinical notes as informative data sources.

Compared to the ***Risk Model 2*** utilizing the structured dataset and the clinical notes processed by CNN machine learning approaches (NLP approach #1), the F-score of the ***Risk Model 3*** utilizing the structured dataset and the Omaha System problems in clinical notes (NLP approach #2) improved by up to 5%. This result is notable, and it might suggest that utilization of comprehensive and extended risk information (such as the Omaha System problems) is superior to utilizing summary indexes (such as the proportion of "*concerning*" notes during the HHC episode). These results require further investigation and validation in other settings with additional data.

Each machine learning algorithm has strengths and weaknesses. In clinical risk models, logistic regression is often applied because clinicians can easily measure and interpret how predictors and outcomes are related. However, this model is not flexible enough to handle more complex non-liner interactions within the data.[32] Random Forest is an example of ensemble methods which is a combination of multiple trees; it often showed good performance for prediction. However, Random Forest is harder to interpret and overfitting issues can easily occur.[33] On the other hand, Bayesian networks can handle complex prediction problems with efficiency, but they do not guarantee cyclic relationships. [34] When the data are complex and multi-dimensional, SVM remains robust to overfitting, but it is difficult to fine-tune the model because it is affected by the kernel parameters.[35] Naive Bayes is a relatively simple algorithm based around conditional probability; it often shows poor performance due to its simplicity,[36] which is supported by our results. To overcome the weaknesses of the machine learning algorithms used in this study and to improve the risk model performance, other ensemble algorithms (e.g., XGBoost) or deep learning models might be needed.

To generate generalizable risk models, we used advanced computational techniques such as stratified random splitting, SMOTE over-sampling, and cross-validation to improve the risk prediction ability on our imbalanced dataset (which refers to the large difference in proportion between case and non-case). We evaluated the risk models with multiple metrics. In this study, two algorithms showed the highest performance: Random Forest when evaluating the performance with PRC area; SVM when evaluating the performance with F-measure. However, because PRC area (i.e., a rank metric) is a more suitable metric than F-score (i.e., threshold metrics) for evaluating predictive performance in the imbalanced data,[45,46] our results indicate that Random Forest model achieved best performance.

4.1. Future clinical implications

Our results indicate the feasibility of developing an HHC-specific risk prediction model. Using such models, early warning systems can be developed to proactively identify HHC

patients at risk for hospitalization or ED visit. These systems can be integrated into HHC clinical workflows to alert HHC nurses about patients at risk, who can subsequently intervene to reduce risks and potentially improve clinical outcomes. In hospital settings, early warning systems were tested and found to improve clinical outcomes.[47] In HHC, however, little is known about such early warning systems and further research is needed to develop usable and clinician-friendly interfaces and to integrate these systems into clinical workflows. Finally, clinical trials are needed to evaluate the effectiveness of early warning systems in HHC settings in terms of improving patient outcomes, such as reducing hospitalizations and ED visits. Further experiments can be conducted to investigate the prediction of the time to the event. Future studies should take into account additional information, such as giving more weight to notes written near the event, timing, or days after the event. Based on longitudinal data, future studies could also examine how early the risk of patients' outcomes can be estimated.

4.2. Limitations

This study has several notable limitations. As this research was a retrospective study, the associations between risk factors and the outcome can be identified but causality cannot be inferred. Furthermore, missing data and other unmeasured confounders that were not available in the dataset might have affected the risk models. In addition, the dataset was derived from a single institution, therefore, it may have an organization-specific template for documentation, or clinical language and jargon patterns, which limits the generalizability of the risk models. The data were collected between 2015 and 2017, the documentation trends in more recent data may not be appropriately reflected. Further limitations are attributable to the OASIS dataset that only captures a snapshot on admission to HHC. Therefore, the fluctuations of conditions during the HHC episode may not have been reflected. Further analysis is needed in order to identify how early the risk can be predicted when the analysis is conducted at the HHC visit level. Lastly, even though our results indicate that the Random Forest model achieved the best performance, the best machine learning algorithm may differ depending on the dataset or by applying other machine learning algorithms.

5. Conclusions

This study demonstrated the contribution of leveraging information included in clinical notes with structured assessment data to identify HHC patients at risk for hospitalization or ED visits. A risk model that utilized both the structured dataset and clinical notes showed improved risk predictive ability by up to 17.8% for hospitalization or ED visits in HHC. In addition, Random Forest showed the best predictive ability among the five machine learning algorithms. Further studies should explore use of early warning systems to prevent hospitalization or ED visits in HHC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Abbreviations:

|	
HHC	Home Health Care
ED	Emergency Department
OASIS	Outcome and Assessment Information Set
EHR	Electronic Health Record
ADL/IADL	Activities of Daily Living/Instrumental Activities of Daily Living function
NLP	Natural Language Processing
CNN	Convolutional Neural Networks
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine; PRC, Precision-Recall Curve

References

[1]. The Medicare Payment Advisory Commission. Report to the congress- Medicare payment policy: Home health care services, 2019; http://www.medpac.gov/docs/default-source/reports/mar19_medpac_entirereport_sec.pdf (accessed December 07, 2020).

[2]. Jarvis WR, Infection control and changing health-care delivery systems, Emerg. Infect. Dis. 7 (2) (2001) 170–173, 10.3201/eid0702.010202. [PubMed: 11294699]

[3]. Hardin L, Mason DJ, Bringing It Home: The Shift in Where Health Care Is Delivered, J. Am. Med. Assoc. 322 (6) (2019) 493–494, 10.1001/jama.2019.11302.

[4]. Mitzner TL, Beer JM, McBride SE, Rogers WA, Fisk AD, Older Adults' Needs for Home Health Care and the Potential for Human Factors Interventions, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 53 (1) (2009) 718–722, 10.1177/154193120905301118. [PubMed: 25530696]

[5]. Landers S, Madigan E, Leff B, Rosati RJ, McCann BA, Hornbake R, MacMillan R, Jones K, Bowles K, Dowding D, Lee T, Moorhead T, Rodriguez S, Breese E, The Future of Home Health Care: A Strategic Framework for Optimizing Value, Home Health Care Manage. Pract. 28 (4) (2016) 262–278, 10.1177/1084822316666368.

[6]. Centers for Medicare and Medicaid Services. Home Health Quality Measures, 2019; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Home-Health-Quality-Measures.html (accessed April 5th, 2021).

[7]. Solberg LI, Ohnsorg KA, Parker ED, et al. , Potentially Preventable Hospital and Emergency Department Events: Lessons from a Large Innovation Project, Perm J. 22 (2018) 17–102, 10.7812/TPP/17-102.

[8]. Zolnoori M, McDonald MV, Barrón Y, Cato K, Sockolow P, Sridharan S, Onorato N, Bowles K, Topaz M, Improving Patient Prioritization During Hospital-Homecare Transition: Protocol for a Mixed Methods Study of a Clinical Decision Support Tool Implementation, JMIR Res. Protocols 10 (1) (2021) e20184, 10.2196/20184.

[9]. Fu L-H, Schwartz J, Moy A, Knaplund C, Kang M-J, Schnock KO, Garcia JP, Jia H, Dykes PC, Cato K, Albers D, Rossetti SC, Development and validation of early warning score system: A systematic literature review, J. Biomed. Inform 105 (2020) 103410, 10.1016/j.jbi.2020.103410. [PubMed: 32278089]

[10]. Tao H, Ellenbecker CH, Is OASIS Effective in Predicting Rehospitalization for Home Health Care Elderly Patients? Home Health Care Manage. Pract 25 (6) (2013) 250–255, 10.1177/1084822313495046.

[11]. Shang J, Russell D, Dowding D, McDonald MV, Murtaugh C, Liu J, Larson EL, Sridharan S, Brickner C, A Predictive Risk Model for Infection-Related Hospitalization Among Home Healthcare Patients, J. Healthcare Quality: Off. Publ. Nat. Assoc. Healthcare Quality 42 (3) (2020) 136–147, 10.1097/JHQ.0000000000000214.

[12]. Lo Y, Lynch SF, Urbanowicz RJ, et al. , Using Machine Learning on Home Health Care Assessments to Predict Fall Risk, Stud. Health Technol. Inform 264 (2019) 684–688, 10.3233/shti190310. [PubMed: 31438011]

[13]. Song J, Woo K, Shang J, Ojo M, Topaz M, Predictive Risk Models for Wound Infection-Related Hospitalization or ED Visits in Home Health Care Using Machine-Learning Algorithms, Adv. Skin Wound Care 34 (8) (2021) 1–12, 10.1097/01.ASW.0000755928.30524.22.

[14]. Collins SA, Cato K, Albers D, Scott K, Stetson PD, Bakken S, Vawdrey DK, Relationship between nursing documentation and patients' mortality, Am. J. Crit. Care: Off. Publ., Am. Assoc. Crit.-Care Nurses 22 (4) (2013) 306–313, 10.4037/ajcc2013426.

[15]. Topaz M, Adams V, Wilson P, Woo K, Ryvicker M, Free-Text Documentation of Dementia Symptoms in Home Healthcare: A Natural Language Processing Study, 2333721420959861, Gerontol. Geriatric Med. 6 (2020), 10.1177/2333721420959861.

[16]. Topaz M, Koleck TA, Onorato N, Smaldone A, Bakken S, Nursing documentation of symptoms is associated with higher risk of emergency department visits and hospitalizations in homecare patients, Nurs. Outlook 69 (3) (2021) 435–446, 10.1016/j.outlook.2020.12.007. [PubMed: 33386145]

[17]. Topaz M, Woo K, Ryvicker M, Zolnoori M, Cato K, Home Healthcare Clinical Notes Predict Patient Hospitalization and Emergency Department Visits, Nurs. Res 69 (6) (2020) 448–454, 10.1097/nnr.0000000000000470. [PubMed: 32852359]

[18]. Woo K, Song J, Adams V, Block LJ, Currie LM, Shang J, Topaz M, Exploring prevalence of wound infections and related patient characteristics in homecare using natural language processing, Int. Wound J 19 (1) (2022) 211–221, 10.1111/iwj.13623. [PubMed: 34105873]

[19]. Topaz M, Murga L, Gaddis KM, McDonald MV, Bar-Bachar O, Goldberg Y, Bowles KH, Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches, J. Biomed. Inform 90 (2019) 103103, 10.1016/j.jbi.2019.103103. [PubMed: 30639392]

[20]. Tullai-McGuinness S, Madigan EA, Fortinsky RH, Validity testing the Outcomes and Assessment Information Set (OASIS), Home Health Care Services Quarterly 28 (1) (2009) 45–57, 10.1080/01621420802716206. [PubMed: 19266370]

[21]. Shang J, Larson E, Liu J, Stone P, Infection in home health care: Results from national Outcome and Assessment Information Set data, Am. J. Infect. Control 43 (5) (2015) 454–459, 10.1016/j.ajic.2014.12.017. [PubMed: 25681302]

[22]. Song J, Ojo M, Bowles KH, McDonald MV, Cato K, Rossetti SC, Adams V, Chae S, Hobensack M, Kennedy E, Tark A, Kang M-J, Woo K, Barrón Y, Sridharan S, Topaz M, Detecting Language Associated with Home Health Care Patient's Risk for Hospitalization and Emergency Department Visit, Nurs. Res (2022), 10.1097/NNR.0000000000000586.

[23]. Topaz M, Golfenshtein N, Bowles KH, The Omaha System: a systematic review of the recent literature, J. Am. Med. Inform. Assoc.: JAMIA 21 (1) (2014) 163–170, 10.1136/amiajnl-2012-001491. [PubMed: 23744786]

[24]. Martin KS, The Omaha System: A Key to Practice, Documentation, and Information Management, Elsevier Saunders, 2005.

[25]. Little RJA, Schluchter MD, Maximum likelihood estimation for mixed continuous and categorical data with missing values, Biometrika 72 (3) (1985) 497–512, 10.1093/biomet/72.3.497.

[26]. Li Q, Shao J, Regularizing LASSO: A Consistent Variable Selection Method, Statistica Sinica 25 (3) (2015) 975–992, 10.5705/ss.2013.001.

[27]. Blagus R, Lusa L, SMOTE for high-dimensional class-imbalanced data, BMC Bioinf 14 (1) (2013) 106, 10.1186/1471-2105-14-106.

[28]. Elreedy D, Atiya AF, A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance, Inf. Sci 505 (2019) 32–64, 10.1016/j.ins.2019.07.070.

[29]. Schaffer C, Selecting a classification method by cross-validation, Machine Learning 13 (1) (1993) 135–143, 10.1007/BF00993106.

[30]. Acker WW, Plasek JM, Blumenthal KG, Lai KH, Topaz M, Seger DL, Goss FR, Slight SP, Bates DW, Zhou L.i., Prevalence of food allergies and intolerances documented in electronic health records, J. Allergy Clin. Immunol. 140 (6) (2017) 1587–1591.e1, 10.1016/j.jaci.2017.04.006. [PubMed: 28577971]

[31]. Sevakula RK, Au-Yeung WTM, Singh JP, Heist EK, Isselbacher EM, Armoundas AA, State-of-the-Art Machine Learning Techniques Aiming to Improve Patient Outcomes Pertaining to the Cardiovascular System, J. Am. Heart Assoc. 9 (4) (2020), e013924, 10.1161/JAHA.119.013924. [PubMed: 32067584]

[32]. Hosmer DW Jr., Lemeshow S, Sturdivant RX, Applied Logistic Regression, vol. 398, John Wiley & Sons, 2013.

[33]. Breiman L, Random Forests, Machine Learning 45 (1) (2001) 5–32, 10.1023/A:1010933404324.

[34]. Friedman N, Geiger D, Goldszmidt M, Bayesian network classifiers, Machine Learning 29 (2) (1997) 131–163, 10.1023/A:1007465528199.

[35]. Joachims T, Making large-scale support vector machine learning practical, in: Advances in kernel methods: support vector learning, MIT Press, 1999, pp. 169–184.

[36]. Rish I, An empirical study of the naive Bayes classifier. Paper presented at: IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001.

[37]. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G, Big data in health care: using analytics to identify and manage high-risk and high-cost patients, Health Affairs (Project Hope) 33 (7) (2014) 1123–1131, 10.1377/hlthaff.2014.0041. [PubMed: 25006137]

[38]. Neuberger L, Silk KJ, Uncertainty and information-seeking patterns: A test of competing hypotheses in the context of health care reform, Health Communication 31 (7) (2016) 892–902, 10.1080/10410236.2015.1012633. [PubMed: 26698558]

[39]. Crockett D, Why predictive modeling in healthcare requires a data warehouse, 2014. https://www.healthcatalyst.com/predictive-modeling-healthcare-requirements/ (accessed October 19, 2020).

[40]. Watson K, Predictive Analytics in Health Care: Emerging Value and Risks, Deloitte Development LLC (2019).

[41]. Monsen KA, Radosevich DM, Kerr MJ, Fulkerson JA, Public health nurses tailor interventions for families at risk, Public Health Nursing (Boston, Mass) 28 (2) (2011) 119–128, 10.1111/j.1525-1446.2010.00911.x. [PubMed: 21732966]

[42]. Ballester N, Parikh PJ, Donlin M, May EK, Simon SR, An early warning tool for predicting at admission the discharge disposition of a hospitalized patient, Am. J. Managed Care 24 (10) (2018) e325–e331.

[43]. Ma C, Shang J, Miner S, Lennox L, Squires A, The Prevalence, Reasons, and Risk Factors for Hospital Readmissions Among Home Health Care Patients: A Systematic Review, Home Health Care Manage. Pract 30 (2) (2017) 83–92, 10.1177/1084822317741622.

[44]. Dieng AB, Ruiz FJR, Blei DM, Topic Modeling in Embedding Spaces, Trans. Assoc. Comput. Linguist 8 (2020) 439–453, 10.1162/tacl_a_00325.

[45]. Saito T, Rehmsmeier M, Brock G, The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, PLoS ONE 10 (3) (2015) e0118432, 10.1371/journal.pone.0118432. [PubMed: 25738806]

[46]. Japkowicz N, Assessment Metrics for Imbalanced Learning. In: Imbalanced Learning, 2013, pp. 187–206.

[47]. Gerry S, Bonnici T, Birks J, et al. , Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology, BMJ 369 (2020) m1501, 10.1136/bmj.m1501. [PubMed: 32434791]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
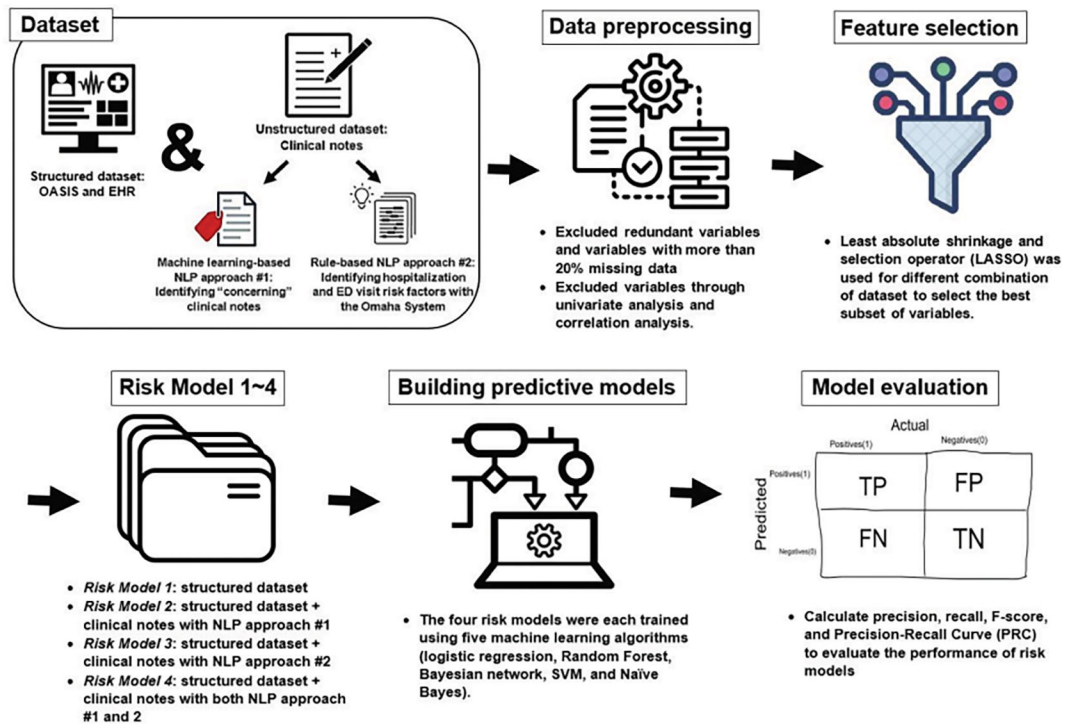
Fig. 1.
Study methods overview.

Table 1

Patient Characteristics and Information Extracted from Clinical Notes Between Patients with Hospitalization/ED Visit and those without (all P-value < 0.05).

	Patients with hospitalization /ED visits (n = 8379)	Patients without hospitalization /ED visits (n = 78,487)
1. Socio-demographic factor		
Age, mean (SD)	78.5 (12.8)	78.9 (11.7)
Gender		
Female [n, (%)] Ethnicity [n, (%)]	5093 (60.8)	50,448 (64.3)
Non-Hispanic White	4750 (56.7)	49,813 (63.5)
Non-Hispanic Black	1825 (21.8)	13,382 (17.0)
Hispanic	1341 (16.0)	10,459 (13.3)
Other	463 (5.5)	4833 (6.2)
Type of insurance [n, (%)]		
Dual eligibility	764 (9.1)	4929 (6.3)
Medicare/Medicaid FFS only	7556 (90.2)	73,013 (93.0)
Any managed care	50 (0.6)	501 (0.6)
Other (e.g., private)	9 (0.1)	44 (0.1)
Living Condition [n, (%)]		
Living with others (Congregate/Other)	5304 (63.3)	47,783 (60.9)
Living alone	3075 (36.7)	30,704 (39.1)
2. Care related factor		
Length of stay greater than 60 days [n, (%)]	2708 (32.3)	7940 (10.1)
3. Medical conditions		
Active Diagnoses [n, (%)] **		
Acute Myocardial Infarction	1735 (20.7)	14,041 (17.9)
Arthritis	1188 (14.2)	19,228 (24.5)
Cancer	336 (4.0)	1151 (1.5)
Cardiac Dysrhythmias	1533 (18.3)	11,796 (15.0)
Dementia	1240 (14.8)	10,154 (12.9)
Diabetes	3175 (37.9)	22,427 (28.6)
Heart Failure	1946 (23.2)	10,506 (13.4)
Pulmonary Disease	1570 (18.7)	11,638 (14.8)
Peripheral Vascular Disease	358 (4.3)	2459 (3.1)
Renal Disease	686 (8.2)	2567 (3.3)
Skin Ulcer	1834 (21.9)	9734 (12.4)
Stroke	797 (9.5)	6726 (8.6)
4. Risk for Hospitalization [n, (%)]		
History of falls in the past 12 months	1856 (22.2)	16,652 (21.2)
Multiple hospitalizations in the past 6 months	3327 (39.7)	18,328 (23.4)
Currently taking 5 or more medications	7007 (83.6)	13,186 (79.0)
5. Sensory Status		

	Patients with hospitalization /ED visits (n = 8379)	Patients without hospitalization /ED visits (n = 78,487)
Frequency of Pain [n, (%)]		
Patient has no pain	2199 (26.2)	18,003 (22.9)
No Interference Activity or Less than Daily	2219 (26.5)	19,776 (25.2)
Daily, Not Constant, or All the Time	3961 (47.3)	40,708 (51.9)
6. Integumentary [n, (%)]		
Having a Risk of Developing Pressure Ulcers	4187 (50.0)	31,541 (40.2)
Having at least one Unhealed Pressure Ulcer at Stage II or Higher	1116 (13.3)	5377 (6.9)
Having Stasis Wound	377 (4.5)	2016 (2.6)
Having Surgical Wound	1576 (18.8)	21,500 (27.4)
Having a Skin Lesion or Open Wound	2218 (26.5)	15,761 (20.1)
7. Respiratory Status		
Short of Breath [n, (%)]		
Never	4112 (49.1)	46,370 (59.1)
When walking more than 20 feet/climbing stairs	3602 (43.0)	29,028 (37.0)
With moderate/minimal exertion or at rest	665 (7.9)	3089 (3.9)
8. Elimination		
Urinary Tract Infection in the past 14 days [n, (%)]	816 (9.7)	5151 (6.6)
9. Neuro, Emotional, and Behavioral Status		
Cognitive Functioning [n, (%)]		
Alert/oriented or prompting	6987 (83.4)	68,587 (87.4)
Requires assistance or totally dependent	1392 (16.6)	9900 (12.6)
When Confused [n, (%)]		
Never	3638 (43.4)	40,550 (51.7)
In new or complex situations	3571 (42.6)	29,392 (37.4)
On awakening and/or at night only, or consistently	1170 (14.0)	8545 (10.9)
10. ADLs / IADLs		
ADL Needed [mean, (SD)]	8.3 (1.4)	8 (1.5)
ADL Severity [mean, (SD)]	17.4 (7.7)	15.5 (6.7)
11. Risk Factors from Clinical Notes (Identified through NLP approaches)		
Proportion of concerning notes (%) [mean, (SD)]	30.9 (13.6)	24.8 (13.4)
Omaha System Problem		
Abuse [n, (%)]	401 (4.8)	1836 (2.3)
Bowel function [n, (%)]	871 (10.4)	3675 (4.7)
Circulation [n, (%)]	3836 (45.8)	26,890 (34.3)
Cognition [n, (%)]	1931 (23.0)	12,639 (16.1)
Infectious condition [n, (%)]	3296 (39.3)	19,300 (24.6)
Consciousness [n, (%)]	616 (7.4)	2053 (2.6)
Digestion/hydration [n, (%)]	1193 (14.2)	5870 (7.5)
Genitourinary function [n, (%)]	532 (6.3)	2236 (2.8)
Health care supervision [n, (%)]	1000 (11.9)	7088 (9.0)

	Patients with hospitalization /ED visits (n = 8379)	Patients without hospitalization /ED visits (n = 78,487)
Income [n, (%)]	383 (4.6)	2577 (3.3)
Interpersonal relationship [n, (%)]	36 (0.4)	150 (0.2)
Medication regimen [n, (%)]	657 (7.8)	4069 (5.2)
Mental health [n, (%)]	3776 (45.1)	24,870 (31.7)
Neglect [n, (%)]	589 (7.0)	2946 (3.8)
Neuro-musculo-skeletal function [n, (%)]	4133 (49.4)	36,073 (46.0)
Nutrition [n, (%)]	1251 (14.9)	5986 (7.6)
Oral health [n, (%)]	14 (0.2)	63 (0.1)
Pain [n, (%)]	4680 (55.9)	37,446 (47.7)
Personal care [n, (%)]	95 (1.1)	331 (0.4)
Respiration [n, (%)]	2979 (35.6)	16,235 (20.7)
Skin [n, (%)]	2786 (33.2)	16,536 (21.1)
Social contact [n, (%)]	1403 (16.8)	14,738 (18.8)
Speech and language [n, (%)]	547 (6.5)	3283 (4.2)
Substance use [n, (%)]	100 (1.2)	459 (0.6)
Average number of problems per HHC episode [mean, (SD)]	4.5 (2.7)	3.2 (2.2)

Table 2

A comparison of the risk prediction ability among the risk models. (1) ***Risk Model 1*** utilizing the structured dataset (i.e., OASIS and EHR); (2) ***Risk Model 2*** utilizing both the structured dataset and clinical notes processed using machine learning-based NLP approaches (*concerning/not concerning note*); (3) ***Risk Model 3*** utilizing both the structured dataset and clinical notes processed with the Omaha System; (4) ***Risk Model 4*** utilizing both the structured dataset and clinical notes processed using both machine learning-based NLP approaches and with the Omaha System.

	Sensitivity (Precision)	PPV (Recall)	F-score	PRC Area
Logistic regression				
Risk Model 1	0.794	0.64	0.709	0.736
Risk Model 2	0.812	0.652	0.723	0.756
Risk Model 3	0.833	0.683	0.751	0.774
Risk Model 4	0.837	0.694	**0.759**	**0.812**
Random Forest				
Risk Model 1	0.896	0.692	0.781	0.818
Risk Model 2	0.909	0.693	0.786	0.84
Risk Model 3	0.918	0.707	0.799	0.845
Risk Model 4	0.927	0.721	**0.811**	**0.864**
Bayes Network				
Risk Model 1	0.721	0.643	0.680	0.71
Risk Model 2	0.749	0.708	0.728	0.757
Risk Model 3	0.815	0.72	0.765	0.795
Risk Model 4	0.827	0.762	**0.793**	**0.836**
SVM				
Risk Model 1	0.801	0.675	0.733	0.765
Risk Model 2	0.82	0.687	0.748	0.784
Risk Model 3	0.902	0.697	0.786	0.807
Risk Model 4	0.922	0.731	**0.815**	**0.821**
Naïve Bayes				
Risk Model 1	0.702	0.65	0.675	0.688
Risk Model 2	0.721	0.677	**0.698**	**0.701**
Risk Model 3	0.692	0.661	0.676	0.682
Risk Model 4	0.702	0.682	0.692	0.684

Note: PPV: positive predictive value; PRC: precision-recall curve; SVM: support vector machine