

Himmelfarb Health Sciences Library, The George Washington University

## Health Sciences Research Commons

---

Clinical Research and Leadership Faculty  
Publications

Clinical Research and Leadership

---

1-1-2016

### Identification and Use of Frailty Indicators from Text to Examine Associations with Clinical Outcomes Among Patients with Heart Failure.

Yijun Shao

April F Mohanty

Ali Ahmed

Charlene R Weir

Bruce E Bray

*See next page for additional authors*

Follow this and additional works at: [https://hsrc.himmelfarb.gwu.edu/smhs\\_crl\\_facpubs](https://hsrc.himmelfarb.gwu.edu/smhs_crl_facpubs)



Part of the [Health and Medical Administration Commons](#)

---

#### APA Citation

Shao, Y., Mohanty, A., Ahmed, A., Weir, C., Bray, B., Shah, R., Redd, D., & Zeng-Treitler, Q. (2016). Identification and Use of Frailty Indicators from Text to Examine Associations with Clinical Outcomes Among Patients with Heart Failure.. *AMIA ... Annual Symposium Proceedings [electronic resource] / AMIA Symposium. AMIA Symposium, 2016* (). <http://dx.doi.org/no doi>

This Journal Article is brought to you for free and open access by the Clinical Research and Leadership at Health Sciences Research Commons. It has been accepted for inclusion in Clinical Research and Leadership Faculty Publications by an authorized administrator of Health Sciences Research Commons. For more information, please contact [hsrc@gwu.edu](mailto:hsrc@gwu.edu).

---

**Authors**

Yijun Shao, April F Mohanty, Ali Ahmed, Charlene R Weir, Bruce E Bray, Rashmee U Shah, Douglas Redd, and Qing Zeng-Treitler

# Identification and Use of Frailty Indicators from Text to Examine Associations with Clinical Outcomes Among Patients with Heart Failure

Yijun Shao, Ph.D.<sup>1</sup>, April F. Mohanty, Ph.D., M.P.H.<sup>2</sup>, Ali Ahmed, M.D., M.P.H.<sup>3</sup>,  
Charlene R. Weir, Ph.D., R.N.<sup>1,2</sup>, Bruce E. Bray, M.D.<sup>1</sup>, Rashmee U. Shah, M.D., M.S.<sup>4</sup>,  
Douglas Redd, Ph.D.<sup>1,2,3,5</sup>, Qing Zeng-Treitler, Ph.D.<sup>1,2,3,5</sup>

<sup>1</sup>Biomedical Informatics, School of Medicine, University of Utah, Salt Lake City, UT, USA;  
<sup>2</sup>Informatics Decision-Enhancement and Analytic Sciences (IDEAS) Center, George E. Wahlen Department; Veterans Affairs Medical Center, Salt Lake City, UT, USA;  
<sup>3</sup>Washington DC VA Medical Center, Washington, DC, USA; <sup>4</sup>University of Utah, Division of Cardiovascular Medicine, Salt Lake City, UT, USA; <sup>5</sup>Department of Clinical Research and Leadership, George Washington University, Washington, DC, USA

## Abstract

*Frailty is an important health outcomes indicator and valuable for guiding healthcare decisions in older adults, but is rarely collected in a quantitative, systematic fashion in routine healthcare. Using a cohort of 12,000 Veterans with heart failure, we investigated the feasibility of topic modeling to identify frailty topics in clinical notes. Topics were generated through unsupervised learning and then manually reviewed by an expert. A total of 53 frailty topics were identified from 100,000 notes. We further examined associations of frailty with age-, sex-, and Charlson Comorbidity Index-adjusted 1-year hospitalizations and mortality (composite outcome) using logistic regression. Frailty ( $\geq 4$  topics versus  $<4$ ) was associated with twice the risk of the composite outcome, Odds Ratio: 2.2, 95% Confidence Interval: (2.0-2.4). This study demonstrates the feasibility of identifying frailty indicators from clinical notes and linking these to clinically relevant outcomes. Future work includes integrating frailty indicators into validated predictive tools.*

## Introduction

Frailty is an important, but frequently overlooked, determinant and indicator of health outcomes in older adults<sup>1-4</sup>. It is distinct from comorbidity and disease, and is an age-related, multidimensional combination of fatigue, weakness, malnutrition, and greater vulnerability to stressors<sup>5-7</sup>. Increasing frailty is often co-existent among patients diagnosed with chronic conditions and it is associated with the severity of acute illness and intensity of healthcare utilization<sup>8,9</sup>. Frailty is widely recognized as a risk factor for morbidity, including cognitive decline, mortality, as well as a barometer of how well patients respond to treatment<sup>7,9</sup>. Therefore, this metric can be valuable for informing healthcare decisions, including recommended surgical procedures, end-of-life care, and living arrangements<sup>7</sup>. In order to provide an efficient and effective patient-centered care, providers should be aware of each patient's frailty status for incorporation into clinical decision making.

Despite its importance, frailty measurements are rarely collected in a quantitative, systematic fashion in routine patient care<sup>9</sup>. Frailty extraction is complex, since information is rarely collected as coded/structured data and clinical note terminology varies significantly. This complexity limits the ability to evaluate the association between frailty severity and clinical outcomes, and prevents healthcare providers from using frailty information in clinical decisions<sup>8</sup>. However, healthcare providers commonly document various aspects of frailty in clinical notes, especially when treatment plans change and/or when patients' quality of life is impacted. Therefore, clinical notes can provide the data for a quantitative frailty metric as well as a longitudinal record of frailty.

Extracting frailty from clinical notes is challenging because frailty is a complex and multi-faceted concept. The use of the term "frail" is not common, while the description of frailty indicators such as "tired," "slow walk," "unsteady gait" and "confused" are abundant. In a prior study, we identified terms describing functional status, one aspect of frailty<sup>10</sup>. Among the hundreds of functional status terms we collected from clinical notes and social media, only a fraction could be mapped to a controlled vocabulary. Furthermore, for the purpose of retrospective analysis or prospective decision support, frailty should not be treated simplistically as a binary variable. The Frailty Scale developed by Dr. Rockwood, for instance, categorized frailty status into 9 levels<sup>11</sup>. As a result, to determine frailty

status from text, we cannot rely on a small number of known terms or the existing controlled vocabulary. Rather, it is necessary for us to identify a rich set of indicators of frailty and differentiate the degrees of frailty.

Topic modeling is a potential approach for identifying indicators of frailty in clinical notes. Briefly, topic modeling is a machine learning method based on statistical models that assess words in documents for identifying hidden themes in a corpus of free text documents<sup>12</sup>. By first focusing on themes rather than individual words/terms, we hypothesize that it would allow us to capture a wide range of text descriptions potentially related to frailty.

Our aims are to 1) demonstrate the feasibility of using topic modeling of clinical notes to identify frailty indicators and 2) using these frailty indicators to examine associations with clinical outcomes among patients diagnosed with heart failure (HF). HF is a highly prevalent condition that results from structural or functional impairment of the heart's ability to fill or eject blood. Recent projections estimate that >8 million Americans (3%) will have HF by 2030<sup>13, 14</sup>. Evidence suggests that frailty is worse among patients with HF compared to similarly aged adults, with similar comorbidities, from the general population<sup>15-17</sup>. Commonly reported HF symptoms such as dyspnea and fatigue can lead to greater frailty<sup>14</sup> and one of the treatment goals is to improve frailty status. Also, since HF patients may be referred for cardiac surgery or other invasive interventions as symptoms worsen (though less common among older HF patients), measurement of frailty can inform treatment strategies and facilitate examination of treatment response<sup>14</sup>. A method to systematically estimate frailty among all HF patients is, therefore, critically important.

## Methods

### *Data*

We used Veterans' electronic medical record (EMR) data which are maintained by the Veterans Administration Informatics and Computing Infrastructure (VINCI) database<sup>18</sup>. VINCI includes data on healthcare utilization (inpatient and outpatient encounters), clinical parameters, and demographics (e.g. age, gender) as well as narrative text from clinical encounters or Text Information Utility (TIU) notes. We used the International Classification of Disease 9th Clinical Modification (ICD-9-CM) codes 428.0-428.9 from inpatient (primary/principal or secondary diagnoses) and outpatient records to identify patients with HF. We identified 12,000 patients with a HF diagnosis in Fiscal Year 2010, among which 4,000 experienced death or  $\geq 2$  HF hospitalizations during the year after diagnosis. Then we extracted Veterans' gender, calculated their age and Charlson Comorbidity Index (CCI) at the first HF diagnosis. The Charlson Comorbidity Index is a measure of disease burden that is determined from comorbid conditions using methods described by Quan *et al.*<sup>19</sup>. These Veteran characteristics were either directly or indirectly ascertained from the structured data tables. We also retrieved all the clinical notes from the patients included in our study dated within one year before the first HF diagnosis. There were a total of 709,389 notes available for the 12,000 Veterans diagnosed with HF. Frailty indicators, which were not available from the structured data, were extracted from these notes using a topic modeling technique. All data access and use was approved through local Institutional Review Board and VINCI policies and no human subjects were contacted for this study.

### *Topic Modeling and Stable Topic Extraction*

We randomly sampled 100,000 notes from the corpus of 709,389 notes for topic modeling. Random samples (50,000 each) were balanced between the 4,000 patients with and the 8,000 patient without 1-year outcomes ( $\geq 2$  HF hospitalizations or mortality). We tokenized these texts by converting all uppercase letters into lowercase, removing all numbers and punctuations except hyphens (" - ") that join two words. We used the Latent Dirichlet Allocation (LDA)<sup>20</sup> program from the MALLET package, a topic modeling tool that is written in Java<sup>22</sup>. LDA is one of the most widely used topic modeling approaches and makes assumptions that topics are probabilistic distributions over words and documents are mixtures of topics. It has found applications in many areas including biomedicine<sup>12, 21</sup>. We set the initial number of topics to be 700 and ran LDA on the 100,000 notes to obtain a list of 700 topics. Then we applied the learned topic model to the full collection of 709,389 notes using the topic inferring tool also included in MALLET. This step yields topic proportions per note for each note. By the design of LDA, every topic has a non-zero proportion within every note. We set 0.02 as the cut-off on the topic proportion in a note to select the topics present in the note.

One major issue when using LDA is the presence of a high amount of noise (irrelevant terms) in some topics, which results in high false positives. To overcome this problem, we designed a method to identify the non-noise topics which we called "stable" topics. Since LDA is an iterative algorithm that initially assigns terms randomly to topics, the topic model it generates varies each time LDA is applied to a test corpus. At the same time, certain topics re-

occur in repeated experiments. For example, in this study we observed the following two topics (only top 10 most probable words are listed here) from two independent runs:

Word	Count	Probability
walker	1440	0.106
gait	1071	0.079
cane	974	0.072
ambulates	433	0.032
steady	433	0.032
ambulating	429	0.032
ambulation	421	0.031
unsteady	370	0.027
walking	320	0.024
ambulate	312	0.023

Run 1, Topic 24

Word	Count	Probability
walker	1799	0.114
gait	1301	0.083
cane	1172	0.074
ambulation	467	0.030
walking	437	0.028
balance	407	0.026
steady	388	0.025
ambulates	386	0.025
ambulating	338	0.021
unsteady	334	0.021

Run 2, Topic 144

Although the topics do not have exactly the same probability distributions over the words, they are very similar, and should be viewed as the “same” topic that recur in different runs.

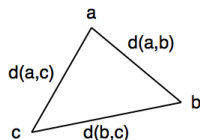
We postulate that non-recurring topics and topics terms are likely to be noise as they are influenced by the initial random seeds. To identify the recurring topics, we proceeded as follows. Given two topics  $a$  and  $b$  with word counts  $ct_a$  and  $ct_b$  respectively, we define the distance between  $a$  and  $b$  to be

$$d(a, b) = 1 - \frac{\sum_w ct_a(w) \times ct_b(w)}{\sqrt{\sum_w ct_a(w)^2} \times \sqrt{\sum_w ct_b(w)^2}}$$

which is known as the cosine-distance. To identify stable topics generated by LDA corresponding to some parameter set (e.g., number of topics, number of iterations, alpha), we ran LDA 3 times under the same parameter set (but with different random seeds) to produce 3 sets of topics, say,  $A, B, C$ . Then, for each topic triple  $(a, b, c)$  with  $a \in A, b \in B, c \in C$  we compute the size of the triple by taking the maximum of the 3 distances:

$$d(a, b, c) = \max [d(a, b), d(a, c), d(b, c)]$$

One can visualize a topic triple as a triangle, so that the terms such as “size”, “small”, “large” make more sense.



We aligned up the three sets of topics using a greedy algorithm as follows. We first found the triple  $(a_1, b_1, c_1)$  of the smallest size from all such triples, and second found the triple  $(a_2, b_2, c_2)$  of the smallest size from all the triples  $(a, b, c)$  with  $a \in A \setminus [a_1], b \in B \setminus [b_1], c \in C \setminus [c_1]$ , and third found the triple  $(a_3, b_3, c_3)$  of the smallest size from all the triples  $(a, b, c)$  with  $a \in A \setminus [a_1, a_2], b \in B \setminus [b_1, b_2], c \in C \setminus [c_1, c_2]$ , and so on. Thus we aligned the three sets of topics by their size from small to large. For any topic triple  $(a, b, c)$  smaller than an adjustable parameter  $r$  we made a new topic by averaging the counts of  $a, b$  and  $c$ :

$$ct_{avg}(w) = \frac{1}{3}[ct_a(w) + ct_b(w) + ct_c(w)]$$

and we called the new topic a stable topic. In this study, we empirically set  $r = 0.7$  and obtained 556 stable topics.

To determine what stable topics were present in each note, we proceeded as follows. First we independently applied each of the 3 learned LDA models to the full collection of 709,389 notes using the topic inference tool included in MALLET. This step yielded 3 topic proportions per note for each note. Next, we defined a stable topic to be present in a note if at least 2 of the 3 topics in the topic triple corresponding to the stable topic had a proportion of 0.02 or higher in that note. This enabled us to make a list of stable topics present in a note for each note.

### *Topic Interpretation and Frailty Topic Identification*

The “topics” generated by topic modeling methods are simply probability distributions over the whole vocabulary (consisting of thousands of words), and every word has a non-zero probability in every topic. Since we need to identify topics related to frailty, the first step is to interpret all the probability distributions over words. Obviously, not all the words are equal for making interpretations: words with higher probabilities are more important than those with lower probabilities. For our project, we list the top 20 most probable words for each topic and ask the human reviewers to focus on the top 10 words to make interpretation and use the next 10 words as auxiliary.

An informatics expert and a medical expert independently reviewed the list of 556 stable topics, interpreted them as described above, and identified those that represent certain aspect(s) of frailty. Then we calculated the inter-rater agreement on the identification of frailty topics. See Table 1 for examples of frailty topics and their interpretations.

### *Counting Frailty Topics for Patients*

We counted the frailty topics for each patient from all the notes dated within one year of the first diagnosis of HF. The topics were counted only once, even if they appeared in multiple notes. These numbers were recorded as the frailty data. We hypothesized that the number of frailty topics a patient has can be used as an indicator of the frailty severity: the higher the number, the more severe the frailty.

### *Cases and Controls*

We divided the 12,000 HF patients into case and control groups. Cases were defined as patients who died or had  $\geq 2$  all-cause hospitalizations within the first year following their first diagnosis of HF (our composite outcome). Controls were patients who did not have the composite outcome within the 1-year period following their initial HF diagnosis. This yielded 5,207 cases and 6,793 controls.

### *Statistical Analyses*

To explore the relevance of the frailty indicators, we ascertained the composite outcomes (mortality/hospitalization) and we conducted several statistical analyses.

#### *1) Correlation Analysis*

We computed the prevalence of patients with  $n$  ( $n=0, 1, 2, \dots$ ) frailty topics in the cases and also in controls. Then we calculated ratio of the prevalence in cases to those in controls for each  $n$ . These ratios measure the population level (i.e., cases vs controls) difference in frailty of various levels of severity. Therefore, the correlation between the ratios and the number of frailty topics (i.e.,  $n$ ) is an indicator of the relationship between frailty and the composite outcome.

#### *2) Logistic Regression Models*

We used logistic regression models to analyze the association of frailty and the composite outcome at the individual level. We adjusted for age, gender, and CCI level. The outcome was modeled as a binary variable: case was coded as 1 and control was coded as 0 (reference). Age (in years) was measured at the time of the first HF diagnosis. CCI was an integer ranging from 0 to 33. Age and CCI were modeled as continuous variables. Gender was modeled as a binary variable: 0 for male and 1 for female. We parameterized the frailty variable in two ways: 1) as a continuous variable and 2) as a binary variable. As a continuous variable, frailty was coded as the number of frailty topics we counted for each patient. As a binary variable, frailty was coded as 0 (reference) if the number of frailty topics was  $\geq 4$  topics and coded as 1 for  $<4$  topics. We chose 4 as the cut-off based on the results of correlation analysis (see Figure 1 and 2). For both cases, we calculated the odds ratio for the composite outcome associated with the frailty variable based on the coefficients output by the logistic regression models.

## **Results**

We obtained 556 stable topics from the topic modeling step. An informatics expert reviewed these topics and identified 53 of them as frailty topics. A medical expert independently identified frailty topics resulted and had an inter-rater agreement of Kappa = 0.818 with the first reviewer, demonstrating that the first expert’s result was reliable. In Table 1, we list eight frailty topics for illustration.

The age, gender, and CCI characteristics of our study population of 12,000 Veterans with new diagnoses of HF are described in Table 2. Mean age was around 70 years old for cases and controls, most were male, and cases generally had a higher CCI compared to controls.

**Table 1.** Eight examples of frailty topics. The first column lists the labels assigned by the informatics expert as human interpretations. The second column lists the top 20 words in order of descending probability.

Topic	Top 20 most probable words
Fall Risk	fall risk score morse gait ambulatory status scale mental falling secondary diagnosis total oriented history aid ability iv lock bedrest
Bedside Condition	bed resting monitor noted distress continue complaints quietly sleeping bedside discomfort comfortably watching tv lying remains vss received unlabored sitting
Bedside Assist	bed call reach light position fall low side locked rails safety assistance lowest room free risk precautions environment monitor place
Home Care	home care services health va referral agency community nursing service aide hospice hha skilled referred fax visits medicare fee start
Family Assist	daughter son home called spoke father call family states law dtr living told stated lives assisted daughters dementia received aware
Mobility Assist	walker cane gait wheelchair ambulation steady ambulates ambulating walking unsteady equipment device home ambulate falls walk assistance balance scooter mobility
Cognitive Status	memory cognitive average speech evaluation range difficulty recall words communication functioning task attention dementia information impaired ability deficits accuracy impairment
Shoe-wear Ability	shoes wear issued fit size instructed wearing prosthetics proper device equipment care stockings compression shoe hose wheelchair scooter measured order

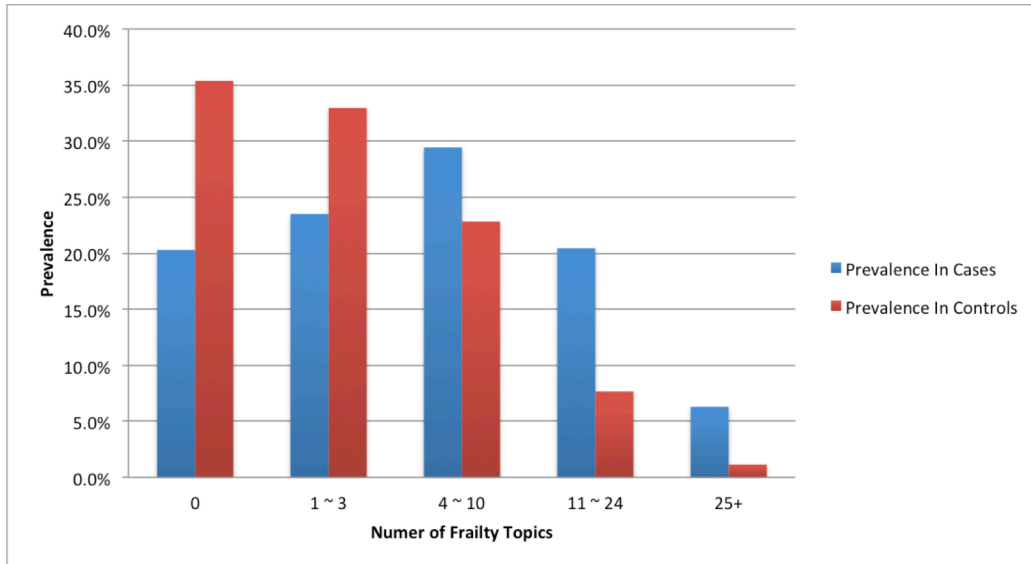
**Table 2.** Study population characteristics

Characteristic	Cases N = 5,207	Controls N = 6,793
Age (year, mean(SD))	70.5 (13.0)	69.1 (11.4)
Men (%)	5096 (97.9)	6647 (97.9)
Charlson Comorbidity Index (%)		
0	1248 (24.0)	2516 (37.0)
1	953 (18.3)	1451 (21.4)
2	722 (13.9)	998 (14.7)
3	661 (12.7)	717 (10.6)
4	543 (10.4)	515 (7.6)
5+	1080 (20.7)	596 (8.8)

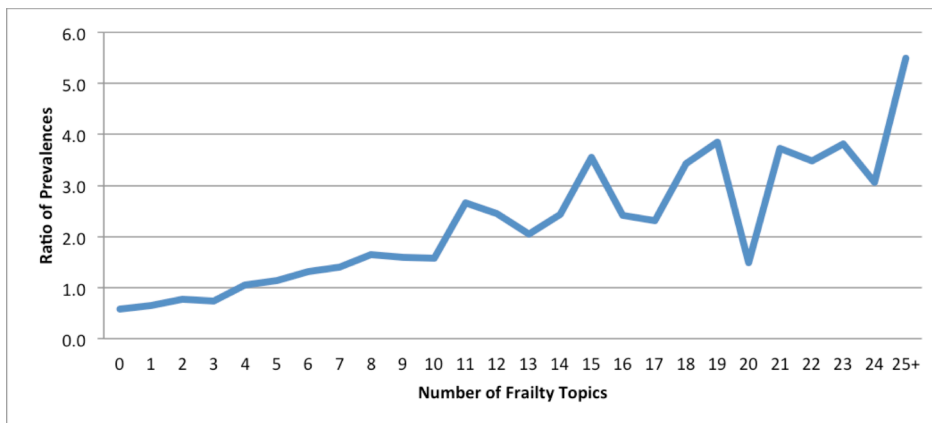
Among cases, 20.3% patients had no frailty topics mentioned in the notes, 8.7% had one frailty topic, 8.4% had two frailty topics, 6.5% had three, 6.5% had four, 5.5% had five, etc. Among controls, 35.4% patients had no frailty topics, 13.4% had one, 10.8% had two, 8.8% had three, 6.1% had four, 4.8% had five, etc.

Comparing the two distributions, when the number of frailty topics was less than 4, the prevalence of patients having that number of topics in the cases was lower than that in controls. However, patients with  $\geq 4$  topics, the prevalence of patients having that number of topics was higher in the cases than in the controls. This is illustrated in Figure 1. This led to the choice of 4 as the cut-off in making frailty into a binary variable in one of logistic regression models.

To make the contrast clearer, we calculated the ratio of the prevalence in cases to that in controls for each number of topics, except for 25 or more topics where we grouped them together into one category. The correlation of this ratio with the number of frailty topics is 0.879.



**Figure 1.** Patient distribution over different ranges of numbers of topics in cases and in controls. The ranges are: 1) 0 topics, 2) 1 ~ 3 topics, 3) 4 ~ 10 topics, 4) 11 ~ 24 topics, 5) 25 or more topics.



**Figure 2.** Ratio of prevalence of patients with various numbers of frailty topics in the cases to that in the controls. Correlation of this ratio to the number of frailty topics was 0.879.

Figure 2 indicates that the ratio was below 1 when the number of frailty topics (n) was below 4, between 1 and 2 when n was between 4 and 10, and above 2 when n was above 11 (the only exception was when n=20 should be viewed as an effect of random sampling).

The results of logistic regression models are presented in Tables 3 and 4, corresponding to treating frailty as a continuous variable and a binary variable (reference, <4 topics), respectively. In both models, frailty was associated with the outcomes and these results were statistically significant. The coefficient (0.07) of frailty in Table 3 can be interpreted as the odds ratio = 1.07 (=exp(0.07)), 95% Confidence Interval (CI): 1.07-1.08, for each additional frailty topic associated with the risk of the composite outcome. Thus, each additional frailty topic was associated with a 7% higher risk of the composite outcome. The coefficient (0.789) corresponds to an odds ratio of 2.20 (95% CI: 2.02-2.39) for the association between  $\geq 4$  frailty topics (vs <4) and the composite outcome, Table 4. In other words,  $\geq 4$



frailty topics was associated with two times the risk of having the composite outcome 1-year following an initial HF diagnosis.

**Table 3.** Logistic regression with frailty modeled as a continuous variable (reference, 0 topics)

	<b>Intercept</b>	<b>Age</b>	<b>Gender</b>	<b>CCI</b>	<b>Frailty</b>
Coefficient	-1.371	0.008	-0.127	0.099	0.070
95% CI	(-1.595, -1.146)	(0.004, 0.011)	(-0.391, 0.138)	(0.080, 0.118)	(0.063, 0.077)
P-value	<0.001	<0.001	0.347	<0.001	<0.001

**Table 4.** Logistic regression with frailty as a binary variable (reference, <4 topics)

	<b>Intercept</b>	<b>Age</b>	<b>Gender</b>	<b>CCI</b>	<b>Frailty</b>
Coefficient	-1.399	0.008	-0.069	0.122	0.789
95% CI	(-1.622, -1.175)	(0.004, 0.011)	(-0.328, 0.191)	(0.104, 0.140)	(0.707, 0.872)
P-value	<0.001	<0.001	0.604	<0.001	<0.001

## Discussion and Conclusion

In this study we demonstrate the feasibility of extracting frailty-related topics from clinical text and that these topics are associated with clinically relevant outcomes. We are unaware of previous studies that have extracted frailty-related topics from clinical notes and associated these topics for clinical outcomes. Our results show that the frailty data we extracted is correlated with mortality and/or hospitalizations at both the population and the individual levels. In particular, the number of frailty topics for a given patient may serve as a proxy for the severity of frailty of the patient. A patient with one or two frailty topics does not necessarily mean the patient is frail: for example, that a patient needs mobility assistance may be a result of injury. However, with more frailty topics, it is more likely that the patient is frail.

Our results have the potential for significant impact. Healthcare providers use risk prediction models to estimate treatment risk. For example, the Society of Thoracic Surgeons uses a calculator to estimate the risk of adverse clinical outcomes following cardiac surgeries. This calculator is inaccurate for older adults—it underestimates risk in frail patients and overestimates risk in healthy patients<sup>23, 24</sup>. A tool that allows frailty to be incorporated into the STS, and other, models could improve the accuracy of risk prediction. We hope that our project can lead to a better measure of frailty that can be useful in routine use of clinical decision support.

Our findings are consistent with previous studies of frailty, not just indicators of frailty, and clinical outcomes. For instance three of the most commonly used measures of frailty include the Fried Frailty Phenotype model<sup>25</sup>, the Frailty Index developed by Rockwood *et al.*<sup>11</sup>, and gait speed<sup>26</sup>. These measures have been found to be associated with future risk of hospitalizations and mortality. However, these measures are not integrated in routine care for all patients, since they can be time consuming to complete and add additional burden to patients and providers. Thus, these measures are often reserved for randomized controlled studies, as in the recent landmark Systolic Blood Pressure Intervention Trial (SPRINT)<sup>27</sup>, or smaller target populations, such as among nursing home residents. Thus, the potential use of clinic notes from the EMR supports opportunities for larger and more generalizable or pragmatic studies of frailty and incorporation into routine healthcare decision making.

We can conclude that the topic modeling approach is capable of extracting the frailty information from text. This approach is also applicable to much larger datasets. Because LDA is a time-consuming program, instead of running LDA directly on all the notes (about 700,000 notes), we sampled a sufficiently large subset (100,000 notes) so that LDA can finish the job in a reasonable time. Then we applied the learned topic model to the all the notes to infer the topic distributions in each of them. The topic inferring program is much less time-consuming than LDA, so the process can finish in a reasonable time while still obtaining the outputs we need: the list of topics and the topic distributions in each note. One novelty of our topic modeling approach is the stable topic identification method.

Topic modeling is an unsupervised machine learning method, which means it does not require human annotations to train the model. Therefore our approach does not require much labor-intensive work such as reading notes and manual annotation. That being said, some manual work – interpreting the computer-extracted topics and identify those that could be indicators of frailty – is still required. However, the human labor is much less than annotating notes, because a reviewer only needs to read 20 words (in this study) for a topic rather than hundreds of words for a note.

Some frailty topics identified by our experts may appear to be cognition-related (see Table 1). This is because there is an association between frailty and cognitive impairment: frailty and cognition interact within a cycle of age-associated decline, therefore some researchers regard cognition deficits as indicators of frailty<sup>2</sup>.

Since this study is our first attempt of identifying frailty from text, there are a number of limitations. 1) A note having a frailty topic could indicate absence of frailty. Using a number of (e.g.  $\geq 4$ ) distinct frailty topics as indicator greatly increases the chance of finding frailty, although the certainty is still not 100%. 2) When we used the number of frailty topics as an indicator of the severity of frailty, we implicitly assumed that all the frailty topics had the same contribution to the severity of frailty. In reality different frailty topics may indicate different levels of severity. 3) Our logistic regression analysis only took into account 3 commonly used predictors in HF patients for our composite outcome: age, gender and CCI. It is likely that many other factors that also correlate to the outcome of mortality/hospitalizations. 4) Although we attempted to identify Veterans with a new or incident diagnosis of HF, patients could have sought care outside the Veterans Health Administration for their initial diagnosis and would have been included in this study of incident HF patients. In addition to addressing these limitations, future studies are needed to replicate our preliminary findings and to extend our findings to other clinical domains. Also, further work is needed to examine changes in frailty indicators from clinical notes over time and to incorporate frailty indicators into validated predictive tools for clinical decision making.

Based on the results of this study, we conclude that it is feasible to identify indicator of frailty from routinely documented clinical notes. We also demonstrate that these indicators are related to clinically relevant outcomes in a population of Veterans diagnosed with HF.

### Acknowledgements

This work is funded by the US Department of Veterans Affairs, Office of Research and Development, Health Services Research and Development grants CHIR HIR 08-374, HIR 08-204, CRE 12-315 and the CREATE: A VHA NLP Software Ecosystem for Collaborative Development and Integration. Dr. Mohanty is supported by the VA Advanced Fellowship Program in Medical Informatics of the Office of Academic Affiliations, Department of Veterans Affairs. We would also like to acknowledge the staff, resources and facilities of the VA Salt Lake City IDEAS Center.

### References

1. Lahousse L, Maes B, Ziere G, et al. Adverse outcomes of frailty in the elderly: the Rotterdam Study. *European Journal Of Epidemiology*. 2014; 29: 419-27.
2. Xue Q-L. The frailty syndrome: definition and natural history. *Clinics In Geriatric Medicine*. 2011; 27: 1-15.
3. Buckinx F, Rolland Y, Reginster J-Y, Ricour C, Petermans J and Bruyère O. Burden of frailty in the elderly population: perspectives for a public health challenge. *Archives Of Public Health = Archives Belges De Santé Publique*. 2015; 73: 19-.
4. Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: a call to action. *Journal Of The American Medical Directors Association*. 2013; 14: 392-7.
5. Clegg A, Young J, Iliffe S, Rikkert MO and Rockwood K. Frailty in elderly people. *Lancet*. 2013; 381: 752-62.
6. Lally F and Crome P. Understanding frailty. *Postgraduate Medical Journal*. 2007; 83: 16-20.
7. Robinson TN, Walston JD, Brummel NE, et al. Frailty for Surgeons: Review of a National Institute on Aging Conference on Frailty for Specialists. *J Am Coll Surg*. 2015; 221: 1083-92.
8. McNallan SM, Singh M, Chamberlain AM, et al. Frailty and healthcare utilization among patients with heart failure in the community. *JACC Heart Fail*. 2013; 1: 135-41.
9. Clegg A, Young J, Iliffe S, Rikkert MO and Rockwood K. Frailty in elderly people. *Lancet*. 2013; 381: 752-62.
10. Kuang J, Mohanty AF, V.H. R, Weir CR, Bray BE and Zeng-Treitler Q. Representation of Functional Status Concepts from Clinical Documents and Social Media Sources by Standard Terminologies. *AMIA Annu Symp Proc*. 2015.
11. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *CMAJ*. 2005; 173: 489-95.

12. Zhao W, Zou W and Chen JJ. Topic modeling for cluster analysis of large biological and medical datasets. *BMC Bioinformatics*. 2014; 15 Suppl 11: S11.
13. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circ Heart Fail*. 2013; 6: 606-19.
14. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*. 2013; 62: e147-239.
15. Cacciatore F, Abete P, Mazzella F, et al. Frailty predicts long-term mortality in elderly subjects with chronic heart failure. *Eur J Clin Invest*. 2005; 35: 723-30.
16. Newman AB, Gottdiener JS, McBurnie MA, et al. Associations of subclinical cardiovascular disease with frailty. *J Gerontol A Biol Sci Med Sci*. 2001; 56: M158-66.
17. Afilalo J, Karunanathan S, Eisenberg MJ, Alexander KP and Bergman H. Role of frailty in patients with cardiovascular disease. *Am J Cardiol*. 2009; 103: 1616-21.
18. VA Informatics and Computing Infrastructure (VINCI)2015: (2012).
19. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005; 43: 1130-9.
20. Blei DM, Ng AY and Jordan MI. Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 2003: 993-1022.
21. Shao Y, Weir C, Zeng-Treitler Q and Estrada N. Identifying Documentation of Delirium in Clinical Notes through Topic Modeling. *Healthcare Informatics (ICHI), 2015 International Conference on*. IEEE, 2015, p. 335-40.
22. McCallum AK. MALLET: A Machine Learning for Language Toolkit. 2002.
23. Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1--coronary artery bypass grafting surgery. *Ann Thorac Surg*. 2009; 88: S2-22.
24. Shahian DM, O'Brien SM, Sheng S, et al. Predictors of long-term survival after coronary artery bypass grafting surgery: results from the Society of Thoracic Surgeons Adult Cardiac Surgery Database (the ASCERT study). *Circulation*. 2012; 125: 1491-500.
25. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001; 56: M146-56.
26. Dodson JA, Arnold SV, Gosch KL, et al. Slow Gait Speed and Risk of Mortality or Hospital Readmission After Myocardial Infarction in the Translational Research Investigating Underlying Disparities in Recovery from Acute Myocardial Infarction: Patients' Health Status Registry. *J Am Geriatr Soc*. 2016.
27. Wright JT, Jr., Williamson JD, Whelton PK, et al. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med*. 2015; 373: 2103-16.