Title
Characterizing shared and distinct symptom clusters in common chronic conditions through natural language processing of nursing notes

Journal
Research in Nursing & Health, 44(6)

ISSN
0160-6891

Authors
Koleck, Theresa A
Topaz, Maxim
Tatonetti, Nicholas P
et al.

Publication Date
2021-12-01

DOI
10.1002/nur.22190

Peer reviewed

HHS Public Access
Author manuscript
Res Nurs Health. Author manuscript; available in PMC 2022 December 01.

Published in final edited form as:
Res Nurs Health. 2021 December ; 44(6): 906–919. doi:10.1002/nur.22190.

Characterizing shared and distinct symptom clusters in common chronic conditions through natural language processing of nursing notes

Theresa A. Koleck[1], **Maxim Topaz**[2,3], **Nicholas P. Tatonetti**[3,4,5,6,7], **Maureen George**[2], **Christine Miaskowski**[8], **Arlene Smaldone**[2,9], **Suzanne Bakken**[2,3,4]

[1]School of Nursing, University of Pittsburgh, Pittsburgh, Pennsylvania

[2]School of Nursing, Columbia University, New York, New York

[3]Data Science Institute, Columbia University, New York, New York

[4]Department of Biomedical Informatics, Columbia University, New York, New York

[5]Department of Systems Biology, Columbia University, New York, New York

[6]Department of Medicine, Columbia University, New York, New York

[7]Institute for Genomic Medicine, Columbia University, New York, New York

[8]School of Nursing, University of California San Francisco, San Francisco, California

[9]College of Dental Medicine, Columbia University, New York, New York

Abstract

Data-driven characterization of symptom clusters in chronic conditions is essential for shared cluster detection and physiological mechanism discovery. This study aims to computationally describe symptom documentation from electronic nursing notes and compare symptom clusters among patients diagnosed with four chronic conditions–chronic obstructive pulmonary disease (COPD), heart failure, type 2 diabetes mellitus, and cancer. Nursing notes (N=504,395; 133,977 patients) were obtained for the 2016 calendar year. We used NimbleMiner, a natural language processing application, to identify the presence of 56 symptoms. We calculated symptom documentation prevalence by note and patient for the corpus. Then, we visually compared documentation for a subset of patients (N=22,657) diagnosed with COPD (n=3,339), heart failure (n=6,587), diabetes (n=12,139), and cancer (n=7,269) and conducted multiple correspondence analysis and hierarchical clustering to discover underlying groups of patients who have similar symptom profiles (i.e., symptom clusters) for each condition. As expected, pain was the most frequently documented symptom. All conditions had a group of patients characterized by no symptoms. Shared clusters included cardiovascular symptoms for heart failure and diabetes; pain and other symptoms for COPD, diabetes, and cancer; and a newly-identified cognitive and neurological symptom cluster for heart failure, diabetes, and cancer. Cancer (gastrointestinal

Correspondence Theresa A. Koleck, School of Nursing, University of Pittsburgh, 440 Victoria Building, 3500 Victoria Street, Pittsburgh, PA 15261, tak150@pitt.edu.

Conflict of Interest
The authors declare no conflicts of interest.

symptoms and fatigue) and COPD (mental health symptoms) each contained a unique cluster. In summary, we report both shared and distinct, as well as established and novel, symptom clusters across chronic conditions. Findings support the use of electronic health record-derived notes and NLP methods to study symptoms and symptom clusters to advance symptom science.

Keywords

chronic conditions; natural language processing; nursing informatics; signs and symptoms; symptom clusters

---

Approximately half of American adults ≥18 years of age live with one or more chronic conditions such as chronic obstructive pulmonary disease (COPD), heart failure, type 2 diabetes mellitus (T2DM), and/or cancer (Ward et al., 2014). Many patients with chronic conditions experience multiple co-occurring symptoms, defined as subjective indications of disease, either related to a chronic condition and/or its treatment. Symptoms are key concepts of nursing interest and include anxiety, depressed mood, disturbed sleep, fatigue, nausea, pain, pruritus, and shortness of breath. Two or more co-occurring, related symptoms are known as symptom clusters (Kim et al., 2005; Miaskowski et al., 2017). Symptom clusters are challenging to manage and burden both the patient and healthcare system (Armstrong, 2003; Kwekkeboom, 2016).

While evidence suggests that patients with all types of chronic conditions are burdened by symptom clusters, the complex relationships within and between symptoms both in a single chronic condition and among chronic conditions (i.e., shared symptom clusters) remain understudied and poorly understood. The majority of symptom cluster research conducted to date is limited to cross-sectional studies of patients diagnosed with cancer, using predetermined symptom inventories. A review of the chronic condition symptom cluster literature conducted by Miaskowski and colleagues (2017) identified 158 studies that evaluated symptom clusters primarily related to cancer or its treatment. The most common symptom clusters identified were a combination of fatigue, pain, depression, and sleep disturbance; nausea and vomiting; and anxiety and depression.

Much less work has been done with other chronic conditions (e.g., heart failure). Research on symptom clusters associated with heart failure is summarized in three review articles published in the past 10 years (DeVon et al., 2017; Herr et al., 2014; Yu et al., 2018). All three reviews highlight an emotional- or depression-related component to heart failure symptom clusters. Specifically, in the systematic review of symptom clusters in heart failure (n=6 studies) by DeVon et al. (2017), symptom clusters frequently had both a physical and emotional/cognitive component. The authors attribute this finding to the known functional decline and cognitive impairment associated with heart failure. Similarly, Yu and colleagues' (2018) systematic review of symptom clusters in advanced heart failure (n=10 studies) identified three common clusters: distress (dyspnea and emotional distress), deconditioning (fatigue, nausea, reduced appetite, and drowsiness), and discomfort (pain). Herr and colleagues' (2014) examination of relationships among heart failure symptoms (n=34 studies) found that depression was related to several other symptoms, including fatigue, anxiety, sleep, and pain.

Even less is known about symptom clusters in patients with COPD and T2DM. A 2018 systematic review of five studies by Jenkins et al. (2019) revealed both psychological and respiratory-related symptom clusters in patients diagnosed with COPD. Severity of combinations of dyspnea, anxiety, depression, and/or fatigue were found to be associated with declines in physical functioning (Park et al., 2013; Park et al., 2012) and increased healthcare use and mortality (Park & Larson, 2014). It is important to note that the specific instrument selection and the number of symptoms assessed was highly variable among studies and resulted in different types of clusters (Jenkins et al., 2019). Only two studies have evaluated symptom clusters in patients with T2DM. The first study was a large population survey of Australian adults that focused on gastrointestinal symptoms (Hammer et al., 2003). The second study investigated symptom clusters in Mexican American adults with T2DM (García et al., 2019). This study identified three clusters of multiple symptoms: (a) sweating, intense thirstiness, urinating more than usual, cravings, dry mouth, tiredness, numbness or tingling, problems sleeping, and dry skin; (b) itching skin, weight gain, discolored skin areas, noise or light sensitivity, trembling, flushing, sexual problems, burning in feet or legs, and hair loss; and (c) constipation, dizziness, indigestion or nausea, sadness, irritability, easily angry, nervous, fidgety, weakness, anxiousness, headache, trouble concentrating, and memory loss. Additional studies have examined relationships between predefined psychological or depressive symptom clusters and outcomes of patients with T2DM, including quality of life (Li et al., 2019), cardiovascular hospitalization (Nefs et al., 2015), and insulin resistance (Khambaty et al., 2014).

Electronic health records (EHRs), especially text-based nursing notes, may serve as a rich source of “real world” symptom information and address current barriers to symptom cluster research, especially lack of access to cohorts of patients diagnosed with chronic conditions and the potential influence of predetermined symptom inventories on symptom cluster results. Despite limitations (e.g. completeness, concordance; Weiskopf et al., 2013), EHRs represent the most comprehensive, longitudinal, population-wide dataset that we have to study symptom clusters. “Real world” practice and utilization data have been used in other domains (e.g., pharmaceutical research) to obtain clinically meaningful results (Bent-Ennakhil et al., 2017; Divino et al., 2014; Feher et al., 2017; Gradman et al., 2013; Johnson et al., 2017; Kolaczynski et al., 2016; Lunacsek et al., 2016; McGovern et al., 2016). Natural language processing (NLP), which is “any computer-based algorithm that handles, augments, and transforms natural language so that it can be represented for computation” (Yim et al., 2016), can be used to study symptom information from text-based nursing notes (Koleck et al. 2019). Our team recently developed a method that combines standardized vocabularies, clinical expertise, and NLP to generate comprehensive symptom vocabularies and identify and extract symptom information in EHR notes in an accurate and scalable manner (Koleck et al., 2020).

Guided by the National Institutes of Health Symptom Science Model (NIH-SSM; Cashion et al., 2016), data-driven characterization of symptom clusters in a variety of chronic conditions is an essential first step toward chronic condition specific analyses; detection of shared symptom clusters and mechanisms; clinical, genetic, and biomarker discovery; and future care delivery. Clinically, data-driven characterization of symptom clusters is important to nursing practice as it informs thorough symptom assessment and planning and evaluation of targeted symptom management nursing interventions (Kwekkeboom, 2016; Tabudlo, 2021). Therefore, the aim of this study was to characterize documentation of 56 naturally documented symptoms in thousands of EHR text-based nursing notes from inpatient, outpatient, and emergency department (ED) encounters and compare symptom documentation and symptom clusters among a subset of patients diagnosed with four common chronic conditions (i.e., COPD, heart failure, T2DM, and cancer). We extracted symptom information from nursing notes using NLP.

Methods

Nursing note corpus

The corpus for this study comprised N=504,395 EHR nursing notes from 133,977 distinct patients. Nursing notes were obtained from the Columbia University Irving Medical Center Clinical Data Warehouse between January 1, 2016, and December 31, 2016. Nursing notes were from inpatient, outpatient, and ED encounters and had various format structures from entirely free-text to predetermined-field forms. This study was approved by the Columbia University Irving Medical Center Institutional Review Board.

We included 20 different nursing note types. The counts and distribution of the note types are detailed in Table 1. We excluded pediatric (e.g., nursing pediatric admission history), obstetric (OB; e.g., nursing OB triage history, nursing neonatal patient history), and operating room (OR) notes (e.g., OR nursing progress note) as well as notes that occurred at a frequency of n<100 (e.g., psychiatric nursing restraint note) or were determined through study team comprehensive manual review to contain minimal symptom-related content (e.g., transfer sending note, transfusion nursing note, bronchoscopy nursing note) or focused on patient education (e.g., CHF nurse phone visit, diabetes nurse education note).

Symptoms

This study focused on identifying 56 diverse symptoms in the nursing notes. Symptoms included: abdominal bloating, abdominal pain, agitation, altered sense of smell, altered sense of taste, anger, anxiety, chest pain, chills, choking sensation, confusion, constipation, cough, decreased appetite, decreased libido, decreased thirst, defecation pain/urgency, depressed mood, difficult painful swallowing, difficulty coping, disturbed sleep, dizziness, dry mouth, dyspareunia, elevated mood, emotional lability, fatigue, hallucinations, headache, hearing impairment, heartburn, hot flashes, impaired attention/executive function, impaired cognition, impaired concentration, impaired memory, increased appetite, increased thirst, malaise, muscle cramp, nausea, pain, palpitations, paresthesia/numbness, pruritus, shortness of breath, suicidal ideation, swelling, tactile perception, tinnitus, unsteadiness, upset stomach, urinary pain/urgency/hesitancy, vaginal dryness/irritation, vision changes, and weakness. Selection of symptoms for inclusion was informed by SNOMED-CT clinical findings medical terminology and study team nurse clinician scientist expertise.

Natural language processing of symptoms

Our team previously generated and validated comprehensive symptom vocabularies using a method that combined standardized vocabularies from the Unified Medical Language

System, clinical expertise of nurse clinician scientists, and NLP (Koleck et al., 2020). Based on this earlier work, we used NimbleMiner, an open-source NLP RStudio application (https://github.com/mtopaz/NimbleMiner), to identify the presence of symptoms in nursing notes at the note level. NimbleMiner was also used to distinguish negated instances of symptom occurrence (e.g., no fatigue, denies dizziness).

After labeling notes with NimbleMiner, we manually reviewed symptoms with an abnormally high frequency of occurrence (i.e., an increase of more than 50% between ranked symptoms) in note types with a predetermined-field form structure. We completed this review because note formatting can lead to labeling errors within the NimbleMiner system. While some of the high occurrence symptoms were true positives (e.g., shortness of breath as an early warning of respiratory distress in rapid response team nursing action notes or mild fatigue over baseline in ambulatory hematology/oncology nursing assessments), we identified systematic false positive labeling errors. Labeling errors were mainly related to negation and fell into one of three categories: (a) negation term was too many words away from the symptom of interest (e.g., no complaint of pain, diarrhea, constipation); (b) negation term occurred after the symptom of interest (e.g., cough no, edema none, does the patient have pain now no); or (c) a numeric indicator of negation was used (e.g., impaired cognition 0, pain scale 0). We created rules to correct instances of systematic false positive labeling in a post hoc manner.

Prevalence of symptom documentation in nursing notes

We calculated the prevalence of positive documentation (i.e., documentation indicated occurrence of the symptom) of each symptom by dividing the number of nursing notes with a symptom documented by the total number of notes in the entire corpus. For example, the prevalence of nausea (5.63%) within notes was calculated by dividing the number of notes with nausea positively documented (n=28,397) by the total number of notes (n=504,395). We evaluated the least and most commonly documented symptoms. We completed this same process at the patient level as well, dividing the number of patients with a symptom documented one or more times (i.e., presence or absence) in a nursing note in 2016 by the total number of patients.

Exploring symptom clusters in patients diagnosed with a chronic condition

We explored symptom clusters in a subset of patients diagnosed with a chronic condition. We used the presence of two or more International Statistical Classification of Diseases and Related Health Problems, 10th revision between January 1, 2016, and December 31, 2016 in the Columbia University Irving Medical Center Clinical Data Warehouse to identify adult (≥ 18 years of age at diagnosis) patients diagnosed with COPD (*J41-J44*), heart failure (*I50*), T2DM (*E11*), and cancer (*C00-C96*). The number of patients (N=22,647) with one or more nursing notes for each chronic condition was as follows: COPD - n=3,399, heart failure - n=6,587, T2DM - n=12,139, and cancer - n=7,269. Chronic condition sample membership was not mutually exclusive (Figure 1). We aggregated symptom data extracted from nursing notes in a binary fashion (i.e., presence or absence) at the patient level for the calendar year. A symptom concept was considered absent for a patient if it was documented but negated or not documented.

Heatmap—For each chronic condition, we calculated the percentage of patients with a symptom documented in one or more nursing notes. We created a heatmap using conditional formatting in Microsoft® Excel for Mac Version 15.23. After grouping symptoms by clinical category (e.g., cardiovascular and respiratory, mental health, gastrointestinal), we visually compared symptom documentation across the four chronic conditions.

Multiple correspondence analysis and hierarchical clustering—We performed multiple correspondence analysis (MCA) and hierarchical clustering for each chronic condition to discover underlying groups of patients who have similar symptom profiles, i.e., symptom clusters (Skerman et al., 2009). We used the FactoMineR and Factoshiny (http://factominer.free.fr/index.html) packages in R Version 3.5.1 (Lê & Husson, 2008). Symptoms with a prevalence of ≥5% for a chronic condition were included in analyses. First, we used MCA, a principal component analysis method for categorical (e.g., anxiety - yes, anxiety - no) variables to perform data preprocessing (Lê & Husson, 2008; Husson, 2016b; Husson, 2020b). Multiple correspondence analysis transforms categorical variables into a few continuous components that can be used as the input for cluster analysis (Kassambara, 2017; Lê & Husson, 2008; Husson, 2016a; Husson, 2020a). Multiple correspondence analysis components are ordered by variance in descending order. For each chronic condition, we retained the number of components that allowed us to preserve 90% of the variance. The distances used to build the clustering model were calculated using the retained components. The remaining components were considered noise.

Next, we built hierarchical clustering trees using Ward's agglomerative clustering method and Euclidean distance and cut the trees to define clusters based on gain/loss of inertia (Kassambara, 2017; Lê & Husson, 2008; Husson, 2016a; Husson, 2020a). The output of the clustering is groups of patients with similar symptom profiles. While each patient is assigned to one group, a symptom variable category (i.e., symptom - yes, symptom - no) can characterize multiple groups. Each patient group/symptom profile represents a symptom cluster. To name the symptom clusters, we evaluated which symptom variable categories (i.e., symptom - yes, symptom - no) best characterized each group of patients with similar symptom profiles using (a) a chi-squared test between the symptom variables and the grouping, (b) the percentage of patients with a symptom variable category belonging to a grouping (e.g., 74% of patients with no unsteadiness belong to COPD profile 1), and (c) the percentage of patients in a grouping with a symptom variable category (e.g., 91% of patients in COPD profile 1 had no unsteadiness).

Results

Prevalence of symptom documentation in nursing notes

The prevalence of symptom documentation varied greatly by symptom. For each symptom, we present (Table 2) the percentage of nursing notes with a symptom documented out of the entire corpus of notes and the percentage of patients with a symptom documented one or more times in 2016 out of the total number of patients. Pain was by far the most frequently documented symptom in nursing notes, with approximately 30% of all notes having one or more positive instances of pain and 64% of patients having pain documented one or more times in 2016. Other commonly documented symptoms in nursing notes included swelling (by note - 6.27%, by patient - 15.81%), shortness of breath (by note - 5.63%, by patient - 12.79%), anxiety (by note - 5.61%, by patient - 8.97%), depressed mood (by note - 5.19%, by patient - 7.63%), and nausea (by note - 5.05%, by patient - 14.64%).

With the exceptions of confusion, impaired cognition, and pruritus, cognition- and perception-related symptoms were not documented frequently. Similarly, while some gastrointestinal symptoms were documented frequently (e.g., nausea, abdominal pain, constipation), others were rarely or never documented (e.g., increased appetite, decreased thirst, defecation pain/urgency). In addition, symptoms related to sexual dysfunction were documented very rarely (e.g., vaginal dryness/irritation, dyspareunia, decreased libido) in our nursing note set.

Symptom clusters in patients diagnosed with a chronic condition

Age, sex/gender, race, and ethnicity for patients diagnosed with chronic conditions are reported in Table 3. A visual comparison of symptom documentation across COPD, heart failure, T2DM, and cancer by clinical category is displayed as a heatmap of the percentage of patients with a symptom documented in a nursing note in 2016 (Figure 2). We observed both shared and distinct patterns of symptom documentation across the four chronic conditions. Pain was the most frequently documented symptom across all chronic conditions. However, the percentage of patients with pain documented varied from 74.73% for COPD to 56.17% for cancer. Gastrointestinal, cognition and perception, general, and genitourinary symptoms were documented at similar rates across the chronic conditions. In contrast, cardiovascular and respiratory symptoms (e.g., shortness of breath, cough, swelling, chest pain) were documented for a higher percentage of patients diagnosed with COPD and heart failure compared to T2DM and cancer. Another distinct pattern of symptom documentation was related to fatigue. Fatigue was documented at least two times more frequently in patients diagnosed with cancer than in those with other chronic conditions. Furthermore, we observed differences in mental health symptoms. Higher percentages of anxiety were documented in patients diagnosed with COPD and heart failure, and higher percentages of depressed mood were documented in patients with COPD. We also noted that symptoms appeared to be documented at lower percentages in patients with cancer overall compared to the other conditions.

Symptom clusters (i.e., groups of patients with similar symptom profiles) and the top five symptoms characterizing a cluster for each of the chronic conditions are displayed in Table 4. We further include all symptom variables that characterize a cluster and relevant statistics (e.g., percentage of patients with a symptom variable belonging to a cluster) in Supplemental Digital Content 1. A total of six symptom clusters (including a symptom profile characterized by no symptoms) were identified. Three symptom clusters were found for COPD and heart failure and four symptom clusters were found for T2DM and cancer. All conditions had a group of patients characterized by no symptoms. Heart failure, T2DM, and cancer included a cognitive and neurological symptom cluster. Impaired cognition and confusion were the symptoms that best characterized this group of patients in all cases. Pain (i.e., pain, chest pain, and/or abdominal pain) and other symptoms (e.g., shortness of breath, unsteadiness, weakness, nausea) was the second cluster shared by three conditions, i.e., COPD, T2DM, and cancer. Both heart failure and T2DM included a cardiovascular symptom cluster that was characterized by symptoms such as shortness of breath, chest pain, and dizziness. Cancer and COPD each contained a unique cluster. A distinct mental health symptom cluster was identified for patients with COPD. This cluster was characterized by suicidal ideation, depressed mood, agitation, and anxiety. A distinct gastrointestinal (i.e., decreased appetite, nausea, abdominal pain, constipation, abdominal bloating) and fatigue (i.e., fatigue, weakness, dizziness, disturbed sleep) symptom cluster was identified in patients with cancer.

Discussion

To our knowledge, this study is the first to use NLP to widely characterize symptom documentation in thousands of nursing notes from inpatient, outpatient, and ED encounters as well as compare symptoms and symptom clusters among large cohorts of patients diagnosed with four common chronic conditions (i.e., COPD, heart failure, T2DM, and cancer). We observed both shared and distinct patterns of symptom documentation in “real world” (i.e., not influenced by predetermined symptom inventories) nursing notes, as well as established and novel symptom clusters from the EHR-derived text-based symptom data, across the four chronic conditions. Overall, our findings support the use of EHR-derived text-based notes and NLP methods to study symptoms and symptom clusters to advance symptom science.

Pain was the most frequently documented symptom in nursing notes overall and within each of the chronic conditions. Furthermore, the presence of pain was a defining characteristic of a pain (i.e., pain, abdominal pain, and/or chest pain) and other symptoms (e.g., unsteadiness, weakness, shortness of breath, nausea) cluster for three out of the four chronic conditions (i.e., COPD, T2DM, and cancer). A number of factors likely contribute to high rates of pain documentation. First, unrelieved acute or chronic pain is among the most common reasons why patients seek medical care (Debono et al., 2013; Vogel et al., 2019). Second, pain is a nursing diagnosis, and nurses play a key role in pain assessment, management, monitoring, and documentation. Third, organizations, like the Joint Commission, have developed and enforce pain assessment and management standards for accredited systems (The Joint Commission, 2021). Gastrointestinal, cognition and perception, general, and genitourinary symptoms were also documented at similar rates across the chronic conditions.

All four chronic conditions shared a group of patients characterized by having no symptoms. This no symptom profile may represent a number of scenarios, including the nurse’s confirmation that the patient did not have any symptoms and/or that the symptom(s) was relieved following treatment. Alternatively, there may be a lack of: patient symptom reporting, patient interaction with the healthcare system, nursing assessment, and/or nursing documentation. It is possible that nurses may chart by exception in some circumstances, potentially biasing documentation towards unexpected symptoms rather than all symptoms. Nevertheless, this finding calls attention to the wide spectrum of symptom experiences associated with chronic conditions and their treatments. Chronic obstructive pulmonary disease (Lopez-Campos et al., 2013), heart failure (Snipelisky et al., 2019), T2DM (Prasad

& Groop, 2019; Udler, 2019), and cancer (Allison & Sledge, 2014) are rather heterogeneous clinical conditions. Additional symptom cluster studies using EHR-derived text-based data may be able to provide more detailed information on disease specific symptom clusters (Ahlqvist et al., 2020; Ahmad et al., 2014; Christensen et al., 2016; Hertzog et al., 2010; Li et al., 2019; Miaskowski et al., 2017; Nagamine et al., 2020; Park et al., 2013). Evaluation of different dimensions of the symptom experience (e.g., severity, frequency, and distress) may provide additional information on condition-specific symptom clusters, as well as symptom clusters that are common across multiple conditions (Miaskowski et al., 2017; Miravitlles & Ribera, 2017; Yu et al., 2018). Likewise, development of computational methods to capture symptom severity, duration, frequency, and distress from the nursing notes would be advantageous.

We noted distinct patterns of symptom documentation and symptom clusters. Compared to other conditions, higher percentages of anxiety were documented in patients diagnosed with COPD and heart failure as well as higher percentages of depressed mood in patients diagnosed with COPD. Chronic obstructive pulmonary disease was the only chronic condition that included a distinct mental health symptom (e.g., suicidal ideation, depressed mood, agitation, anxiety) cluster. This finding is consistent with the previous reports that found that mental health symptoms, including anxiety and depressed mood, occurred in combination with dyspnea and fatigue in patients with COPD (Park et al., 2012; Park et al., 2013).

While we did not find a specific mental health symptom cluster in patients diagnosed with heart failure, T2DM, or cancer (DeVon et al., 2017; Herr et al., 2014; Miaskowski et al., 2017; Yu et al., 2018), our analysis identified a novel cluster of cognitive (e.g., impaired cognition, confusion) and neurological symptoms (e.g., paresthesia/numbness, unsteadiness, weakness) common to heart failure, T2DM, and cancer. This cognitive and neurological cluster included mental health symptoms (i.e., depressed mood and anxiety), fatigue, and sleep disturbance (see Table 4 and Supplemental Digital Content 1). A related psychoneurologic symptom cluster of depressed mood, cognitive disturbance, fatigue, sleep disturbance, and pain was described previously in patients with cancer (Kim & Malone, 2019; Kim, Barsevick, Beck, & Dudley, 2012; Kim, Barsevick, Fang, & Miaskowski, 2012; Levkovich et al., 2018). However, this pre-specified symptom cluster did not include physical/functional neurological symptoms (e.g., paresthesia/numbness, unsteadiness, weakness). Similarly, the aforementioned mental health symptom cluster for patients with COPD was partially characterized by cognitive symptoms (e.g., confusion, impaired cognition) and fatigue, but largely lacks the physical/functional neurological symptoms.

Kim, Barsevick, Fang, & Miaskowski (2012) proposed that the underlying mechanisms for the psychoneurologic cluster in cancer patients included: activation of proinflammatory cytokines, alterations in the hypothalamic-pituitary-adrenal axis, and changes in the monoamine neurotransmission system. High levels of proinflammatory cytokines have been found in patients with cancer (Seruga et al., 2008), as well as in patients with COPD (Barnes, 2009), heart failure (Liu et al., 2014), and T2DM (Guest et al., 2008). Moreover, alterations in proinflammatory cytokines play a role in physical/functional neurological diseases, such as myopathy (De Paepe et al., 2009; Salomonsson & Lundberg, 2006; Szodoray et al., 2010) and neuropathy (Fregnan et al., 2012; Hung et al., 2017). Additional research is warranted to explore the presence, antecedents, and outcomes of physical/functional neurological symptoms within the context of or external to a psychoneurologic symptom cluster and elucidate shared physiological mechanisms underlying symptom development of this cluster across chronic conditions.

In addition, we report a number of expected differences in symptom documentation and symptom clusters among the chronic conditions. These expected findings confer validity to the use of EHR-derived text-based notes and NLP methods to study symptoms and symptom clusters in chronic conditions. For example, cardiovascular and respiratory symptoms (e.g., shortness of breath, cough, swelling, chest pain) were documented for a greater percentage of patients diagnosed with COPD and heart failure compared to T2DM and cancer. We anticipated these results based on the pathophysiologic changes associated with these conditions. Another distinct pattern of symptom documentation was related to fatigue. Fatigue was documented at least two times more frequently in patients with cancer compared to the other three conditions. Likewise, we observed a distinct gastrointestinal symptoms and fatigue cluster in cancer patients. Fatigue and gastrointestinal symptoms, like nausea and decreased appetite, are commonly experienced within the context of cancer and/or its treatment (Berger et al., 2015; Dev et al., 2017; Walsh et al., 2017). However, symptoms were documented with a lower frequency in cancer patients overall compared to other chronic conditions. This documentation pattern was particularly surprising as cancer is the chronic condition in which most symptom and symptom cluster research was historically conducted (Miaskowski et al., 2017). We hypothesize that cancer symptoms may be documented using structured symptom assessments, tools, or checklists (Cooley & Siefert, 2016; Kirkova et al., 2006). Alternatively, symptom prevalence may be diluted in the cancer sample due to our broad inclusion of all cancer types, including skin, which have highly variable clinical presentations and treatments as discussed previously.

Limitations

Our analysis had several limitations in addition to those already discussed. First, the system that we used to label notes, NimbleMiner, has limitations related to negation that required manual, post hoc mitigation. While we believe that we were able to remedy the majority of systematic false positives, labeling errors may have impacted results. Second, several cognition and perception (e.g., hallucinations, impaired memory, altered sense of taste), gastrointestinal (e.g., increased appetite, decreased thirst, defecation pain/urgency), and sexual dysfunction (e.g., vaginal dryness/irritation, dyspareunia, decreased libido) symptoms were rarely documented in our nursing note set. Some symptoms may actually occur at a lower prevalence while other may not be as frequently reported, assessed, and/or documented. Inclusion of notes authored by other types of clinicians (e.g., physicians, therapists) in addition to nurses and incorporation of structured data sources (e.g., checklists, laboratory test results, medications) may mitigate this limitation. Regardless, it is important to thoughtfully consider the chronic condition and research question when using symptom information extracted from nursing notes and the EHR in general. Third, we explored the presence of symptom clusters using binary (i.e., presence or absence) patient-level symptom data from a calendar year of interaction with the healthcare system. We did not account for the number of times a symptom was documented, notes a patient had, and encounters with the healthcare system or length of stay of inpatient encounters. While some symptoms from visits unrelated to the chronic condition may be included, it would be challenging to distinguish a completely unrelated acute (or other chronic) disease process from the condition of interest. Likewise, we did not control for patient comorbidities or other patient characteristics (e.g., sociodemographic features). Chronic condition sample membership in our study was not exclusive. While this inclusivity represents the real world, the impact of overlapping categorization and other comorbidities is unknown. Much more research is needed to evaluate the influence of comorbid occurrence of two or more of four chronic conditions featured in this study, as well as other comorbid conditions (e.g., stroke, kidney disease, arthritis), on shared and distinct groupings of patient symptom profiles. Finally, longitudinal analyses should be conducted to explore symptom cluster trajectories.

Conclusion and future directions

In summary, we used NLP methods to extract extensive symptom information from EHR-derived text-based nursing notes. We characterized symptom documentation in nursing notes and compared symptoms and symptom clusters for patients diagnosed with four common chronic conditions, namely: COPD, heart failure, T2DM, and cancer. We observed both shared and distinct patterns of symptom documentation and symptom clusters among patients diagnosed with these four chronic conditions. We were able to confirm expected or previous findings (e.g., high prevalence of pain documentation; gastrointestinal symptoms and fatigue symptom cluster for cancer patients) as well as identify a novel cluster of cognitive and neurological symptoms shared by heart failure, T2DM, and cancer. These findings support the use of EHR-derived text-based notes and NLP methods to study symptoms and symptom clusters, advancing the phenotypic characterization step of the NIH-SSM. Electronic health record-derived symptom data may help to overcome current barriers to symptom research, including lack of access to large cohorts of patients diagnosed with a chronic condition and the influence of predetermined symptom inventories. The investigation of symptom documentation patterns and symptom clusters in chronic conditions has the potential to elucidate underlying physiological mechanisms of shared symptom clusters and improve clinical management. Despite its limitations, the EHR offers nurse scientists the ability to connect symptom and symptom cluster information extracted from nursing notes with additional biological, behavioral, environmental, social, and lifestyle factors for thousands of patients diagnosed with one or more chronic conditions in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding Information

National Institutes of Health, National Institute of Nursing Research, Grant/Award Numbers: K99NR017651, R00NR017651, and P30NR016587.

References

Ahlqvist E, Prasad RB, & Groop L (2020). Subtypes of type 2 diabetes determined from clinical parameters. Diabetes, 69(10), 2086–93. 10.2337/dbi20-0001 [PubMed: 32843567]

Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Piña IL, Kitzman DW, Lee KL, O'Connor CM, & Felker MG (2014). Clinical implications of chronic heart failure phenotypes defined by cluster analysis. Journal of the American College of Cardiology, 64(17), 1765–74. 10.1016/j.jacc.2014.07.979 [PubMed: 25443696]

Allison KH, & Sledge GW (2014). Heterogeneity and cancer. Oncology (Williston Park, N.Y.), 28(9), 772–8.

Armstrong TS (2003). Symptoms experience: A concept analysis. Oncology Nursing Forum, 30(4), 601–6. 10.1188/03.ONF.601-606 [PubMed: 12861321]

Barnes PJ (2009). The cytokine network in chronic obstructive pulmonary disease. American Journal of Respiratory Cell and Molecular Biology, 41(6), 631–8. 10.1165/rcmb.2009-0220TR [PubMed: 19717810]

Bent-Ennakhil N, Coste F, Xie L, Aigbogun MS, Wang Y, Kariburyo F, Harty A, Baser O, Neumann P, & Fillit H (2017). A real-world analysis of treatment patterns for cholinesterase inhibitors and memantine among newly-diagnosed Alzheimer's disease patients. Neurology and Therapy, 6(1), 131–44. 10.1007/s40120-017-0067-7 [PubMed: 28508250]

Berger AM, Mooney K, Alvarez-Perez A, Breitbart WS, Carpenter KM, Cella D, Cleeland C, Dotan E, Eisenberger MA, Escalante CP, Jacobsen PB, Jankowski C, LeBlanc T, Ligibel JA, Loggers ET, Mandrell B, Murphy BA, Palesh O, Pirl WF, … National Comprehensive Cancer Network. Cancer-Related Fatigue, Version 2.2015. Journal of the National Comprehensive Cancer Network, 13(8), 1012–39. doi: 10.6004/jnccn.2015.0122 [PubMed: 26285247]

Cashion AK, Gill J, Hawes R, Henderson WA, & Saligan L (2016). NIH Symptom Science Model sheds light on patient symptoms. Nursing Outlook, 64(5), 499–506. 10.1016/j.outlook.2016.05.008 [PubMed: 27349632]

Christensen VL, Rustøen T, Cooper BA, Miaskowski C, Henriksen AH, Bentsen SB, & Holm AM (2016). Distinct symptom experiences in subgroups of patients with COPD. International Journal of Chronic Obstructive Pulmonary Disease, 11(1), 1801–9. 10.2147/COPD.S105299 [PubMed: 27536092]

Cooley ME, & Siefert ML (2016). Assessment of multiple co-occurring cancer symptoms in the clinical setting. Seminars in Oncology Nursing, 32(4), 361–72. 10.1016/j.soncn.2016.08.003 [PubMed: 27776830]

De Paepe B, Creus KK, & De Bleecker JL (2009). Role of cytokines and chemokines in idiopathic inflammatory myopathies. Current Opinion in Rheumatology, 21(6), 610–6. 10.1097/BOR.0b013e3283317b31 [PubMed: 19726994]

Debono DJ, Hoeksema LJ, & Hobbs RD (2013). Caring for patients with chronic pain: Pearls and pitfalls. The Journal of the American Osteopathic Association, 113(8), 620–7. 10.7556/jaoa.2013.023 [PubMed: 23918913]

Dev R, Wong A, Hui D, & Bruera E (2017). The evolving approach to management of cancer cachexia. Oncology (Williston Park, N.Y.), 31(1), 23–32.

DeVon HA, Vuckovic K, Ryan CJ, Barnason S, Zerwic JJ, Pozehl B, Schulz P, Seo Y, & Zimmerman L (2017). Systematic review of symptom clusters in cardiovascular disease. European Journal of Cardiovascular Nursing, 16(1), 6–17. 10.1177/1474515116642594 [PubMed: 27034451]

Divino V, DeKoven M, Hallinan S, Varol N, Wirta SB, Lee WC, & Reaney M (2014). Glucagon-like Peptide-1 receptor agonist treatment patterns among type 2 diabetes patients in six European countries. Diabetes Therapy: Research, Treatment and Education of Diabetes and Related Disorders, 5(2), 499–520. 10.1007/s13300-014-0087-6

Feher M, Vega-Hernandez G, Mocevic E, Buysse B, Myland M, Power GS, Husemoen LLN, Kim J, & Witte DR (2017). Effectiveness of liraglutide and lixisenatide in the treatment of type 2 diabetes: real-world evidence from The Health Improvement Network (THIN) database in the United Kingdom. Diabetes Therapy: Research, Treatment and Education of Diabetes and Related Disorders, 8(2), 417–31. 10.1007/s13300-017-0241-z

Fregnan F, Muratori L, Simões AR, Giacobini-Robecchi MG, & Raimondo S (2012). Role of inflammatory cytokines in peripheral nerve injury. Neural Regeneration Research, 7(29), 2259–66. 10.3969/j.issn.1673-5374.2012.29.003 [PubMed: 25538747]

García AA, Bose E, Zuñiga JA, & Zhang W (2019). Mexican Americans' diabetes symptom prevalence, burden, and clusters. Applied Nursing Research, 46, 37–42. 10.1016/j.apnr.2019.02.002 [PubMed: 30853074]

Gradman AH, Parisé H, Lefebvre P, Falvey H, Lafeuille M, & Duh MS (2013). Initial combination therapy reduces the risk of cardiovascular events in hypertensive patients: a matched cohort study. Hypertension, 61(2), 309–18. 10.1161/HYPERTENSIONAHA.112.201566 [PubMed: 23184383]

Guest CB, Park MJ, Johnson DR, & Freund GG (2008). The implication of proinflammatory cytokines in type 2 diabetes. Frontiers in Bioscience: A Journal and Virtual Library, 13(13), 5187–94. 10.2741/3074 [PubMed: 18508580]

Hammer J, Howell S, Bytzer P, Horowitz M, & Talley NJ (2003). Symptom clustering in subjects with and without diabetes mellitus: A population-based study of 15,000 Australian adults. The American Journal of Gastroenterology, 98(2), 391–8. 10.1111/j.1572-0241.2003.07236.x [PubMed: 12591060]

Herr JK, Salyer J, Lyon DE, Goodloe L, Schubert C, & Clement DG (2014). Heart failure symptom relationships: A systematic review. The Journal of Cardiovascular Nursing, 29(5), 416–22. 10.1097/JCN.0b013e31829b675e [PubMed: 23839572]

Hertzog MA, Pozehl B, & Duncan K (2010). Cluster analysis of symptom occurrence to identify subgroups of heart failure patients: A pilot study. The Journal of Cardiovascular Nursing, 25(4), 273–83. 10.1097/JCN.0b013e3181cfbb6c [PubMed: 20539162]

Hung AL, Lim M, & Doshi TL (2017). Targeting cytokines for treatment of neuropathic pain. Scandinavian Journal of Pain, 17(1), 287–93. 10.1016/j.sjpain.2017.08.002 [PubMed: 29229214]

Husson F [Francois Husson]. (2016a, 12 16). Clustering with FactoMineR [Video]. YouTube. https://www.youtube.com/watch?v=EJqYTDTJJug.

Husson F [Francois Husson]. (2016b, 11 30). Multiple correspondence analysis with FactoMineR [Video]. YouTube. https://www.youtube.com/watch?v=Sl1-UYD6iac&t=8s

Husson F [Francois Husson]. (2020a, 3 3). Clustering with R (FactorMineR & Factoshiny) [Video]. YouTube. https://www.youtube.com/watch?v=pmFCosCzRN8&t=129s

Husson F [Francois Husson]. (2020b, 2 27). Tutorial on MCA - multiple correspondence analysis - with R (Factoshiny & FactoMineR) [Video]. YouTube. https://www.youtube.com/watch?v=0z30RDikYaw&t=75s

Jenkins BA, Athilingam P, & Jenkins RA (2019). Symptom clusters in chronic obstructive pulmonary disease: a systematic review. Applied Nursing Research, 45, 23–9. 10.1016/j.apnr.2018.11.003 [PubMed: 30683247]

Johnson JF, Parsa R, & Bailey RA (2017). Real-world clinical outcomes among patients with type 2 diabetes receiving canagliflozin at a specialty diabetes clinic: Subgroup analysis by baseline HbA1c and age. Clinical Therapeutics, 39(6), 1123–31. 10.1016/j.clinthera.2017.04.006 [PubMed: 28495028]

Kassambara A (2017). Articles - principal component methods in R: practical guide. HCPC - hierarchical clustering on principal components: essentials. http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/117-hcpc-hierarchical-clustering-on-principal-components-essentials/

Khambaty T, Stewart JC, Muldoon MF, & Kamarck TW (2014). Depressive symptom clusters as predictors of 6-year increases in insulin resistance: Data from the Pittsburgh Healthy Heart Project. Psychosomatic Medicine, 76(5), 363–9. 10.1097/PSY.0000000000000063 [PubMed: 24846000]

Kim H, & Malone PS (2019). Roles of biological and psychosocial factors in experiencing a psychoneurological symptom cluster in cancer patients. European Journal of Oncology Nursing, 42(4), 97–102. 10.1016/j.ejon.2019.08.005 [PubMed: 31479847]

Kim H, Barsevick AM, Beck SL, & Dudley W (2012). Clinical subgroups of a psychoneurologic symptom cluster in women receiving treatment for breast cancer: a secondary analysis. Oncology Nursing Forum, 39(1), E20–30. 10.1188/12.ONF.E20-E30 [PubMed: 22201665]

Kim H, Barsevick AM, Fang CY, & Miaskowski C (2012). Common biological pathways underlying the psychoneurological symptom cluster in cancer patients. Cancer Nursing, 35(6), E1–20. 10.1097/NCC.0b013e318233a811

Kim H, McGuire DB, Tulman L, & Barsevick AM (2005). Symptom clusters: concept analysis and clinical implications for cancer nursing. Cancer Nursing, 28(4), 270–82. 10.1097/00002820-200507000-00005 [PubMed: 16046888]

Kirkova J, Davis MP, Walsh D, Tiernan E, O'Leary N, LeGrand SB, Lagman RL, & Russell KM (2006). Cancer symptom assessment instruments: a systematic review. Journal of Clinical Oncology, 24(9), 1459–73. 10.1200/JCO.2005.02.8332 [PubMed: 16549841]

Kolaczynski WM, Hankins M, Ong SH, Richter H, Clemens A, & Toussi M (2016). Microvascular outcomes in patients with type 2 diabetes treated with vildagliptin vs. sulfonylurea: A retrospective study using German electronic medical records. Diabetes Therapy: Research, Treatment and Education of Diabetes and Related Disorders, 7(3), 483–96. 10.1007/s13300-016-0177-8

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(4), 364–379. 10.1093/jamia/ocy173 [PubMed: 30726935]

Koleck TA, Tatonetti NP, Bakken S, Mitha S, Henderson MM, George M, Miaskowski C, Smaldone A, & Topaz M (2020). Identifying symptom information in clinical notes using natural language processing. Nursing Research. Advance online publication. 10.1097/NNR.0000000000000488

Kwekkeboom KL (2016). Cancer symptom cluster management. Seminars in Oncology Nursing, 32(4), 373–82. 10.1016/j.soncn.2016.08.004 [PubMed: 27789073]

Lê S, Josse J & Husson F (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 25(1), 1–18.

Levkovich I, Cohen M, Alon S, Kuchuk I, Nissenbaum B, Evron E, Pollack S, & Fried G (2018). Symptom cluster of emotional distress, fatigue and cognitive difficulties among young and older breast cancer survivors: The mediating role of subjective stress. Journal of Geriatric Oncology, 9(5), 469–75. 10.1016/j.jgo.2018.05.002 [PubMed: 29759914]

Li H, Ji M, Scott P, & Dunbar-Jacob JM (2019). The effect of symptom clusters on quality of life among patients with type 2 diabetes. The Diabetes Educator, 45(3), 287–94. 10.1177/0145721719837902 [PubMed: 30873908]

Liu M, Chen J, Huang D, Ke J, & Wu W (2014). A meta-analysis of proinflammatory cytokines in chronic heart failure. Heart Asia, 6(1), 130–6. 10.1136/heartasia-2013-010484 [PubMed: 27326188]

Lopez-Campos JL, Calero C, & Quintana-Gallego E (2013). Symptom variability in COPD: A narrative review. International Journal of Chronic Obstructive Pulmonary Disease, 8, 231–8. 10.2147/COPD.S42866 [PubMed: 23687444]

Lunacsek OE, Ravelo A, Coutinho AD, Hazard SJ, Green MR, Willey J, Eaddy M, & Goertz H (2016). First-line treatment with bevacizumab and platinum doublet combination in non-squamous non-small cell lung cancer: A retrospective cohort study in US oncology community practices. Drugs - Real World Outcomes, 3(3), 333–43. 10.1007/s40801-016-0090-5 [PubMed: 27747837]

McGovern A, Hinton W, Correa A, Munro N, Whyte M, & de Lusignan S (2016). Real-world evidence studies into treatment adherence, thresholds for intervention and disparities in treatment in people with type 2 diabetes in the UK. BMJ Open, 6(11), e012801. 10.1136/bmjopen-2016-012801

Miaskowski C, Barsevick A, Berger A, Casagrande R, Grady PA, Jacobsen P, Kutner J, Patrick D, Zimmerman L, Xiao C, Matocha M, & Marden S (2017). Advancing symptom science through symptom cluster research: Expert panel proceedings and recommendations. Journal of the National Cancer Institute, 109(4), djw253. 10.1093/jnci/djw253

Miravitlles M, & Ribera A (2017). Understanding the impact of symptoms on the burden of COPD. Respiratory Research, 18(1), 67. 10.1186/s12931-017-0548-3 [PubMed: 28431503]

Nagamine T, Gillette B, Pakhomov A, Kahoun J, Mayer H, Burghaus R, Lippert J, & Saxena M (2020). Multiscale classification of heart failure phenotypes by unsupervised clustering of unstructured electronic medical record data. Scientific Reports, 10(1), 21340. 10.1038/s41598-020-77286-6 [PubMed: 33288774]

Nefs G, Pop VJM, Denollet J, & Pouwer F (2015). Depressive symptom clusters differentially predict cardiovascular hospitalization in people with type 2 diabetes. Psychosomatics, 56(6), 662–73. 10.1016/j.psym.2015.06.006 [PubMed: 26481961]

Park SK, & Larson JL (2014). Symptom cluster, healthcare use and mortality in patients with severe chronic obstructive pulmonary disease. Journal of Clinical Nursing, 23(17–18), 2658–71. 10.1111/jocn.12526 [PubMed: 24460846]

Park SK, Meldrum CA, & Larson JL (2013). Subgroup analysis of symptoms and their effect on functioning, exercise capacity, and physical activity in patients with severe chronic obstructive pulmonary disease. Heart & Lung, 42(6), 465–72. 10.1016/j.hrtlng.2013.08.008 [PubMed: 24054947]

Park SK, Stotts NA, Douglas MK, Donesky-Cuenco D, & Carrieri-Kohlman V (2012). Symptoms and functional performance in Korean immigrants with asthma or chronic obstructive pulmonary disease. Heart & Lung, 41(3), 226–37. 10.1016/j.hrtlng.2011.09.014 [PubMed: 22054719]

Prasad RB, & Groop L (2019). Precision medicine in type 2 diabetes. Journal of Internal Medicine, 285(1), 40–8. 10.1111/joim.12859 [PubMed: 30403316]

Salomonsson S, & Lundberg IE (2006). Cytokines in idiopathic inflammatory myopathies. Autoimmunity, 39(3), 177–90. 10.1080/08916930600622256 [PubMed: 16769651]

Seruga B, Zhang H, Bernstein LJ, & Tannock IF (2008). Cytokines and their relationship to the symptoms and outcome of cancer. Nature Reviews. Cancer, 8(11), 887–99. 10.1038/nrc2507 [PubMed: 18846100]

Skerman HM, Yates PM, & Battistutta D (2009). Multivariate methods to identify cancer-related symptom clusters. Research in Nursing & Health, 32(3), 345–360. 10.1002/nur.20323 [PubMed: 19274688]

Snipelisky D, Chaudhry S, & Stewart GC (2019). The many faces of heart failure. Cardiac Electrophysiology Clinics, 11(1), 11–20. 10.1016/j.ccep.2018.11.001 [PubMed: 30717842]

Szodoray P, Alex P, Knowlton N, Centola M, Dozmorov I, Csipo I, Nagy AT, Constantin T, Ponyi A, Nakken B, & Danko K (2010). Idiopathic inflammatory myopathies, signified by distinctive peripheral cytokines, chemokines and the TNF family members B-cell activating factor and a proliferation inducing ligand. Rheumatology (Oxford, England), 49(10), 1867–77. 10.1093/rheumatology/keq151

Tabudlo JB (2021). A proposed theory of symptom cluster management. Belitung Nursing Journal, 7(2), 78–87. 10.33546/bnj.1359

The Joint Commission. (2021). Pain management standards for accredited organizations. https://www.jointcommission.org/resources/patient-safety-topics/pain-management-standards-for-accredited-organizations/

Udler MS (2019). Type 2 diabetes: Multiple genes, multiple diseases. Current Diabetes Reports, 19(8), 55–9. 10.1007/s11892-019-1169-7 [PubMed: 31292748]

Vogel JA, Rising KL, Jones J, Bowden ML, Ginde AA, & Havranek EP (2019). Reasons patients choose the emergency department over primary care: A qualitative metasynthesis. Journal of General Internal Medicine, 34(11), 2610–9. 10.1007/s11606-019-05128-x [PubMed: 31428988]

Walsh D, Davis M, Ripamonti C, Bruera E, Davies A, & Molassiotis A (2017). 2016 updated MASCC/ESMO consensus recommendations: Management of nausea and vomiting in advanced cancer. Supportive Care in Cancer, 25(1), 333–40. 10.1007/s00520-016-3371-3 [PubMed: 27534961]

Ward BW, Schiller JS, & Goodman RA (2014). Multiple chronic conditions among US adults: A 2012 update. Preventing Chronic Disease, 11, E62. 10.5888/pcd11.130389 [PubMed: 24742395]

Weiskopf NG, Hripcsak G, Swaminathan S, & Weng C (2013). Defining and measuring completeness of electronic health records for secondary use. Journal of Biomedical Informatics, 46(5), 830–6. 10.1016/j.jbi.2013.06.010 [PubMed: 23820016]

Yim WW, Yetisgen M, Harris WP, & Kwan SW (2016). Natural language processing in oncology: A review. JAMA Oncology, 2(6), 797–804. 10.1001/jamaoncol.2016.0213 [PubMed: 27124593]

Yu DS, Li PW, & Chong SO (2018). Symptom cluster among patients with advanced heart failure: A review of its manifestations and impacts on health outcomes. Current Opinion in Supportive and Palliative Care, 12(1), 16–24. 10.1097/SPC.0000000000000316 [PubMed: 29176333]

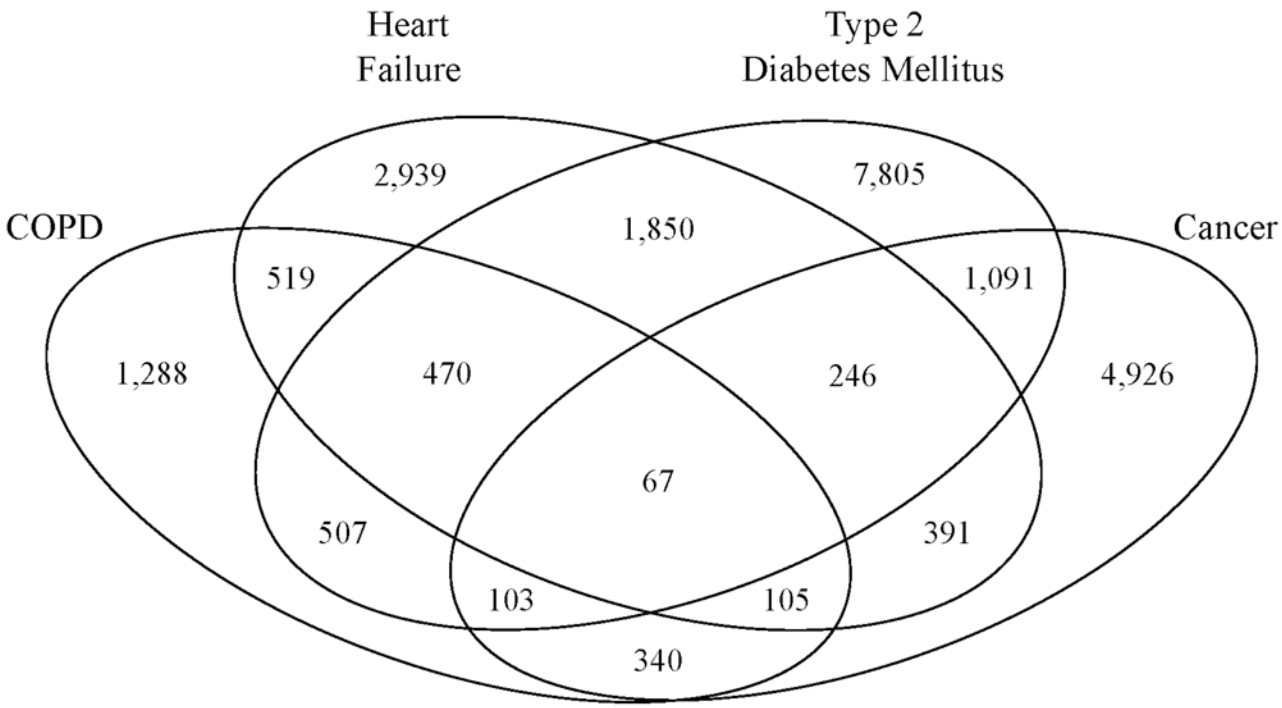

Figure 1. Overlap in Patients Diagnosed with Chronic Conditions of Interest (N=22,647)
Note. COPD = chronic obstructive pulmonary disease.

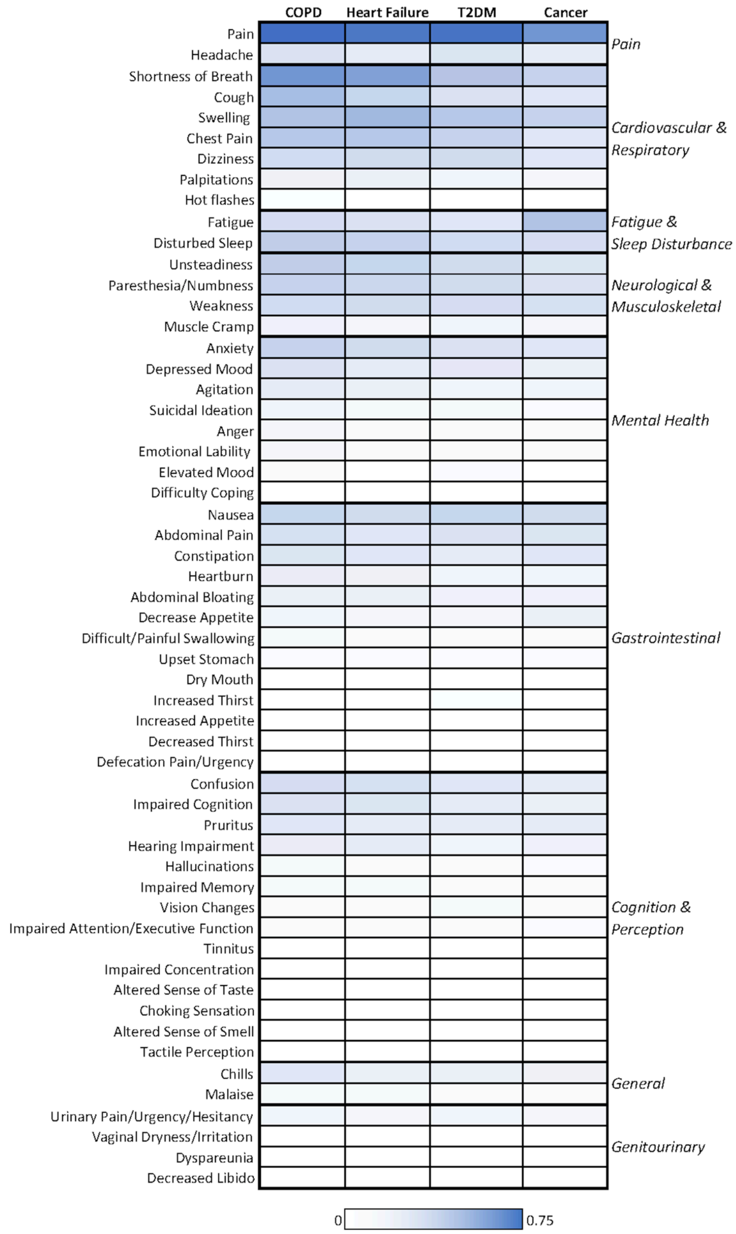

Figure 2. Percentage of Patients with a Symptom Documented in Nursing Notes Organized by Chronic Condition and Symptom Clinical Category

Note. COPD = chronic obstructive pulmonary disease. T2DM = type 2 diabetes mellitus.

Table 1

Nursing Note Types (N=504,395)

Note Type	n	%
Miscellaneous inpatient nursing note[a]	213,062	42.2
ED nursing assessment note	150,372	29.8
Nursing adult admission history	36,738	7.3
Ambulatory nursing note	29,980	5.9
Psychiatric nursing progress note	23,190	4.6
Ambulatory hematology/oncology nursing assessment	22,614	4.5
Psychiatric nursing note	9,493	1.9
ED psychiatric RN discharge note	4,229	0.8
ED psychiatric nursing assessment	4,225	0.8
Wound, ostomy, and continence nursing note	2,148	0.4
Gastrointestinal nursing note	1,739	0.3
Ambulatory Associates in Internal Medicine Practice nursing note	1,672	0.3
Psychosocial nursing note	1,151	0.2
Psychiatric nursing admission assessment	997	0.2
Rapid response team nursing action	922	0.2
Cardiovascular nursing note	515	0.1
Respiratory nursing note	408	0.1
Psychiatric clinical nurse specialist note	359	0.1
Genital urinary nursing note	337	0.1
RN initiated observation note	244	<0.1

[a]Miscellaneous inpatient nursing note encompass notes written for a variety of reasons in the inpatient setting. Examples include: *Pt* [patient] *complaining of nausea, had episode of vomiting. Zofran given.; Pt complaining of chest pain. Vital signs taken. MD aware.; Pt received sitting on bed. Pt denies headache, pain, and/or blurry vision. Pt denies any distress.*

Table 2

Prevalence of Documented Symptom Occurrence by Note and Patient

By note (*N*=504,395 notes)		By patient (*N*=133,977 patients)	
Symptom	**%**	**Symptom**	**%**
Pain	29.99	Pain	63.55
Swelling	6.27	Swelling	15.81
Nausea	5.63	Shortness of Breath	14.64
Shortness of Breath	5.61	Anxiety	12.79
Abdominal Pain	5.19	Depressed Mood	10.60
Chest Pain	5.05	Nausea	10.05
Headache	4.46	Fatigue	10.03
Cough	3.77	Chest Pain	9.15
Paresthesia/Numbness	3.75	Abdominal Pain	9.08
Anxiety	3.56	Pruritus	8.97
Dizziness	3.36	Paresthesia/Numbness	8.64
Depressed Mood	3.29	Cough	7.36
Disturbed Sleep	3.27	Headache	6.83
Pruritus	3.20	Dizziness	6.59
Unsteadiness	2.69	Disturbed Sleep	6.24
Fatigue	2.48	Unsteadiness	5.83
Weakness	2.25	Agitation	5.24
Muscle Cramp	2.08	Confusion	4.54
Chills	1.96	Elevated Mood	4.24
Confusion	1.95	Weakness	4.21
Impaired Cognition	1.78	Emotional Lability	3.98
Agitation	1.75	Impaired Cognition	3.35
Urinary Pain/Urgency/Hesitancy	1.47	Suicidal Ideation	3.06
Constipation	1.46	Muscle Cramp	3.03
Abdominal Bloating	1.28	Abdominal Bloating	2.74
Palpitations	1.22	Chills	2.72
Suicidal Ideation	1.10	Constipation	2.39
Emotional Lability	0.91	Urinary Pain/Urgency/Hesitancy	2.12
Vision Changes	0.90	Hallucinations	1.80
Hearing Impairment	0.86	Palpitations	1.80
Decreased Appetite	0.78	Decreased Appetite	1.78
Heartburn	0.77	Anger	1.75
Anger	0.72	Heartburn	1.62
Hallucinations	0.62	Hearing Impairment	1.45
Difficult/Painful Swallowing	0.56	Vision Changes	1.42
Malaise	0.41	Difficult/Painful Swallowing	0.88
Impaired Attention/Executive Function	0.26	Malaise	0.75
Elevated Mood	0.21	Impaired Memory	0.74

By note (*N*=504,395 notes)		**By patient (*N*=133,977 patients)**	
Symptom	**%**	**Symptom**	**%**
Impaired Memory	0.20	Impaired Attention/Executive Function	0.63
Upset Stomach	0.15	Upset Stomach	0.41
Vaginal Dryness/Irritation	0.08	Vaginal Dryness/Irritation	0.29
Impaired Concentration	0.07	Impaired Concentration	0.21
Difficulty Coping	0.05	Tinnitus	0.14
Hot flashes	0.04	Difficulty Coping	0.14
Tinnitus	0.04	Hot flashes	0.13
Increased Thirst	0.03	Increased Thirst	0.11
Dry Mouth	0.03	Dry Mouth	0.09
Increased Appetite	0.01	Altered Sense of Taste	0.03
Choking Sensation	0.01	Increased Appetite	0.02
Dyspareunia	0.01	Choking Sensation	0.01
Altered Sense of Taste	0	Dyspareunia	0.01
Altered Sense of Smell	0	Altered Sense of Smell	0
Decreased Libido	0	Decreased Libido	0
Decreased Thirst	0	Decreased Thirst	0
Defecation Pain/Urgency	0	Defecation Pain/Urgency	0
Tactile Perception	0	Tactile Perception	0

Note. For notes, prevalence was calculated by dividing the number of nursing notes with a symptom documented by the total number of notes in the entire corpus. For patients, prevalence was calculated by dividing the number of patients with a symptom documented one or more times in a nursing note in 2016 by the total number of patients. Prevalence calculations were completed for each symptom.

Table 3

Demographics for Patients Diagnosed with Chronic Conditions

	COPD		Heart Failure		T2DM		Cancer		Total	
	***n*=3,399**		***n*=6,587**		***n*=12,139**		***n*=7,269**		***N*=22,647**	
	$\bar{x}$	***med***	$\bar{x}$	***med***	$\bar{x}$	***med***	$\bar{x}$	***med***	$\bar{x}$	***med***
Age (years)	68.2	69.0	70.3	72.0	65.2	66.0	65.2	67.0	65.5	67.0
	n	***%***	***n***	***%***	***n***	***%***	***n***	***%***	***n***	***%***
Sex/Gender										
Male	1,609	47.3	3,606	54.7	5,820	47.9	3,491	48.0	10,977	48.5
Female	1,789	52.6	2,981	45.3	6,318	52.0	3,778	52.0	11,669	51.5
Other	1	0	-		1	0	-		1	0
Race										
White	1,094	32.2	2,434	37.0	3,034	25.0	3,291	45.3	7,468	33.0
Black	374	11.0	798	12.1	1,285	10.6	737	10.1	2,353	10.4
Unknown/Declined	1,871	55.0	3,177	48.2	7,433	61.2	2,906	40.0	12,075	53.3
Other	60	1.8	178	2.7	387	3.2	335	4.6	751	3.3
Ethnicity										
Hispanic/Latinx	374	11.0	729	11.1	1,762	14.5	849	11.7	2,829	12.5
Non-Hispanic/Latinx	915	26.9	2,092	31.8	2,422	20.0	2,939	40.4	6,300	27.8
Unknown/Declined	1,993	58.6	3,493	53.0	7,283	60.0	3,174	43.7	12,492	55.2
Other	117	3.4	273	4.1	672	5.5	307	4.2	1,026	4.5

Note. med = median. COPD = chronic obstructive pulmonary disease. T2DM = type 2 diabetes mellitus.

Table 4

Comparison of Symptom Clusters (i.e., Groups of Patients with Similar Symptom Profiles) Identified for Patients Diagnosed with COPD, Heart Failure, T2DM, and Cancer and the Top Five Symptoms Characterizing the Cluster When Identified for a Chronic Condition

Chronic Condition	Symptom Clusters: No Symptoms	Cognitive & Neurological Symptoms	Pain & Other Symptoms	Cardiovascular Symptoms	Mental Health Symptoms	Gastrointestinal Symptoms & Fatigue
COPD	Unsteadiness-no	*Not identified*	Pain-yes	*Not identified*	Suicidal ideation-yes	*Not identified*
	Pain-no		Unsteadiness-yes		Depressed mood-yes	
	Nausea-no		Nausea-yes		Agitation-yes	
	Anxiety-no		Weakness-yes		Anxiety-yes	
	Disturbed Sleep-no		Shortness of breath-yes		Confusion-yes	
Heart Failure	Shortness of breath-no	Impaired cognition-yes	*Not identified*	Shortness of breath-yes	*Not identified*	*Not identified*
	Pain-no	Confusion-yes		Chest pain-yes		
	Chest pain-no	Paresthesia/numbness-yes		Pain-yes		
	Unsteadiness-no	Agitation-yes		Dizziness-yes		
	Swelling-no	Fatigue-yes		Nausea-yes		
T2DM	Shortness of breath-no	Impaired cognition-yes	Shortness of breath-yes	Palpitations-yes	*Not identified*	*Not identified*
	Pain-no	Confusion-yes	Pain-yes	Dizziness-yes		
	Nausea-no	Agitation-yes	Nausea-yes	Shortness of breath-yes		
	Disturbed sleep-no	Unsteadiness-yes	Chest pain-yes	Chest pain-yes		
	Chest pain-no	Fatigue-yes	Abdominal pain-yes	Pain-yes		
Cancer	Pain-no	Impaired cognition-yes	Pain-yes	*Not identified*	*Not identified*	Decreased appetite-yes
	Shortness of breath-no	Confusion-yes	Shortness of breath-yes			Fatigue-yes
	Nausea-no	Unsteadiness-yes	Nausea-yes			Nausea-yes
	Unsteadiness-no	Paresthesia/numbness-yes	Chest pain-yes			Abdominal pain-yes
	Decreased appetite-no	Weakness-yes	Abdominal pain-yes			Weakness-yes

Note. COPD = chronic obstructive pulmonary disease. T2DM = type 2 diabetes mellitus.