Advances in Conceptual and Methodological Issues in Symptom Cluster Research: A Twenty Year Perspective

Carolyn S. Harris, RN BSN[1], **Marylin Dodd, RN, PhD**[1], **Kord M. Kober, PhD**[1], **Anand A. Dhruva, MD**[2], **Marilyn Hammer, RN, PhD**[3], **Yvette P. Conley, PhD**[4], **Christine A. Miaskowski, RN, PhD**[1,2]

[1]School of Nursing, University of California, San Francisco, CA, USA

[2]School of Medicine, University of California, San Francisco, CA, USA

[3]Dana-Farber Cancer Institute, Boston, MA, USA

[4]School of Nursing, University of Pittsburgh, Pittsburgh, PA, USA

INTRODUCTION

Symptom science was transformed by two landmark papers that suggested the existence of "symptom clusters" in oncology patients.[1,2] Prior to these papers, symptom research focused primarily on an evaluation of the prevalence and severity of single symptoms in patients with chronic conditions.[3] Building on the clinical reality that symptoms rarely occur alone, researchers and clinicians were challenged to evaluate for and manage co-occurring symptoms and/or symptom clusters.

Given that these two studies published in 2001 are credited with launching the field of "symptom cluster" research,[1,2] they warrant careful evaluation twenty years later. In the first study,[2] the relationships between pain and fatigue and the co-occurrence of 20 other symptoms were evaluated in a heterogeneous sample of newly diagnosed oncology patients over one year. In the second study,[1] the effect of a pre-specified symptom cluster (i.e., pain, fatigue, sleep disturbance) on oncology patients' functional status was evaluated over three cycles of chemotherapy. Of note, in this paper,[1] the first definition of a symptom cluster was proposed to be "three or more concurrent symptoms" that "are related to each other….The symptoms within a cluster are not required to share the same etiology," (pp465).

While these studies provided a stimulus and new directions for symptom science research, several limitations warrant consideration. First, only two symptoms (i.e., pain, fatigue) were evaluated in one study[2] and three symptoms (i.e., pain, fatigue, sleep insufficiency) in the other study.[1] In both studies, the symptom cluster was pre-specified, not created "de novo". Third, both studies evaluated for associations between single symptoms and a distal outcome, not with the "symptom cluster" as a whole.

Corresponding author: Carolyn S. Harris, RN, BSN, Department of Physiological Nursing, School of Nursing, University of California, San Francisco, 2 Koret Way, N605E, San Francisco, CA 94143, USA, Phone: 415-476-9407, Fax: 415-476-8899, carolyn.harris@ucsf.edu.

While symptom cluster research has grown considerably since the publication of these two relatively "simplistic" studies,[1,2] as noted in the most recent expert panel report,[4] this field is relatively new and ongoing conceptual issues warrant consideration. One key question is a rather simple one, namely: "What constitutes symptom cluster research?" As noted by Miaskowski and colleagues in 2007,[5] two conceptual approaches to evaluate symptom clusters evolved over a period of five years, namely: "clustering" symptoms (equates with a variable-centered analytic approach) and "clustering" patients (equates with a person-centered analytic approach) (Figure 1). The use of the word "clustering" for both approaches has led to confusion in the literature on symptom cluster research. For example, it is not uncommon to find publications that have described "symptom clusters" when patients were grouped based on an evaluation of a pre-specified symptom cluster that consisted of two or more symptoms.[6,7] Given this confusion, it is imperative to use the correct terminology as outlined below.

As noted in Figure 1A, variable-centered approaches (e.g., exploratory factor analysis (EFA)) identify *symptoms* that cluster together empirically through the use of an analytic approach that creates distinct groups of related symptoms (i.e., symptom clusters).[5] These approaches are based on the hypothesis that symptoms cluster together because they may share a common underlying mechanism(s).[8,9]

Patient-centered approaches (Figure 1B; e.g., latent class analysis (LCA)) identify subgroups of *patients* with distinct symptom profiles using one or more symptoms or a pre-specified symptom cluster (e.g., pain, fatigue, depression, sleep disturbance[10]). With these approaches, it is important to note that in the context of symptom cluster research, a symptom cluster must be pre-specified. These patient-centered analyses can be used to identify subgroups of patients with distinct symptom(s) profiles (i.e., lower versus higher symptom burden) and associated risk factors (e.g., demographic, clinical, biomarkers).[5]

Previous reviews have evaluated the conceptual, methodological, and clinical basis for symptom cluster research.[5,11–15] In a concept analysis that included a review of symptom cluster research across psychiatry, medicine, and nursing, Kim and colleagues[14] identified five key attributes of a symptom cluster (e.g., co-occurrence of symptoms within a cluster, stability, shared or common etiology). Based on research findings and clinical evidence, both Kim and colleagues[14] and Aktas[11] argued for the definition of a symptom cluster to be modified to include a minimum of two symptoms. Kim and Abraham[13] and Skerman and colleagues[15] examined the application of various statistical methods to identify symptom clusters and reviewed the conceptual and methodological challenges of each method. Building on a previous paper by Miaskowski and colleagues[5] that described the two conceptual approaches for symptom cluster research, Barsevick[12] examined the application of qualitative approaches to symptom cluster research and expanded on the concept of stability in symptom cluster research.

In the most recent state of the science report,[4] an expert panel called for the identification of symptom clusters using newer analytic techniques and for an investigation of the underlying mechanisms for symptom clusters. In addition, they suggested that additional research is warranted to clarify the "de novo" approach to the identification of symptom

clusters versus the grouping of patients with distinct symptom cluster profiles based on a "pre-specified" symptom cluster. Given the recent application of newer methods to symptom cluster research (e.g., network analysis (NA),[16] natural language processing (NLP[17])), a review of the conceptual basis for these older and newer methods in the context of symptom cluster research is warranted. Therefore, the purposes of this paper are to review the conceptual basis for symptom cluster research; compare and contrast the conceptual basis for using variable-centered versus patient-centered analytic approaches in symptom cluster research; review the strengths and weaknesses of the most common variable-centered and patient-centered analytic approaches for symptom cluster research; and compare the various applications of each approach in symptom cluster research.

DEFINITION OF A SYMPTOM CLUSTER

As the science of symptom cluster research has advanced over the past 20 years, the definition of a symptom cluster has gone through multiple revisions.[1,12,14] In the most recent revision by an expert panel,[4] several characteristics of both a symptom and a symptom cluster were identified (Table 1). While some debate continues on the minimum number of symptoms that constitutes a symptom cluster,[11,12] a minimum of two symptoms in a cluster is generally accepted. However, clarification and/or refinement of the other characteristics are needed. For example, in terms of "stability," neither the definition of nor the methods to assess stability exist. This issue is particularly important when one considers the temporal dimension of symptom clusters. Does stability refer to whether or not the various types of symptom clusters (e.g., psychological, gastrointestinal) remain "stable" or whether or not the symptoms within each cluster (e.g., sad, irritable, angry) remain "stable" over time? We propose that the term "stable" be used to describe whether the symptom clusters change over time and/or across symptom dimensions. Alternatively, the term "consistent" should be used to describe whether the specific symptoms within a cluster remain the same over time and/or across symptom dimensions. For both stability and consistency, the assessment methods and numeric criteria need to be determined.[18]

Equally important is the question of whether or not symptom clusters need to be independent of other clusters. Given the recent use of NA, that demonstrates that symptoms within one cluster are related to symptoms in other clusters,[16] this criterion may need to be reconsidered. Equally important, research is needed to support the criteria that symptom clusters may share common underlying mechanisms and may have shared outcomes.

TWO BROAD APPROACHES TO SYMPTOM CLUSTER RESEARCH

De Novo Identification of Symptom Clusters

Variable-centered approaches explore the relationships among symptoms using either regression-based techniques[19] or measures of similarity[13] and create symptom clusters "de novo." As a first step, participants need to complete one or more symptom assessment instruments or a symptom inventory (Figure 1A).[5] Then, a variable-centered analytic approach is used to identify the symptom clusters. Historically, four statistical approaches were used to identify symptom clusters, namely: cluster analysis, EFA, confirmatory factor analysis (CFA), and principal components analysis (PCA).[14]

Following the recommendations of Skerman and colleagues,[15] EFA is the most common approach used to identify symptom clusters in oncology research, followed by hierarchical cluster analysis (HCA).[14,18,20] In contrast, PCA is the most common approach used to identify symptom clusters in other chronic conditions (e.g., chronic obstructive pulmonary disease (COPD),[21] human immunodeficiency virus[22]). However, PCA uses a data-reduction approach to analyze symptoms and does not assume any causal relationship between the symptoms within a cluster.[15,23] Given that one hypothesis underlying symptom cluster research is that symptoms cluster together due to a shared, underlying mechanism,[8,9] the use of PCA is not consistent with this hypothesis.

A non-exhaustive search of the Cumulative Index of Nursing and Allied Health Literature (CINAHL) and PubMed databases was conducted to explore the use of different variable-centered approaches for studying symptom clusters. Exemplars for each statistical method are described in Supplemental Table 1. As noted below, compared to studies of oncology patients, research on symptom clusters in patients with other chronic conditions is much less common. Therefore, exemplar studies conducted in samples with other chronic conditions are highlighted in Supplemental Table 1 to stimulate growth in symptom cluster research within these patient populations.

Hierarchical cluster analysis.—HCA is one type of cluster analysis that has been used in symptom cluster research across a variety of chronic conditions.[20,22,24] It is important to note that depending on the research question, HCA can be used to group symptoms or patients.[13]

Two types of HCA can be used: agglomerative or divisive.[25] Starting with all of the symptoms in individual clusters, agglomerative HCA is used to identify and successively group pairs or groups of similar symptoms into mutually exclusive clusters of related symptoms.[26] In contrast, divisive HCA starts with all of the symptoms in a single cluster. Then, it systematically partitions the cluster into smaller groups of similar symptoms.[25] The hierarchical clustering of symptoms continues in a stepwise fashion until a certain level of groupings that have clinical meaning and interpretability are selected.[15] These steps are displayed graphically on a dendrogram. Measures of similarity for interval data include correlation coefficients or squared Euclidean distances,[13] while coefficients of association can be used for binary data.[15]

HCA has several limitations.[13,15] First, it is important to note that cluster analytic methods are not based on the underlying assumption of shared causality. Rather, they seek to identify groupings based on statistical measures of similarity.[13] Second, because cluster analytic methods strive to identify mutually exclusive groups of similar symptoms, a symptom can belong to only one cluster.[15] Given that a single symptom may be related to multiple symptoms that associate into different clusters, this limitation does not allow for an examination of symptoms that cross-load on other clusters. In addition, it impedes our ability to identify common and distinct underlying mechanisms. Third, using HCA, the determination of the final number of clusters is highly subjective. This subjectivity may lead to bias, as well as variability in both the number and types of symptom clusters identified across studies.

Thirty-nine studies were identified that evaluated for symptom clusters "de novo" using HCA. While 74.4% of these studies were conducted in patients with cancer, exemplars of studies that used HCA to identify symptom clusters in patients with other chronic conditions are provided in Supplemental Table 1.

Exploratory factor analysis.—The common factor model consists of two factor analytic methods: EFA and CFA. Factor analytic methods are used to discover unobserved or latent factors (i.e., symptom clusters) that account for the common variance among multiple, observed variables (i.e., symptoms).[27] The underlying conceptual framework for factor analytic methods is that variables within a latent factor covary due to a common, underlying cause. The "strength and direction of the influence"[23](pp10) of the latent factors on the variables in the common factor model are estimated with factor loadings. Because of the exploratory nature of EFA, no assumptions are made a priori about the nature of the relationships between the observed variables.[23]

A unique feature of EFA is that symptoms can load on more than one factor (i.e., symptom cluster).[23] Given the possibility that one symptom can influence symptoms on different clusters, the ability for a symptom to load on more than one cluster has conceptual utility. For example, in a study that evaluated for symptom clusters in patients with lung cancer,[28] difficulty concentrating and feeling nervous cross-loaded on a sickness behavior and a psychological cluster. However, a lack of consensus exists on whether a symptom can load on multiple factors. For example, in a recent review of studies that evaluated for symptom clusters in patients receiving adjuvant chemotherapy,[18] only 58.8% of the studies that used EFA allowed for symptoms to cross-load.

Compared to HCA where 39 studies were identified, 89 studies used EFA to identify symptom clusters "de novo." Of these studies, 66.3% were conducted in patients with cancer. This pattern is consistent with previous reviews that identified EFA as the most common statistical approach for identifying symptom clusters in oncology patients.[15,18,20] Exemplars of studies that used EFA to identify symptom clusters in patients with other chronic conditions are provided in Supplemental Table 1.

Confirmatory factor analysis.—This approach is used to test hypotheses on the relationships between latent factors and observed variables.[27] More specifically, all of the model's assumptions (e.g., number of factors, pattern of variable to factor loadings) must be specified a priori. These hypotheses must be rooted in theory and/or empirical evidence.

Given that the conceptual basis for CFA is to confirm hypotheses, it can be used to confirm the number and types of symptom clusters previously identified using another variable-centered approach (e.g., EFA).[15] For example, in a study that evaluated for symptom clusters in children and adolescents receiving myelosuppressive therapy,[29] EFA was used to identify symptom clusters. Then, CFA was used to confirm the structure of the findings. Given the continued need to evaluate and compare different statistical methods to identify symptom clusters "de novo,"[4] CFA may be one approach to validate the stability and/or consistency of symptom clusters.

Use of variable-centered approaches to investigate underlying biological mechanisms.—Relatively few studies have used a variable-centered approach to evaluate the underlying biological mechanisms of symptom clusters.[30,31] In one study,[31] EFA was used to identify symptom clusters in oncology patients using the severity dimension. Then, a factor severity score was calculated for each of the three symptom clusters that were identified (i.e., mood-cognitive, sickness-behavior, and treatment-related symptom). These scores were used in regression analyses to identify associations between each symptom cluster and polymorphisms in cytokine genes.

Another study used EFA to identify two symptom clusters in patients with COPD.[30] Next, symptom cluster severity scores were calculated for each cluster. Subgroups of patients were identified based on their average symptom cluster severity score. Inflammatory biomarkers were used in logistic regression analyses to identify associations between subgroup membership and levels of C-reactive protein.

A Priori Identification of Symptom Clusters and Associated Symptom Cluster Profiles

Patient-centered analytic approaches evaluate for relationships among individuals using the principles of structural equation modeling[19] or measures of similarity.[13] Similar to variable-centered approaches, participants complete one or more symptom assessment instruments or a symptom inventory (Figure 1B).[5] In the context of symptom cluster research, a symptom cluster must be identified a priori (e.g., pain, fatigue, sleep disturbance, and depression). Then, with this pre-specified symptom cluster, groups of patients with distinct symptom cluster profiles are identified using patient-centered analytic approaches. Because these methods allow for the identification of subgroups of patients based on their experiences with a pre-specified symptom cluster, a variety of phenotypic and molecular risk factors can be identified that distinguish the various patient subgroups.

A search of the CINAHL and PubMed databases identified 31 studies that evaluated the symptom profiles of patients experiencing a pre-specified symptom cluster. Exemplars for each statistical method are provided in Supplemental Table 2.

Hierarchical cluster analysis.—As mentioned previously, cluster analysis methods like HCA can be used to "cluster" symptoms or patients. With the latter approach, subgroups of patients are identified based on similar symptom cluster profiles using a pre-specified symptom cluster.[13] Eight studies were identified that used HCA to evaluate for subgroups of patients based on a clearly defined pre-specified symptom cluster. While the majority of these studies were conducted in patients with cancer (75%), exemplar studies that used HCA to identify subgroups of patients with a distinct symptom cluster profile in other chronic conditions are provided in Supplemental Table 2.

Latent variable modeling (LVM).—LVM is used to identify subgroups or classes of individuals within a sample or population who have similar attributes or symptom experiences.[19] The underlying conceptual framework for LVM is that subgroup membership is based on an unobserved, latent variable (i.e., pre-specified symptom cluster) whose "value indicates what group the individual belongs to"[25](pp819). Common types of LVM include LCA for categorical data (e.g., symptom occurrence) and latent profile analysis for

continuous data (e.g., symptom severity). In addition, latent transition analysis can be used to evaluate for changes in subgroup membership over time.[19]

The identification of subgroups of patients based on their distinct symptom cluster profiles using LVM has multiple advantages. First, differences in salient characteristics (e.g., demographics, stress, resilience) between the subgroups can be identified. Second, LVM can be used to evaluate how patient outcomes (e.g., functional status, quality of life) differ by class membership.[25]

While the use of both HCA and LVM results in the identification of subgroups of patients with distinct symptom cluster profiles, the methods differ in a few key ways. First, with LVM, multiple models are evaluated using fit indices prior to selecting the final model.[25] In contrast, selection of the final solution for HCA is highly subjective. Second, because LVM tends to be computationally more challenging than HCA,[25] fewer variables may be included in the LVM analysis.

Twenty-three studies have used a form of LVM to identify subgroups of patients with a distinct symptom cluster profile. While most of these studies were conducted in oncology patients (56.5%), exemplar studies that used LVM to identify subgroups of patients with a distinct symptom cluster profile in other chronic conditions are provided in Supplemental Table 2.

Use of patient-centered analytic approaches to investigate underlying biological mechanisms.—Ten studies have used a patient-centered analytic approach to evaluate the underlying biological mechanism(s) for a pre-specified symptom cluster (exemplars in Supplemental Table 2). In one study,[32] latent profile analysis was used to identify three distinct subgroups of breast cancer patients based on their experience with a pain, fatigue, sleep disturbance, and depression cluster. Multiple associations were found between latent class membership and cytokine gene polymorphisms. Another study used HCA to identify subgroups of patients with advanced cancer based on their experience with the symptom cluster of pain, fatigue, depression, and sleep disturbance.[10] Higher serum levels of IL-6 were associated with an increased risk for membership in the moderate-to-high symptom subgroup.

EMERGING METHODS IN SYMPTOM CLUSTER RESEARCH

Network Analysis

One novel approach that can be used to identify symptom clusters "de novo" is NA. Based on the principles of graph theory,[33] NA is used to evaluate the relationships between a set of variables (i.e., symptoms). The structure of these relationships is presented in graphs. Within these graphs, symptoms are represented as nodes and the relationship(s) between symptoms are represented as edges (Figure 2A). The presence (i.e., a relationship between the symptoms) and strength (e.g., correlation, conditional association) of these edges are calculated from the data. While firmly based in mathematical and statistical methods, a strength of NA is that it allows for a qualitative (i.e., visual) appraisal of the data.

One challenge with NA is the determination of the importance of nodes or groups of nodes within a network. Various types of centrality indices are used to aid in the interpretation of which nodes (i.e., symptoms) may have the largest influence on a network.[33–35] These highly influential nodes are sometimes referred to as "core" or "sentinel" nodes[16] and have the potential to serve as targets for therapeutic interventions.

Following the network's construction, community detection algorithms are used to identify clusters of symptoms (i.e., nodes) that are closely connected relative to other symptoms or clusters.[36] Various types of community detection algorithms are available and selection of the appropriate algorithm depends on multiple factors, including the network's size.[37]

One of the advantages of NA over other analytic approaches is that you can visualize the relationships between symptom clusters and how symptoms within one cluster relate to symptoms in another cluster. In addition, this approach allows for the identification of core or sentinel symptoms. However, a variety of approaches exist to create the networks and selection of the appropriate algorithms to estimate and evaluate the networks warrant consideration.

Three studies were identified that used NA to evaluate symptoms and/or symptom clusters in patients with cancer.[16,38,39] In one study,[16] NA was used to identify symptom clusters using multiple dimensions of the symptom experience (i.e., occurrence, severity, distress) in a heterogeneous sample of oncology patients. While five symptom clusters were identified across all three symptom dimensions (i.e., psychological, hormonal, respiratory, nutritional, chemotherapy-related), two additional symptom clusters (i.e., gastrointestinal, epithelial) were identified using distress (Figure 2B). The authors hypothesized that these results suggest that distress is a unique dimension of the patients' symptom experience. Because nausea and lack of appetite had the highest centrality index scores, the authors suggested that targeting these symptoms may decrease the other symptoms within the network.

In another study,[39] a network was constructed using severity scores for eight symptoms and serum concentrations for 13 cytokines. Two communities were identified: a symptom cluster with five symptoms and another cluster with all 13 cytokines. While an evaluation of the associations between symptoms and biomarkers warrants additional research, findings from this study illustrate the challenges with incorporating heterogenous types of data (i.e., symptom severity scores and cytokine levels) into a NA.

A third study used HCA and PCA to identify symptom clusters in a sample of patients receiving chemotherapy.[38] Three common symptom clusters were identified over five assessments. Then, using only the 12 symptoms that were identified in the initial analyses, NA identified comparable symptom clusters that were found using PCA only at one timepoint. Fatigue, anxiety, and depression were identified as the most central symptoms in the network.

Bayesian Networks Analysis

Bayesian NA incorporates Bayesian statistics with NA to allow for an evaluation of the strength and direction of the relationships among symptoms.[40] While both types of networks

contain nodes (i.e., symptoms) and edges (i.e., relationships between the symptoms), Bayesian NA graphically displays these relationships in a causal model (i.e., directed acyclic graph). Conditional dependencies are estimated for each node (i.e., symptom). The strength and direction of these relationships are calculated with joint probability distributions.[41]

Bayesian NA approaches offer many advantages for symptom cluster research. First, in addition to identifying "sentinel" symptoms, Bayesian NA can be used to elucidate the direction and the flow of a symptom's influence on other symptoms within a network.[41] Second, similar to EFA and LVM, Bayesian NA can identify latent variables.[42,43] However, given the complexity of the relationships between symptoms, interpretation of these relationships on an acyclic graph may be challenging. In addition, Bayesian NA methods are computationally expensive,[44] particularly with large sample sizes or with large symptom inventories.

While Bayesian NA is used extensively in bioinformatics[45] and health sciences[46] research, only one study was identified that used Bayesian NA to examine the relationships between symptoms within a pre-specified cluster (i.e., sleep disturbance, fatigue, depressive symptoms) and their effect on cognitive performance and quality of life in breast cancer patients receiving chemotherapy.[47] Findings from this analysis suggest that the relationships among symptoms changed across time. For example, while mood directly impacted fatigue prior to the start of treatment and at the end of chemotherapy, previous levels of fatigue and sleep disturbance and current quality of life directly impacted the severity of fatigue one year after the start of chemotherapy.

Application of NLP to Symptom Cluster Research

An ongoing issue in symptom cluster research is to determine the optimal number of common symptoms that need to be assessed across chronic conditions.[18] The determination of a consistent, comprehensive, and clinically meaningful list of symptoms would enable the identification of common symptom clusters across chronic conditions, as well as their common underlying mechanisms. Because of this lack of consensus, inventories with a large number of symptoms are administered to patients to evaluate for symptom clusters, with a potential for increased burden. A variety of new and emerging data science approaches (e.g., machine learning, NLP) have the potential to resolve this issue. The application of one of these approaches in symptom cluster research is described below.

NLP is a data extraction method that uses computer-based algorithms to acquire, process, and modify natural language obtained from "Big Data" (e.g., electronic health record (EHR)) for computational analyses.[48] Systematic extraction of "real world" symptom data from EHRs and its subsequent evaluation has the potential to not only lessen the burden on patients with chronic conditions, but provide researchers with the "most comprehensive, longitudinal, population-wide dataset"[17](pp907) available. NLP methodologies have the potential to provide novel information on symptoms and symptom management throughout and beyond treatment of chronic conditions.[49]

Two recent publications describe the use of NLP in symptom science research. In the first publication,[50] the authors used a free and open-source NLP software (i.e., NimbleMiner)

to find and extract data on five symptoms (i.e., constipation, depressed mood, disturbed sleep, fatigue, palpitations) from the EHR. While this method was piloted using only five symptoms, it can be expanded to include a larger symptom "vocabulary."

In the second study,[17] Koleck and colleagues used NLP to extract 56 symptoms from the EHR nursing notes of 22,647 patients across four common chronic conditions (i.e., cancer, COPD, heart failure, type II diabetes). Then, HCA was used to identify subgroups of patients with distinct symptom profiles for each chronic condition. While condition-specific symptom profiles were identified (e.g., gastrointestinal symptoms and fatigue for cancer, mental health symptoms for COPD), multiple symptom profiles were identified across two or more chronic conditions (e.g., cognitive and neurological). Given the strength of their results and the ability of NLP software tools to accurately identify and obtain specific symptom data, ongoing development of these methods has the potential to be applied to symptom cluster research and to advance symptom science.

FUTURE DIRECTIONS

In their report,[4] the expert panel called for an examination of symptom clusters across various chronic conditions. These types of comparative studies are needed to determine whether or not "generic" symptom clusters occur across chronic conditions. To accomplish this goal, a comprehensive symptom assessment, as well as consistent methods, need to be used. Equally important, with the emergence of NA and NLP, studies are needed that compare symptom clusters that are created "de novo" using various analytic approaches.

Based on the literature reviews for each analytic approach, notable gaps in symptom cluster research were identified. In general, the study samples were homogeneous in terms of race or ethnicity, gender identity, socioeconomic status, and educational attainment. Given that each of these characteristics can impact an individual's symptom experience, health outcomes, and quality of life, this lack of diversity and evaluation of a limited number of social determinants of health limits our understanding of how these factors may influence the relationships with and among symptoms and symptom clusters. Future research that evaluates for symptom clusters in diverse and/or underserved samples, across a variety of acute and chronic conditions, is needed. Exemplars of studies that evaluated for differences in symptom clusters in relationship to age, gender, socioeconomic status, or ethnicity are provided in Supplemental Tables 1 and 2.

While the definition of a symptom cluster has evolved over the past 20 years, multiple issues remain that warrant careful consideration to move this area of scientific inquiry forward (Table 1). Specifically, clear criteria need to be developed to determine the stability and consistency of symptom clusters. The establishment of these criteria will allow researchers to determine within studies whether symptom clusters change over time and/or across dimensions of the symptom experience. In addition, they can be used to evaluate stability and consistency of symptom clusters across studies of patients with similar and different chronic conditions. Additional research is needed to determine whether symptoms in a cluster must be independent or can cross-load on more than one cluster. Given that previous studies that used EFA and NA demonstrated that symptoms may load on multiple clusters,

or that symptoms within clusters and the clusters themselves are related, this characteristic of a symptom cluster may need to be revised. One way to resolve this issue would be to evaluate common and distinct mechanisms that underlie various symptom clusters that include symptoms that cross-load on more than one cluster.

CONCLUSION

As symptom cluster research continues to evolve, the use of both variable-centered and patient-centered analytic approaches are needed to move the science forward. While each approach has unique strengths and weaknesses, conceptual clarity is needed when a study is designed and the research question should inform the selection of the appropriate method. The conceptual approaches illustrated in Figure 1 can serve as a guide for future studies. Variable-centered approaches identify symptom clusters and are based on the hypothesis that symptoms cluster together because they may share a common underlying mechanism(s). The terminology "symptom clusters" should be used when symptom clusters are created with this approach (Figure 1A). Patient-centered analyses identify subgroups of patients with distinct symptom cluster profiles and associated risk factors. Researchers should clearly specify when they are "clustering" patients (Figure 1B) that they have used a pre-specified symptom cluster and identified "subgroups of patients with distinct symptom cluster profiles."

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Dodd M, Miaskowski C, Paul S. Symptom clusters and their effect on the functional status of patients with cancer. Oncol Nurs Forum. 2001;28(3):465–470. [PubMed: 11338755]

2. Given CW, Given B, Azzouz F, Kozachik S, Stommel M. Predictors of pain and fatigue in the year following diagnosis among elderly cancer patients. J Pain Symptom Manage. 2001;21(6):456–466. [PubMed: 11397603]

3. Dodd M, Janson S, Facione N, et al. Advancing the science of symptom management. J Adv Nurs. Mar 2001;33(5):668–76. doi:10.1046/j.1365-2648.2001.01697.x [PubMed: 11298204]

4. Miaskowski C, Barsevick A, Berger A, et al. Advancing symptom science through symptom cluster research: Expert panel proceedings and recommendations. J Natl Cancer Inst. Apr 2017;109(4)doi:10.1093/jnci/djw253

5. Miaskowski C, Aouizerat BE, Dodd M, Cooper B. Conceptual issues in symptom clusters research and their implications for quality-of-life assessment in patients with cancer. J Natl Cancer Inst Monogr. 2007;(37):39–46. doi:10.1093/jncimonographs/lgm003 [PubMed: 17951230]

6. Hsu HT, Lin KC, Wu LM, et al. Symptom cluster trajectories during chemotherapy in breast cancer outpatients. J Pain Symptom Manage. 2017;53(6):1017–1025. doi:10.1016/j.jpainsymman.2016.12.354 [PubMed: 28196783]

7. Woods NF, Cray LA, Mitchell ES, Farrin F, Herting J. Polymorphisms in estrogen synthesis genes and symptom clusters during the menopausal transition and early postmenopause: Observations from the Seattle Midlife Women's Health Study. Biol Res Nurs. Mar 2018;20(2):153–160. doi:10.1177/1099800417753536 [PubMed: 29334760]

8. Cleeland CS, Bennett GJ, Dantzer R, et al. Are the symptoms of cancer and cancer treatment due to a shared biologic mechanism? A cytokine-immunologic model of cancer symptoms. Cancer. Jun 1 2003;97(11):2919–25. doi:10.1002/cncr.11382 [PubMed: 12767108]

9. Miaskowski C, Dodd M, Lee K. Symptom clusters: The new frontier in symptom management research. J Natl Cancer Inst Monogr. 2004;(32):17–21. doi:10.1093/jncimonographs/lgh023 [PubMed: 15263036]

10. Ji YB, Bo CL, Xue XJ, et al. Association of inflammatory cytokines with the symptom cluster of pain, fatigue, depression, and sleep disturbance in Chinese patients with cancer. J Pain Symptom Manage. Dec 2017;54(6):843–852. doi:10.1016/j.jpainsymman.2017.05.003 [PubMed: 28797869]

11. Aktas A. Cancer symptom clusters: Current concepts and controversies. Curr Opin Support Palliat Care. Mar 2013;7(1):38–44. doi:10.1097/SPC.0b013e32835def5b [PubMed: 23287418]

12. Barsevick A. Defining the symptom cluster: How far have we come? Semin Oncol Nurs. Nov 2016;32(4):334–350. doi:10.1016/j.soncn.2016.08.001 [PubMed: 27776831]

13. Kim HJ, Abraham IL. Statistical approaches to modeling symptom clusters in cancer patients. Cancer Nurs. 2008;31(5):E1–E10.

14. Kim HJ, McGuire DB, Tulman L, Barsevick AM. Symptom clusters: Concept analysis and clinical implications for cancer nursing. Cancer Nurs. Jul–Aug 2005;28(4):270–82. [PubMed: 16046888]

15. Skerman HM, Yates PM, Battistutta D. Multivariate methods to identify cancer-related symptom clusters. Res Nurs Health. Jun 2009;32(3):345–60. doi:10.1002/nur.20323 [PubMed: 19274688]

16. Papachristou N, Barnaghi P, Cooper B, et al. Network analysis of the multidimensional symptom experience of oncology. Article. Sci Rep. 2019;9(1):2258. doi:10.1038/s41598-018-36973-1 [PubMed: 30783135]

17. Koleck TA, Topaz M, Tatonetti NP, et al. Characterizing shared and distinct symptom clusters in common chronic conditions through natural language processing of nursing notes. Res Nurs Health. 2021;44:906–919. doi:https://doi-org.ucsf.idm.oclc.org/10.1002/nur.22190 [PubMed: 34637147]

18. Harris CS, Kober KM, Conley YP, Dhruva AA, Hammer M, Miaskowski CA. Symptom clusters in patients receiving chemotherapy: A systematic review. BMJ Support Palliat Care. 2021;Preprint. doi:10.1136/bmjspcare-2021-003325

19. Muthen B, Muthen LK. Intergrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. Alcohol Clin Exp Res. 2000;24(6):882–891. [PubMed: 10888079]

20. Sullivan CW, Leutwyler H, Dunn LB, Miaskowski C. A review of the literature on symptom clusters in studies that included oncology patients receiving primary or adjuvant chemotherapy. J Clin Nurs. Feb 2018;27(3–4):516–545. doi:10.1111/jocn.14057 [PubMed: 28859255]

21. Jenkins BA, Athilingam P, Jenkins RA. Symptom clusters in chronic obstructive pulmonary disease: A systematic review. Appl Nurs Res. Feb 2019;45:23–29. doi:10.1016/j.apnr.2018.11.003 [PubMed: 30683247]

22. Zhu Z, Zhao R, Hu Y. Symptom clusters in people living with HIV: A systematic review. J Pain Symptom Manage. Jul 2019;58(1):115–133. doi:10.1016/j.jpainsymman.2019.03.018 [PubMed: 30951828]

23. Fabrigar LR, Wegener DT. Exploratory Factor Analysis. In: Understanding Statistics. Oxford University Press; 2012.

24. DeVon HA, Vuckovic K, Ryan CJ, et al. Systematic review of symptom clusters in cardiovascular disease. Eur J Cardiovasc Nurs. Jan 2017;16(1):6–17. doi:10.1177/1474515116642594 [PubMed: 27034451]

25. Woo SE, Jebb AT, Tay L, Parrigon S. Putting the "person" in the center: Review and synthesis of person-centered approaches and methods in organizational science. Organ Res Methods. 2018;21(4):814–845. doi:10.1177/1094428117752467

26. Everitt BS, Landau S, Leese M, Stahl D. Cluster Analysis. 5th ed. Wiley Series in Probability and Statistics. John Wiley and Sons, Ltd.; 2011.

27. Brown TA. Confirmatory Factor Analysis for Applied Research. 2nd ed. Methodology in the Social Sciences. Guilford Press; 2015.

28. Russell J, Wong ML, Mackin L, et al. Stability of symptom clusters in patients with lung cancer receiving chemotherapy. J Pain Symptom Manage. 2019;57(5):909–922. doi:10.1016/j.jpainsymman.2019.02.002 [PubMed: 30768960]

29. Baggott C, Cooper BA, Marina N, Matthay KK, Miaskowski C. Symptom cluster analyses based on symptom occurrence and severity ratings among pediatric oncology patients during myelosuppressive chemotherapy. Cancer Nurs. Jan–Feb 2012;35(1):19–28. doi:10.1097/NCC.0b013e31822909fd [PubMed: 21921793]

30. Yang Z, Cui M, Zhang X, et al. Identification of symptom clusters and their influencing factors in subgroups of Chinese patients with acute exacerbation of chronic obstructive pulmonary disease. J Pain Symptom Manage. Sep 2020;60(3):559–567. doi:10.1016/j.jpainsymman.2020.03.037 [PubMed: 32276100]

31. Miaskowski C, Conley YP, Mastick J, et al. Cytokine gene polymorphisms associated with symptom clusters in oncology patients undergoing radiation therapy. J Pain Symptom Manage. Sep 2017;54(3):305–316. doi:10.1016/j.jpainsymman.2017.05.007 [PubMed: 28797847]

32. Doong SH, Dhruva A, Dunn LB, et al. Associations between cytokine genes and a symptom cluster of pain, fatigue, sleep disturbance, and depression in patients prior to breast cancer surgery. Biol Res Nurs. May 2015;17(3):237–47. doi:10.1177/1099800414550394 [PubMed: 25304131]

33. Newman M. Networks: An introduction. Oxford University Press; 2010.

34. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behav Res Methods. Feb 2018;50(1):195–212. doi:10.3758/s13428-017-0862-1 [PubMed: 28342071]

35. Freeman L. Centrality in social networks conceptual clarification. Soc Networks. 1979;1:215–239.

36. Orman GK, Labatut V. A comparison of community detection algorithms on artificial networks. In: Gama J, Costa VS, Jorge AM, Brazdil PB, eds. Lecture Notes in Computer Science. Springer; 2009:242–256.

37. Yang Z, Algesheimer R, Tessone CJ. A comparative analysis of community detection algorithms on artificial networks. Sci Rep. Aug 1 2016;6:30750. doi:10.1038/srep30750 [PubMed: 27476470]

38. Rha SY, Lee J. Stable symptom clusters and evolving symptom networks in relation to chemotherapy cycles. J Pain Symptom Manage. 2021;61(3):544–554. doi:10.1016/j.jpainsymman.2020.08.008 [PubMed: 32828931]

39. Henneghan A, Wright ML, Bourne G, Sales AC. A cross-sectional exploration of cytokine-symptom networks in breast cancer survivors using network analysis. Can J Nurs Res. Jun 1 2020:1–13. doi:10.1177/0844562120927535

40. Puga JL, Krzywinski M, Altman N. Points of Significance. Bayesian networks. Nat Methods. Sep 2015;12(9):799–800. doi:10.1038/nmeth.3550 [PubMed: 26554085]

41. Su C, Andrew A, Karagas MR, Borsuk ME. Using Bayesian networks to discover relations between genes, environment, and disease. BioData Mining. 2013;6(6):1–21. [PubMed: 23294634]

42. Gao T, Ji Q. Constrained local latent variable discovery. Paper presented at: 25th International Joint Conference on Artificial Intelligence; July 9–15, 2016; New York, NY.

43. Lazic N, Bishop C, Winn J. Structural expectation propagation (SEP): Bayesian structure learning for networks with latent variables. Paper presented at: Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics. In: Proceedings of Machine Learning Research. 2013;31:379–387. Available from https://proceedings.mlr.press/v31/lazic13a.html.

44. Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR. A primer on learning in Bayesian networks for computational biology. PLoS Comput Biol. Aug 2007;3(8):e129. doi:10.1371/journal.pcbi.0030129 [PubMed: 17784779]

45. Cooper GF, Bahar I, Becich MJ, et al. The center for causal discovery of biomedical knowledge from big data. J Am Med Inform Assoc. Nov 2015;22(6):1132–6. doi:10.1093/jamia/ocv059 [PubMed: 26138794]

46. Kyrimi E, McLachlan S, Dube K, Neves MR, Fahmi A, Fenton N. A comprehensive scoping review of Bayesian networks in healthcare: Past, present and future. Artif Intell Med. Jul 2021;117:102108. doi:10.1016/j.artmed.2021.102108 [PubMed: 34127238]

47. Xu S, Thompson W, Ancoli-Israel S, Liu LQ, Palmer B, Natarajan L. Cognition, quality-of-life, and symptom clusters in breast cancer: Using Bayesian networks to elucidate complex relationships. Psychooncology. Mar 2018;27(3):802–809. doi:10.1002/pon.4571 [PubMed: 29055062]

48. Yim WW, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: A review. JAMA Oncol. Jun 1 2016;2(6):797–804. doi:10.1001/jamaoncol.2016.0213 [PubMed: 27124593]

49. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. J Am Med Inform Assoc. Apr 1 2019;26(4):364–379. doi:10.1093/jamia/ocy173 [PubMed: 30726935]

50. Koleck TA, Tatonetti NP, Bakken S, et al. Identifying symptom information in clinical notes using Natural Language Processing. Nurs Res. May-Jun 01 2021;70(3):173–183. doi:10.1097/NNR.0000000000000488 [PubMed: 33196504]
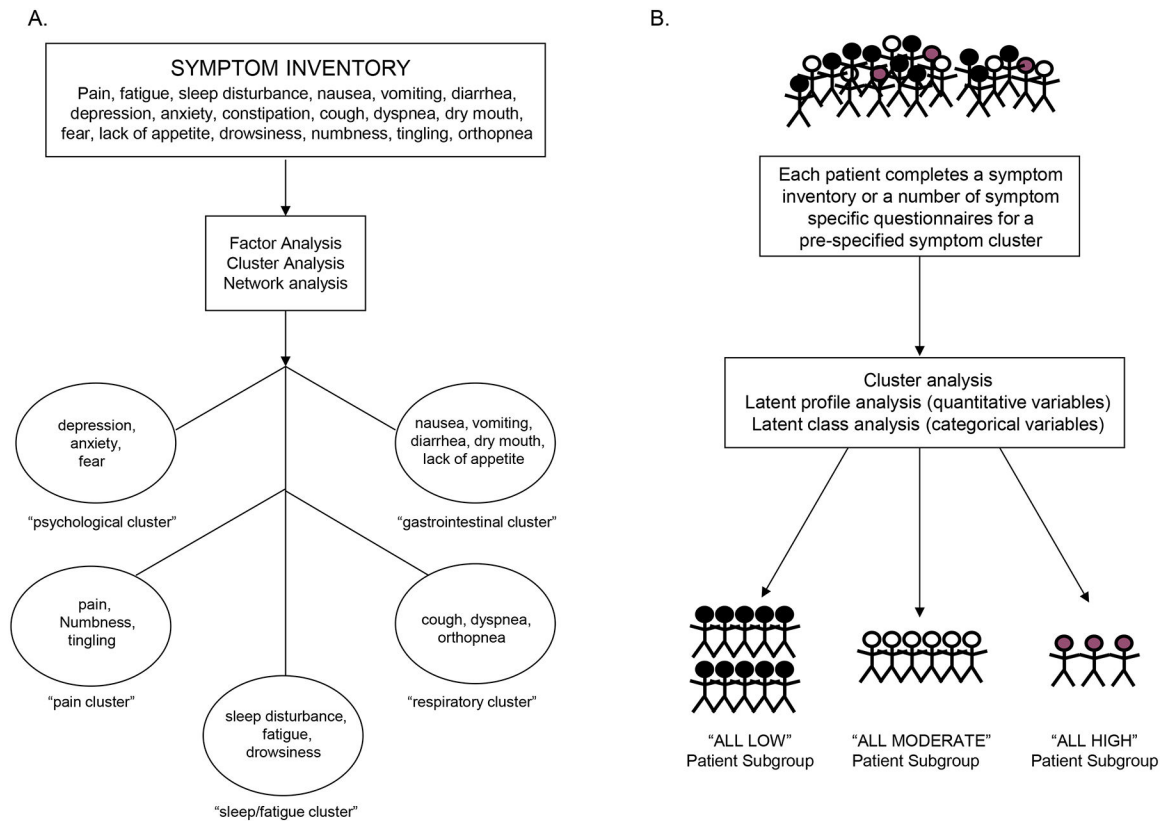
Figure 1.

Two conceptual approaches to symptom cluster research. A) Illustrates the identification of symptom clusters using a variable-centered approach. B) Illustrates the identification of subgroups of patients based on their experience with a pre-specified symptom cluster (e.g., pain, fatigue, sleep disturbance, depression). Adapted from Miaskowski C, Aouizerat BE, Dodd M, Cooper B. Conceptual issues in symptom clusters research and their implications for quality-of-life assessment in patients with cancer. *J Natl Cancer Inst Monogr*. 2007;(37):39–46. doi:10.1093/jncimonographs/lgm003. Reprinted with permission from the Journal of the National Cancer Institute Monographs.
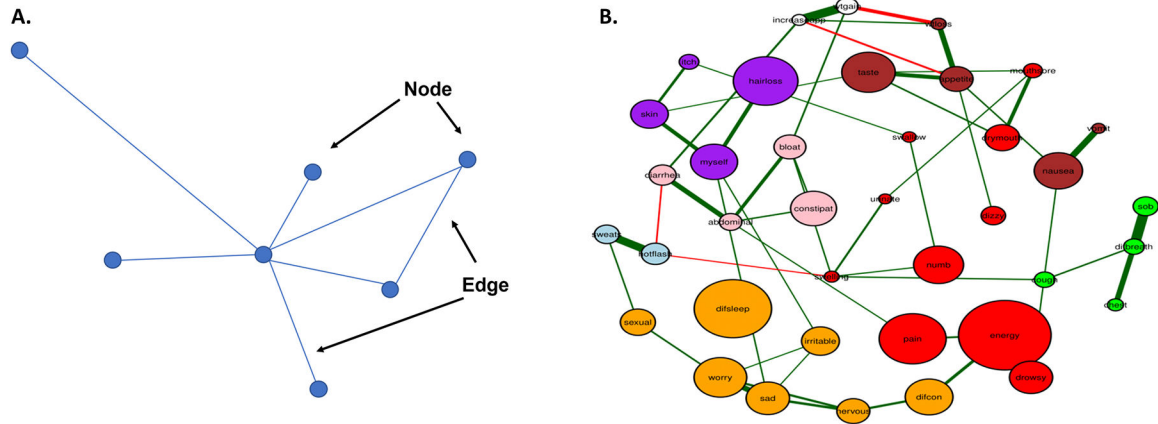
Figure 2.
A) An undirected graphical model with seven nodes. Each node represents a symptom. The presence of an edge between two nodes indicates a relationship between them. B) This figure represents the estimated network of 38 cancer symptoms across the "distress" symptom dimension. In this figure, the node size corresponds to the symptom distress scores and the strength of the relationship between nodes is illustrated by the thickness of the edges. Green edges indicate positive relationships and red edges indicate negative relationships. Symptom clusters were identified using a community detection algorithm and are identified by the color of the symptoms within each cluster. Adapted from Papachristou N, Barnaghi P, Cooper B, et al. Network analysis of the multidimensional symptom experience of oncology. Article. *Sci Rep*. 2019;9(1):2258. doi:10.1038/s41598-018-36973-1.

Table 1.

Areas of ongoing development in the definition of a symptom cluster

Symptom[*]	Symptom cluster Same characteristics as a symptom – plus:	Exemplars of areas for future research and development:
Subjective perception	Two or more concurrent symptoms	Consensus is needed on the specific characteristics that encompass the definition of a symptom cluster within and across acute and chronic conditions
May vary over time	Stable group of symptoms	The definition of and criteria for stability and consistency need to be established and evaluated. In addition, the conditions or circumstances when symptom clusters may or may not be stable warrants additional research (e.g., across symptom dimensions, within and across symptom dimensions over time)
Has antecedents	Independent of other clusters	The inter-relationships between and among symptoms and symptom clusters warrant detailed evaluation
Influences outcomes	May have shared underlying mechanism(s)	How do the mechanisms that underlie single symptoms within a cluster differ from mechanisms that underlie the entire cluster?
May be influenced by an intervention	May have shared outcome(s)	Do symptom clusters influence patient outcomes similarly or differently?
Has an underlying mechanism	Temporal dimension	When and how do symptom clusters change over time?

[*]
Symptoms are subjective sensations. Signs are objective indications of some medical characteristics.