Published in final edited form as:

Nurs Res. 2020 ; 69(6): 448–454. doi:10.1097/NNR.0000000000000470.

Home Health Care Clinical Notes Predict Patient Hospitalization and Emergency Department Visits

Maxim Topaz¹, Kyungmi Woo¹, Miriam Ryvicker², Maryam Zolnoori¹, Kenrick Cato¹

¹Columbia University School of Nursing, New York, New York

²Center for Home Care Policy & Research, Visiting Nurse Service of New York, New York, New York

Abstract

Background: About 30% of home health care patients are hospitalized or visit an emergency department (ED) during a home health care (HHC) episode. Novel data science methods are increasingly used to improve identification of patients at risk for negative outcomes.

Objectives: To identify patients at heightened risk hospitalization or ED visits using HHC narrative data (clinical notes).

Methods: This study used a large database of HHC visit notes ($n = 727,676$) documented for 112,237 HHC episodes (89,459 unique patients) by clinicians of the largest nonprofit home health care agency in the United States. Text mining and machine learning algorithms (Naïve Bayes, decision tree, random forest) were implemented to predict patient hospitalization or ED visits using the content of clinical notes. Risk factors associated with hospitalization or ED visits were identified using a feature selection technique (gain ratio attribute evaluation).

Results: Best performing text mining method (random forest) achieved good predictive performance. Seven risk factors categories were identified, with clinical factors, coordination/communication, and service use being the most frequent categories.

Discussion: This study was the first to explore the potential contribution of HHC clinical notes to identifying patients at risk for hospitalization or an ED visit. Our results suggest that HHC visit notes are highly informative and can contribute significantly to identification of patients at risk. Further studies are needed to explore ways to improve risk prediction by adding more data elements from additional data sources.

Keywords

home health care; natural language processing; nursing informatics; risk prediction; text mining

Every year, more than 11,000 home health care (HHC) agencies across the United States provide care to more than 5 million older adults (MedPac, 2014). Currently, about one in three HHC patients are hospitalized or visit an emergency department (ED) during the 30–

Corresponding author: Maxim Topaz, Columbia University School of Nursing, 560 West 168th Street, New York, NY 10032. mt3315@columbia.edu.

The authors have no conflicts of interest to report.

60 day HHC episode (Centers for Medicare and Medicaid Services, 2019). These numbers have not improved over the last several years (Centers for Medicare and Medicaid Services, 2019; MedPac, 2014), despite several national and local efforts to improve quality of care and patient outcomes (Agency for Healthcare Research & Quality, 2016; Busby et al., 2015; Hodgson et al., 2019).

In settings other than HHC, novel, data science methods increasingly are used to improve identification of patients at risk. For example, in hospital settings, recent systematic reviews (Jayasundera et al., 2018; Linnen et al., 2019) found that early warning systems are effective in identifying patients who are likely to die or be transferred to an intensive care unit (ICU). Studies developing early warning systems often use advanced data analytics methods to identify specific patterns in electronic health record (HER) documentation that are a proxy of a clinician's concern about a patient. Some researchers demonstrate the ability to predict patient deterioration up to 48 hours before adverse outcomes happen using these methods (Collins & Vawdrey, 2012; Dykes et al., 2017; Linnen et al., 2019).

Researchers have used previously unexplored data sources to improve further identification of patients at risk. Several recent studies showed that information from narrative clinical notes, extracted via natural language processing (NLP), is a key data source about patient risk (Korach et al., 2019; Woo et al., 2019). For instance, some researchers have found that a patient's social risk factors (Navathe et al., 2018) or poor self-management status (Topaz et al., 2016) extracted automatically from inpatient clinical notes are significantly associated with rehospitalizations. Other researchers have shown that information extracted from inpatient clinical notes, including nursing notes, can be used to identify patients at risk for mortality (Lehman et al., 2012; Marafino et al., 2015; Waudby-Smith et al., 2018). These previous studies suggest that NLP can be used to develop early warning systems in the hospital setting by identifying key risk factors expressed in the narrative notes and using that information to trigger real-time interventions to mitigate risk and prevent poor outcomes.

Compared to the hospital setting, developing early warning systems in HHC is challenging due to several major differences between hospitals and HHC settings. First, care encounters in HHC (e.g., nursing visits) typically happen a few days apart, compared to minute or hourly encounters in hospitals. Second, most of the data in HHC is generated by nurses and physical or occupational therapists compared to a larger array of health professionals in the inpatient setting. Our study is the first to explore the potential for HHC narrative data to help identify patients at heightened risk for poor outcomes. Specifically, we used a large patient sample from the largest nonprofit HHC agency in the U.S. to examine whether information documented in HHC visit notes can help predict patient hospitalization or ED visits. Answering the study question is critical in order to determine the feasibility of implementing early warning systems in HHC using innovative data science methods, such as those developed for the hospital setting, to improve care for the highest-risk patients. Thus, we explored the potential for HHC narrative nursing notes to help identify patients at high risk for hospitalization or ED visits.

Methods

This was secondary data analysis using a retrospective cohort data of narrative HHC visit notes and patient hospitalization or ED visits outcome from the Outcomes Assessment and Information Set (OASIS). HHC notes and the OASIS were not collected specifically for this study. The study was approved by the institutional review board at Columbia University and the HHC agency.

Data Collection

We used HHC visit notes ($n = 1,149,586$) documented for 112,237 HHC episodes (89,459 unique patients) by clinicians of the largest nonprofit HHC agency in New York, NY between January 1, 2014 and December 31, 2014. The patient sample had a mean age of 70.8, was 60.8% female, and 37% lived alone. The sample was racially and ethnically diverse, with 43% White, 27% Black or African American, 24% Hispanic/Latino, and 6% Asian.

Visit notes were completed by HHC clinicians (e.g., nurses, physical or occupational therapists, social workers, etc.) using the agency's EHR. We used the narrative part of the HHC visit description, which ranged from lengthy admission notes (often written by a registered nurse) to shorter progress notes (e.g., nurse follow-up notes, physical therapy progress notes, etc.). The average note length was 150 words. Since all visit notes were in the HHC agency's EHR, the text was in an electronic and computer readable format (txt) rather than Portable Document Format (PDF). Each HHC episode in the sample had at least one visit note, thus there was no missing data.

Study endpoints consisted of patient hospitalization or ED visits and were extracted from OASIS. The OASIS assessments are federally mandated of all HHC agencies certified to accept the Centers for Medicare and Medicaid Services payments. OASIS assessments are conducted at the patient's HHC admission and discharge. We used OASIS item "M2300: Emergent Care" from the discharge assessment to document study outcomes.

Clinical Notes Selection

Our goal was to generate a predictive model identifying patient's risk for hospitalization or ED visit early during the HHC episode. Early identification of risk affords enough time for nursing or other care management interventions to potentially reduce hospitalizations and ED visits. To accomplish our goal, we generated a subset of notes documented early during the HHC episode and omitted notes documented relatively close to an event (i.e., hospitalization or ED visit or regular discharge from HHC). Specifically, we reduced the original data set to a subsample of 70% "initial" visit notes documented early in the HHC episode. For example, when a patient had 10 visit notes, we used only the first seven chronological visit notes documented during the episode (time 1 $\rightarrow$ time 7) and omitted the last three visit notes (time 8 $\rightarrow$ time 10). There were 18.5 visit notes per episode on average (median = 7) and the final study sample included 727,676 visit notes.

Data Analysis

To identify patients at high risk of ED visit and hospitalization, we developed an analytical pipeline that included text mining and machine learning methods (see Figure 1). This pipeline can be applied to a large body of clinical notes to: (a) identify patients with high risk of ED visit and hospitalization; and (b) identify key risk factors associated with ED visit and hospitalization, and subsequently, to develop early warning to trigger real time interventions for high-risk patients. The analytical pipeline consists of five components:

1. text preprocessing;
2. feature selection;
3. training of machine learning algorithms;
4. evaluation of the predictive performance of machine learning algorithms; and
5. identification of the most important features.

In general, no special computing resources were needed. All study methods were implemented in KNIME open-source data analytics platform that can be downloaded for free at <https://www.knime.com/downloads> (KNIME, n.d.). In the following sections, we provide detailed information on each component.

Component 1: Text Preprocessing—Text preprocessing is an integral part of any text mining system. Appropriate text preprocessing is vital for the downstream analytical pipeline components. To improve identification of risk factors expressed in the text, we conducted the following text preprocessing tasks: omitting punctuation, omitting common stop words, omitting terms with high frequency, and converting all words to lowercase. See Appendix A (Supplemental Digital Content) for rational and examples of the text preprocessing steps.

Component 2: Feature Selection—Feature selection is the process of selecting a subset of words and expressions from clinical notes that are significantly associated with the outcome of the study. For example, a term “severe pain” might be highly associated with a patient’s risk for hospitalization or emergency department visit. Further, feature selection might improve the accuracy of text classification by excluding “unimportant” words from the clinical notes. We used n-gram as a feature. N-grams help identify common expressions that contain one word or more (e.g., “pain” is a 1-gram, “severe pain” is a 2-gram, and “very severe pain” is a 3-gram). Using n-grams can help identify expressions with relatively high predictive value for projecting hospitalizations and ED visits (Schonlau & Guenther, 2016).

Component 3: Training of Machine Learning Algorithms—Unlike other statistical models that are designed to infer causal relationships between study variables and outcomes (e.g., logistic or linear regression), many machine learning algorithms are designed to identify correlational patterns in the data that help make the most accurate prediction of the study outcome. We used three machine learning algorithms—Naïve Bayes, Decision tree, and Random Forest—to develop prediction models from clinical notes. Although other text mining methods exist (e.g., artificial neural networks), we selected these specific methods

because they are: commonly used in biomedical informatics literature (Sun et al., 2018); can be easily replicated; and provide a good baseline assessment of text mining feasibility.

Naïve Bayes is a probabilistic classifier developed based on the Bayes Theorem (Witten et al., 2011) to evaluate the association between identified features with the outcome class (hospitalization and ED visit) assigned to the document (clinical notes; Witten et al., 2011). Decision Tree algorithm uses decision tree models to go from observations about features such as “severe pain” to assigned class to the document (Witten et al., 2011). Random Forest generates numerous independent trees operating as an ensemble to predict accurate class for a document (Witten et al., 2011). See Appendix B (Supplemental Digital Content) for more information about the machine learning algorithms.

These three algorithms use numeric data to develop prediction models. Therefore, as a first step, we converted all clinical notes into numerical values; a process referred to as vectorization. For this purpose, we used a “one-hot” encoding method, which computes the frequency of each word and expression in a set of clinical notes. For instance, “pain” could be represented as [1, 0, 4, ...], indicating that it was repeated 1 time in the first clinical note, 0 times in the second clinic note, 4 times in the third clinical note, and so on.

To create the training data set for training the machine learning algorithms, we first stratified the sample by the study outcome. This created two subsamples of all HHC episodes—episodes with hospitalization or ED visit versus episodes without hospitalization or ED visit. Next, we randomly selected 70% of the data from each of the two subsamples ($n = 78,566$ episodes) to create the training data set. We saved the remaining 30% of data as a test data set (see Component 4). We used the KNIME platform to train the machine learning algorithms using the training data set. All machine learning algorithms were applied with default settings, including: (a) Algorithm Naïve Bayes was applied without debugging or Kernel estimators; (b) Algorithm Decision Tree (specific type: J48, recursive partitioning) was applied with gain ratio (tree quality measure), reduced error pruning, and minimum number records per node = 2; and (c) Algorithm Random Forest was applied with number of iterations = 100, minimum number of instances = 1, minimum variance for split = $1e-3$, depth = unlimited. See Appendix B (Supplemental Digital Content) for more information about machine learning algorithms applied in this study.

Component 4: Evaluation of Predictive Performance Machine Learning

Algorithms—Once machine learning models were developed on the training data set, we applied these models on the test data set (30% all HHC episodes, $n = 33,671$). Specifically, we generated a machine learning predicted class for each HHC episode (predicted hospitalization or ED visit vs. none) and compared it with the observed HHC class (actual hospitalization or ED visit vs. none). For each machine learning method, we computed standard weighted metrics (Beger, 2016) of: precision—the number of true positives out of the total number of predicted positives; recall—the number of true positives out of actual number of positives; *F*-score, the weighted harmonic mean of the precision, and recall; and Area Under the Precision-Recall Curve (AUC) —a single scalar value that measures the overall performance of a binary classifier. For all metrics, scores range between 0 and 1,

with higher scores indicating better predictive performance. In general, an F -score > 0.8 would indicate that a machine learning algorithm achieves good predictive performance.

Component 5: Identification of the Most Important Features—To identify words and expressions strongly associated with risk for hospitalizations and ED visits, we used gain ratio attribute evaluation technique (Witten et al., 2011), which measure the predictive power of each word and expression for the outcome class (hospitalization and ED visit). We set the gain ratio threshold to > 0 to ensure that only the most informative features were retained by the method. Two PhD prepared nurses with experience in the HHC setting independently reviewed the list of final selected words and expressions and categorized them into clinically meaningful categories (e.g., clinical factors, social/environmental factors). We computed the Kappa interrater agreement (McHugh, 2012) for the category assignments, which was high (Kappa interrater agreement = .94), indicating high initial agreement between the reviewers. All disagreements were resolved via a consultation with an additional PhD-prepared nurse with expertise in clinical informatics.

Results

Overall, 19,301 HHC episodes (17.2% of all episodes) resulted in hospitalization or ED visits. Table 1 shows the performance of machine learning methods, with a best method (random forest) achieving an F -measure of .83 and AUC of .76, indicating good predictive performance.

Using gain ratio attribute evaluation, 388 words and expressions were identified as most associated with patient hospitalization or ED visits. Two HHC nurse reviewers used the six categories seen in Table 2 to classify each word or expression. The interrater agreement was very high (Kappa interrater agreement = 96%) and final consensus of 100% was achieved on all category assignments via reviewer discussion. Table 2 summarizes the distribution of categories associated with patient hospitalization or ED visits. The most prevalent categories were clinical factors (28%), and coordination/communication (23%), while almost one third of the words and expressions were marked as “other” (28%). Table 3 shows the top 10 words and expressions (sorted by gain ratio value) identified in each category.

Discussion

This study is the first to explore the potential contribution of narrative clinical notes in HHC to identify patients at risk for hospitalization or ED visits. Using only HHC visit notes documented relatively early during the HHC episode, we were able to build machine learning models with good predictive performance. These results have several critical implications. First, ability to predict negative outcomes from the HHC EHR data highlight the value of HHC clinical documentation. Although several previous studies in HHC attempted to create models predicting a patient’s risk for hospitalization (Fortinsky et al., 2014; Lohman et al., 2018; Ma et al., 2018; Rosati et al., 2003; Rosati & Huang, 2007), their methodological approaches had several marked limitations. In a recent systematic review, Ma et al. (2018) concluded that major limitations of these studies were due to a focus on patients with a specific disease (often heart failure), reliance on a single administrative data

source, and small sample size. The predictive performance of the existing models also remains low and often not reported (Fortinsky et al., 2014; Lohman et al., 2018; Ma et al., 2018). Our findings suggest that using routinely collected narrative data from the HHC EHR is a critical step in improving early identification of patients at risk for poor outcomes. Research from hospital settings supports these findings (Jayasundera et al., 2018; Linnen et al., 2019).

Our results support the feasibility of creating early warning systems in the HHC setting. We have purposefully used a subsample of 70% of notes documented first (chronologically) during the HHC episode and showed good predictive performance. In practice, this means that warning signs of patient deterioration can be identified from clinical documentation before an adverse outcome occurs. Our findings can be used to develop an early warning system that will notify HHC clinicians of an increasing patient risk, which in turn should trigger risk reduction interventions.

Our results suggest that it is feasible to implement such text predictive pipelines in the HHC clinical settings. Ideally, such risk prediction models would be developed jointly by HHC clinicians and informatics experts and implemented by either the HHC software vendors or other companies that specialize in predictive health analytics.

We also explored a subset of factors highly associated with patient risk predictions and found that clinical factors, such as dehydration or vomiting, were the most prevalent category. Other related factors included care coordination and communication activities (e.g., a conversation with a physician), health service use (e.g. outpatient clinic or physical therapy), or social factors (e.g. living alone or mentions of income. Importantly, many of these factors are found mostly in narrative notes rather than structured data elements within many EHR systems. Other examples of factors found only in notes include patient symptoms (Koleck et al., 2019) or social risk factors (Navathe et al., 2018). For instance, several researchers found that social isolation is documented in the clinical notes (Conway et al., 2019; Zhu et al., 2019) and is associated with rehospitalizations (Navathe et al., 2018).

Our findings also support and extend previous categorizations of risk factors in HHC. For example, in a systematic review focused on identifying factors affecting HHC patient risk for rehospitalization, the authors found that of all socioeconomic factors, only insurance type was predictive of patient rehospitalization (Ma et al., 2018). In another instance, functional limitations were identified as risk factors in HHC previously, and our results further enrich this category adding specific devices like a cane or wheelchair.

Our results also suggest that temporal text features (e.g., “this a.m.”) are associated with negative outcomes. One possible explanation is that these expressions are often used in conjunction with other clinical risk factors (e.g., “fall this a.m.”; Topaz et al., 2019) and thus highlight other risk factors. However, the importance of temporal expressions needs further examination. Finally, about one third of the predictors did not fit into one particular category, requiring further review and potential classification.

Text mining methods, similar to those we used, were applied in settings other than HHC to generate early warnings of a patient’s risk. For example, Mehrabi et al. (2015) developed a

text mining system to identify patients with risk factors of pancreatic cancer in clinical notes. The system provided early warnings that helped clinicians identify patients at risk of pancreatic cancer for further screening and early interventions. In other work, text mining systems were used to identify adverse drug reactions by Zolnoori et al. (2019). Text mining systems were used to identify patients with likely peripheral artery disease based on findings described in radiology reports by Savova et al. (2010). In HHC, our text mining methods can be used to extract risk factors from text and present them to HHC nurses to improve timely risk identification and inform interventions.

Our results suggest that it is feasible to build early warning systems in HHC using routinely collected data from electronic health records. Further studies might explore improvement of risk prediction models by adding more data elements, such as specific symptoms, social risk factors, patterns of previous HHC visits, patient characteristics from standard assessment tools, etc. Researchers might also apply advanced statistical methods that will allow for better adjustment for time-dependence and other potential interdependence between the data element used as predictors, for example by applying time-series analysis and/or mixed effects modeling. Finally, more research is needed in order to understand whether and which early warning systems are effective in the HHC setting.

Limitations

This study has several limitations. First, our results are based on one data set of visit notes and generalizability of our findings to other HHC agencies is limited. Further studies should explore the methods' generalizability by using data from several HHC agencies. Second, some of the outcome data might have been missing due to limited ability of HHC clinicians to follow-up with the discharged patients. In the future, service utilization data from the Centers of Medicare and Medicaid Services might be used to improve outcome completeness. In addition, more granular outcome data might be used to develop different risk models: One for hospitalizations and one for ED visits. Third, our methods only allow for partial understanding of important risk factors and we might have missed other risk categories documented less frequently in visit notes. Additional machine learning methods, such as latent Dirichlet allocation for topic modelling (Blei et al., 2003), can help identify additional categories of risk associated with negative outcomes. In addition, machine learning algorithms might be better adjusted to the data by fine-tuning the model parameters rather than using the default settings (e.g., random forest algorithm's number of iterations might be increased). Finally, this study only used data from clinical narratives; however, integrating other data sources (e.g., structured data about patient medications and diseases) might have resulted in better risk prediction performance.

Conclusion

This study was the first to explore the potential contribution of HHC narrative notes to identifying patients at risk for hospitalization or ED visits. Our results suggest that HHC visit notes are highly informative and can contribute significantly to identification of patients at risk. In addition, predictive risk factors extracted from the text are clinically meaningful and should be used by other studies exploring HHC patient risk factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Kyungmi Woo is supported by the Comparative and Cost-Effectiveness Research (T32NR014205) grant through the National Institute of Nursing Research. The study received ethical review board approval from the Visiting Nurse Services of New York.

References

- Agency for Healthcare Research & Quality. (2016). Potentially avoidable hospitalizations. Retrived date: 1/7/2019 <https://www.ahrq.gov/research/findings/nhqrdr/chartbooks/carecoordination/measure3.html>
- Beger A (2016). Precision-recall curves. SSRN. 10.2139/ssrn.2765419
- Blei DM, Ng AY, & Jordan MI (2003). Latent dirichlet allocation [PDF file]. Journal of Machine Learning, 3, 993–1022. <http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- Busby J, Purdy S, & Hollingworth W (2015). A systematic review of the magnitude and cause of geographic variation in unplanned hospital admission rates and length of stay for ambulatory care sensitive conditions. BMC Health Services Research, 15, 324 10.1186/s12913-015-0964-3 [PubMed: 26268576]
- CMS.gov. (2019). Home health quality measures. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Home-Health-Quality-Measures>
- Collins SA, & Vawdrey DK (2012). “Reading between the lines” of flow sheet data: Nurses’ optional documentation associated with cardiac arrest outcomes. Applied Nursing Research, 25, 251–257. 10.1016/j.apnr.2011.06.002 [PubMed: 22079746]
- Conway M, Keyhani S, Christensen L, South BR, Vali M, Walter LC, Mowery DL, Abdelrahman S, & Chapman WW (2019). Moonstone: A novel natural language processing system for inferring social risk from clinical narratives. Journal of Biomedical Semantics, 10, 6 10.1186/s13326-019-0198-0 [PubMed: 30975223]
- Dykes PC, Rozenblum R, Dalal A, Massaro A, Chang F, Clements M, Collins S, Donze J, Fagan M, Gazarian P, Hanna J, Lehmann L, Leone K, Lipsitz S, McNally K, Morrison C, Samal L, Mlaver E, Schnock K, ... Bates DW (2017). Prospective evaluation of a multifaceted intervention to improve outcomes in intensive care: The promoting respect and ongoing safety through patient engagement communication and technology study. Critical Care Medicine, 45, e806–e813. 10.1097/CCM.0000000000002449 [PubMed: 28471886]
- Fortinsky RH, Madigan EA, Sheehan TJ, Tullai-McGuinness S, & Kleppinger A (2014). Risk factors for hospitalization in a national sample of Medicare home health care patients. Journal of Applied Gerontology, 33, 474–493. 10.1177/0733464812454007 [PubMed: 24781967]
- Hodgson K, Deeny SR, & Steventon A (2019). Ambulatory care-sensitive conditions: Their potential uses and limitations. *BMJ Quality & Safety*, 28, 429–433. 10.1136/bmjqs-2018-008820
- Jayasundera R, Neilly M, Smith TO, & Myint PK (2018). Are early warning scores useful predictors for mortality and morbidity in hospitalised acutely unwell older patients? A systematic review. Journal of Clinical Medicine, 7, 309 10.3390/jcm7100309
- KNIME. (n.d.). Home page. <https://www.knime.com/>
- Koleck TA, Dreisbach C, Bourne PE, & Bakken S (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. Journal of the American Medical Informatics Association, 26, 364–379. 10.1093/jamia/ocy173 [PubMed: 30726935]
- Korach ZT, Cato KD, Collins SA, Kang MJ, Knaplund C, Dykes PC, Wang L, Schnock KO, Garcia JP, Jia H, Chang F, Schwartz JM, & Zhou L (2019). Unsupervised machine learning of topics documented by nurses about hospitalized patients prior to a rapid-response event. Applied Clinical Informatics, 10, 952–963. 10.1055/s-0039-3401814 [PubMed: 31853936]

- Lehman L, Saeed M, Long W, Lee J, & Mark R (2012). Risk stratification of ICU patients using topic models inferred from unstructured progress notes. In AMIA Annual Symposium Proceedings, 2012, 505–511. [PubMed: 23304322]
- Linnen DT, Escobar GJ, Hu X, Scruth E, Liu V, & Stephens C (2019). Statistical modeling and aggregate-weighted scoring systems in prediction of mortality and ICU transfer: A systematic review. *Journal of Hospital Medicine*, 14, 161–169. 10.12788/jhm.3151 [PubMed: 30811322]
- Lohman MC, Scherer EA, Whiteman KL, Greenberg RL, & Bruce ML (2018). Factors associated with accelerated hospitalization and re-hospitalization among Medicare home health patients. *Journals of Gerontology: Series A*, 73, 1280–1286. 10.1093/gerona/glw335
- Ma C, Shang J, Miner S, Lennox L, & Squires A (2018). The prevalence, reasons, and risk factors for hospital readmissions among home health care patients: A systematic review. *Home Health Care Management & Practice*, 30, 83–92. 10.1177/1084822317741622
- Marafino BJ, Boscardin WJ, & Dudley RA (2015). Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to ICU risk stratification from nursing notes. *Journal of Biomedical Informatics*, 54, 114–120. 10.1016/j.jbi.2015.02.003 [PubMed: 25700665]
- McHugh ML (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22, 276–282. <https://hrcak.srce.hr/89395> [PubMed: 23092060]
- Medicare.gov. (2019). Home health compare: Find a home health agency. <https://www.medicare.gov/homehealthcompare/search.html>
- MedPac. (2014). Home health care services In, Report to the Congress: Medicare payment policy (pp. 213–238). Medicare Payment Advisory Commission http://www.medpac.gov/docs/default-source/reports/mar14_entirereport.pdf
- Mehrabi S, Krishnan A, Roch AM, Schmidt H, Li D, Kesterson J, Beesley C, Dexter P, Schmidt M, Palakal M, & Liu H (2015). Identification of patients with family history of pancreatic cancer-investigation of an NLP system portability. *Studies in Health Technology and Informatics*, 216, 604–608. [PubMed: 26262122]
- Navathe AS, Zhong F, Lei VJ, Chang FY, Sordo M, Topaz M, Navathe SB, Rocha RA, & Zhou L (2018). Hospital readmission and social risk factors identified from physician notes. *Health Services Research*, 53, 1110–1136. 10.1111/1475-6773.12670 [PubMed: 28295260]
- Rosati RJ, & Huang L (2007). Development and testing of an analytic model to identify home healthcare patients at risk for a hospitalization within the first 60 days of care. *Home Health Care Services Quarterly*, 26, 21–36. 10.1300/J027v26n04_03
- Rosati RJ, Huang L, Navaie-Waliser M, & Feldman PH (2003). Risk factors for repeated hospitalizations among home healthcare recipients. *Journal for Healthcare Quality*, 25, 4–11. 10.1111/j.1945-1474.2003.tb01038.x
- Savova GK, Fan J, Ye Z, Murphy SP, Zheng J, Chute CG, & Kullo IJ (2010). Discovering peripheral arterial disease cases from radiology notes using natural language processing. *AMIA Annual Symposium Proceedings*, 2010, 722–726. [PubMed: 21347073]
- Schonlau M, & Guenther N (2016). Text mining using N-grams. *SSRN Electronic Journal*. 10.2139/ssrn.2759033
- Sun W, Cai Z, Li Y, Liu F, Fang S, & Wang G (2018). Data processing and text mining technologies on electronic medical records: A review. *Journal of Healthcare Engineering*, 2018. 10.1155/2018/4302425
- Topaz M, Murga L, Gaddis KM, McDonald MV, Bar-Bachar O, Goldberg Y, & Bowles KH (2019). Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches. *Journal of Biomedical Informatics*, 90, 103103. 10.1016/J.JBI.2019.103103 [PubMed: 30639392]
- Topaz M, Radhakrishnan K, Lei V, & Zhou L (2016). Mining clinicians' electronic documentation to identify heart failure patients with ineffective self-management: A pilot text-mining study. *Studies in Health Technology and Informatics*, 225, 856–857. 10.3233/978-1-61499-658-3-856 [PubMed: 27332377]

- Waudby-Smith IER, Tran N, Dubin JA, & Lee J (2018). Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. PLoS ONE, 13, e0198687 10.1371/journal.pone.0198687 [PubMed: 29879201]
- Witten IH, Frank E, & Hall MA (2011). Data mining: Practical machine learning tools and techniques (3rd ed.). Morgan Kaufmann.
- Woo K, Topaz M, Cato K, Shang J, Rossetti S, Fu L-H, McDonald M, Adams V, & Wilson P (2019). Using natural language processing to examine early signs of urinary tract infection in homecare notes H3IT: International Conference on Home Healthcare, Hospice, and Information Technology, Seattle, WA <https://h3it.org/programs/H3IT-2019-program.html>
- Zhu VJ, Lenert LA, Bunnell BE, Obeid JS, Jefferson M, & Halbert CH (2019). Automatically identifying social isolation from clinical narratives for patients with prostate cancer. BMC Medical Informatics and Decision Making, 19, 43 10.1186/s12911-019-0795-y [PubMed: 30871518]
- Zolnoori M, Fung KW, Patrick TB, Fontelo P, Kharrazi H, Faiola A, Shah ND, Wu YSS, Eldredge CE, Luo J, Conway M, Zhu J, Park SK, Xu K, & Moayyedy H (2019). The PsyTAR dataset: From patients generated narratives to a corpus of adverse drug events and effectiveness of psychiatric medications. Data in Brief, 24, 103838 10.1016/j.dib.2019.103838 [PubMed: 31065579]

Figure 1: Study methods overview

The figure visualizes data analytics methods applied in the study. The analytical pipeline consists of five components: 1) text pre-processing, 2) feature selection, 3) training of machine learning algorithms, 4) evaluation of the predictive performance of machine learning algorithms, and 5) identification of the most important features.

Table 1

Performance of machine learning methods

Machine learning method	Recall	Precision	<i>F</i> -score	Area under precision-recall curve (AUC)
Naïve Bayes	0.59	0.71	0.62	0.61
Decision Tree (J48)	0.73	0.74	0.73	0.68
Random Forest	0.81	0.83	0.82	0.76

Table 2

Category definitions for words and expressions associated with hospitalization or emergency department visits among HHC patients

Category name	Category description
Clinical factors	Any words or expressions suggesting clinical problems, including symptoms, diseases, functional status limitations, clinical interventions, etc.
Coordination/communication	Any words or expressions mentioning coordination/communication activities related to patient care.
Service use	Any words or expressions mentioning care settings or professional services.
Social/environmental factors	Any words or expressions mentioning social or environmental factors like patient's living situation, family support, etc.
Temporal	Any words or expressions mentioning time.
Device/equipment	Any words or expressions mentioning medical device/equipment
Other	Other ambiguous words or expressions.

Table 3

Examples of top 10 words and expressions identified in each category

Clinical factors	Coordination/ communication	Service use	Social/ environmental factors	Temporal	Device/ equipment	Other
dehydrated	d/c	to er	shopping	in am	catheter	goals met
vomited	with ns	no skilled	lives alone	started on	diaper	negotiate
nausea	md called	to hospital	in apt	today vn	with cane	all goals
wound care	c[are] planning	skilled nursing	income	last night	Foley [catheter]	final
incision	plan discussed	outpatient	pt lives with	this am	rollator	no further
kidney	pt in agreement	no further skilled	private insurance	today for	in chair	experienced
cleansed	c[are] plan	hospitalized for	children	when he	with rollator	large
worsening	supplies ordered	surgery	spouse	at pm	rolling walker	render
unsteady gait	vn called	physical therapy	child	next week	device	yo female
hx falls	vn unable to	pharmacy	house	tomorrow	wheel chair	pt was

Note. Hx = history; d/c = discharge; ns = nursing service; pt = patient; vn = visiting nurse; er = emergency room; yo = years old