Original Article

Congruence Between Latent Class and K-Modes Analyses in the Identification of Oncology Patients With Distinct Symptom Experiences

CrossMark

Nikoloas Papachristou, PhD(c), Payam Barnaghi, PhD, Bruce A. Cooper, PhD, Xiao Hu, PhD, Roma Maguire, BN, MSc, PhD, Kathi Apostolidis, Jo Armes, RN, PhD, Yvette P. Conley, PhD, Marilyn Hammer, PhD, Stylianos Katsaragakis, RN, PhD, Kord M. Kober, PhD, Jon D. Levine, MD, PhD, Lisa McCann, Bsc (Hon), MSc, PhD, Elisabeth Patiraki, RN, PhD, Steven M. Paul, PhD, Emma Ream, RN, PhD, Fay Wright, RN, PhD, and Christine Miaskowski, RN, PhD

School of Health Sciences (N.P., P.B., E.R.), University of Surrey, Guilford, UK; School of Nursing (B.A.C., X.H., K.M.K., S.M.P., C.M.), University of California, San Francisco, California, USA; Department of Computer and Information Sciences (R.M., L.M.), University of Strathclyde, Glasgow, UK; European Cancer Patient Coalition (K.A.), Brussels, Belgium; Florence Nightingale Faculty of Nursing and Midwifery (J.A.), King's College, London, UK; School of Nursing (Y.P.C.), University of Pittsburgh, Pittsburgh, Pennsylvania, USA; Department of Nursing (M.H.), Mount Sinai Medical Center, New York, New York, USA; Faculty of Nursing (S.K.), University of Peloponnese, Efstathiou & Stamatikis Valioti and Plateon, PC, Sparti, Greece; School of Medicine (J.D.L.), University of California, San Francisco, California, USA; School of Health Sciences (E.P.), National and Kapodistrian University of Athens, Athens, Greece; and School of Nursing (F.W.), Yale University, New Haven, Connecticut, USA

Abstract

Context. Risk profiling of oncology patients based on their symptom experience assists clinicians to provide more personalized symptom management interventions. Recent findings suggest that oncology patients with distinct symptom profiles can be identified using a variety of analytic methods.

Objectives. The objective of this study was to evaluate the concordance between the number and types of subgroups of patients with distinct symptom profiles using latent class analysis and K-modes analysis.

Methods. Using data on the occurrence of 25 symptoms from the Memorial Symptom Assessment Scale, that 1329 patients completed prior to their next dose of chemotherapy (CTX), Cohen's kappa coefficient was used to evaluate for concordance between the two analytic methods. For both latent class analysis and K-modes, differences among the subgroups in demographic, clinical, and symptom characteristics, as well as quality of life outcomes were determined using parametric and nonparametric statistics.

Results. Using both analytic methods, four subgroups of patients with distinct symptom profiles were identified (i.e., all low, moderate physical and lower psychological, moderate physical and higher Psychological, and all high). The percent agreement between the two methods was 75.32%, which suggests a moderate level of agreement. In both analyses, patients in the all high group were significantly younger and had a higher comorbidity profile, worse Memorial Symptom Assessment Scale subscale scores, and poorer QOL outcomes.

Conclusion. Both analytic methods can be used to identify subgroups of oncology patients with distinct symptom profiles. Additional research is needed to determine which analytic methods and which dimension of the symptom experience provide the most sensitive and specific risk profiles.   J Pain Symptom Manage 2018;55:318−333. © *2017 American Academy of Hospice and Palliative Medicine. Published by Elsevier Inc. All rights reserved.*

Key Words
Symptom clusters, cancer, latent class analysis, machine learning, clustering, chemotherapy, k-modes analysis

Introduction

Both clinical experience and research findings suggest that oncology patients experience significant interindividual variability in their symptom experience.[1,2] In the era of precision medicine,[3] which focuses on the identification of patients who are at greater risk for chronic conditions like cancer, it is imperative that the optimal methods to risk profile patients based on their symptom burden are identified. In two reviews of the state of the science in symptom clusters research,[4,5] it was noted that future studies need to focus on an evaluation of the concordance between the various analytic methods that can be used to identify patients who are at greatest risk for a higher symptom burden.

Recent findings from our group[6–14] and others[15–18] have identified subgroups of patients with distinct symptom experiences using approaches like hierarchical cluster analysis and latent class analysis (LCA). In the earliest of these studies,[6,7,15,16] different clustering methods were used to create the patient subgroups. In the later studies,[9–14,18] LCA was the preferred analytic approach. While across these 13 studies, the number of subgroups ranged from two to five, a common finding across all these studies was the identification of a group of patients who reported low levels of symptoms and a group of patients who reported high levels of symptoms. However, none of these studies determined whether the use of two different analytic approaches produces congruent results (e.g., the percentages of patients in the "all high" groups are equal and are the same patients).

As noted in a recent review,[5] machine learning techniques may provide useful approaches to identify subgroups of patients with distinct symptom profiles. Some specific machine learning techniques that can be used for this purpose include: K-means,[19] K-modes,[20,21] spectral clustering,[22] birch,[23] or agglomerative hierarchical clustering (AHC).[24,25] For binary variables (e.g., symptom occurrence), K-means and K-modes are two centroid-based algorithms that calculate the distance between each pair of data points using Euclidean distance or a simple dissimilarity measure (e.g., Hamming distance), respectively. The clusters derived from K-means and K-modes analyses are described by the "centroid," which is the multidimensional mean and mode, respectively, of the samples inside them.[19,21] Spectral clustering is a graph distance–based algorithm that performs a dimensionality reduction before clustering the lower dimension data set in a similar fashion to K-means. It is used when the clusters are not linearly separated in the original space, providing better results than algorithms such as K-means (which tends to find spherical clusters).[26] Birch is a hierarchical clustering algorithm that can provide an advantage in data sets that are nonuniformly distributed, and every data point is not equally important.

It concentrates on densely occupied partitions and follows a hierarchical order of analysis that focuses on calculating and updating measurements that capture the natural closeness of data. Therefore, it is more robust to "noise" (i.e., data points that are not part of the underlying pattern).[23] Finally, AHC is a decision tree, bottom-up clustering method that starts with every single data point in a single cluster. In each successive iteration, it agglomerates (merges) the closest pair of clusters by satisfying a similarity criterion, until all the data are in one cluster. A matrix tree plot visually demonstrates the hierarchy within the final cluster, where each merger is represented by a binary tree. AHC can be both informative for data display and helpful for the discovery of smaller clusters.[24]

No studies were identified that evaluated for congruence between two methods of classifying oncology patients based on their distinct experiences with common symptoms associated with cancer treatment. Based on how well the machine learning methods described previously performed during our initial analyses,[27] for this article, K-modes was selected as the method to compare with LCA. The purpose of this study, in a sample of patients ($n = 1329$) who were undergoing chemotherapy (CTX) for breast, lung, gastrointestinal, or gynecological cancers, was to evaluate the concordance between the number and types of subgroups of patients with distinct symptom experiences that were identified using LCA and K-modes analyses. We hypothesized that the number and types of subgroups would be similar using these two analytic methods.

Methods

Patients and Settings

This study is part of a longitudinal study of the symptom experience of oncology outpatients receiving CTX. The methods for this study are described in detail elsewhere.[13,28,29] According to the study's eligibility criteria: patients were ≥18 years of age; had a diagnosis of breast, gastrointestinal, gynecological, or lung cancer; had received CTX within the preceding four weeks; were scheduled to receive at least two additional cycles of CTX; were able to read, write, and understand English; and gave written informed consent. Patients were recruited from two Comprehensive Cancer Centers, one Veteran's Affairs hospital, and four community-based oncology programs.

Instruments

A demographic questionnaire obtained information on age, gender, ethnicity, marital status, living arrangements, education, employment status, and income. The Karnofsky Performance Status (KPS) scale[30] was used to evaluate patients' functional status. The Self-administered Comorbidity Questionnaire (SCQ)[31]

evaluated the occurrence, treatment, and functional impact of 13 common comorbid conditions (e.g., diabetes and arthritis).

A modified version of the Memorial Symptom Assessment Scale (MSAS) was used to evaluate the occurrence, severity, frequency, and distress of 38 symptoms commonly associated with cancer and its treatment. In this study, six symptoms were added to the original list of 32 MSAS symptoms (i.e., hot flashes, chest tightness, difficulty breathing, abdominal cramps, increased appetite, and weight gain). The MSAS is a self-report questionnaire designed to measure the multidimensional experience of symptoms. Patients were asked to indicate whether or not they had experienced each symptom in the past week (i.e., symptom occurrence). If they had experienced the symptom, they were asked to rate its frequency of occurrence, severity, and distress. The reliability and validity of the MSAS is well established in oncology patients.[32,33]

Three subscale scores (i.e., physical [MSAS-PHYS], psychological [MSAS-PSYCH], and Global Distress Index [MSAS-GDI]) were calculated. The MSAS-PHYS is the average of the frequency, severity, and distress ratings for 12 physical symptoms (i.e., lack of energy, feeling drowsy, pain, nausea, vomiting, change in the way food tastes, lack of appetite, dry mouth, constipation, feeling bloated, dizziness, and weight loss). The MSAS-PSYCH is the average of the frequency, severity, and distress ratings for six psychological symptoms (i.e., worrying, feeling sad, feeling nervous, feeling irritable, difficulty in sleeping, and difficulty concentrating). The MSAS-GDI is the average of the distress ratings for six physical symptoms (i.e., lack of energy, feeling drowsy, pain, lack of appetite, dry mouth, and constipation) and the frequency ratings for four psychological symptoms (i.e., worrying, feeling sad, feeling nervous, and feeling irritable).

Quality of life (QOL) was evaluated using disease-specific (i.e., Quality of Life Scale-Patient Version [QOL-PV])[34–36] and generic (i.e., Medical Outcomes Study-Short Form-12 [SF-12])[37] measures. The QOL-PV is a 41-item instrument that measures four dimensions of QOL (i.e., physical, psychological, social, and spiritual well-being) in oncology patients, as well as a total QOL score. Each item is rated on a 0 to 10 numeric rating scale with higher scores indicating a better QOL. The QOL-PV has established validity and reliability.[36,38–40]

The SF-12 consists of 12 questions that evaluate physical, mental, and overall health status. Individual items on the SF-12 are evaluated. In addition, the instrument is scored into physical component summary (PCS) and mental component summary (MCS) scores. These scores can range from 0 to 100. Higher PCS and MCS scores indicate a better QOL. The SF-12 has well-established validity and reliability.[37]

Study Procedures

The study was approved by the Committee on Human Research at the University of California, San Francisco, and by the Institutional Review Board at each of the study sites. Written informed consent was obtained from all patients. For this analysis, symptom occurrence data from the enrollment assessment, that asked patients to report on their symptom experience for the week prior to the administration of the next cycle of CTX, were analyzed (i.e., recovery from previous CTX cycle).

Data Analyses

Symptom Occurrence Data. In order to have a sufficient number of patients who endorsed each symptom, the LCA and K-modes analyses were done with the 25 symptoms that occurred in ≥30% of the patients (i.e., difficulty concentrating, pain, lack of energy, cough, feeling nervous, hot flashes, dry mouth, nausea, numbness or tingling in hands or feet, feeling drowsy, difficulty sleeping, feeling bloated, diarrhea, feeling sad, sweats, problems with sexual interest or activity, worrying, lack of appetite, dizziness, feeling irritable, hair loss, constipation, change in the way food tastes, I do not look like myself, and changes in skin).

Latent Class Analysis. LCA identifies latent classes based on an observed response pattern.[41,42] It is a statistical method for finding subtypes of related cases (i.e., latent classes) from multivariate categorical data. The LCA was performed using Mplus™, version 7.[43] Estimation was carried out with robust the maximum-likelihood and the expectation-maximization algorithms.[44] The optimal number of latent classes for this LCA was selected based on the Bayesian Information Criterion (BIC), the Vuong, Lo, Mendel, and Rubin (VLMR) likelihood ratio test, and entropy. Theoretically, the best fitting LCA model has the lowest BIC. Nevertheless, the BIC can be supplemented by an evaluation of the VLMR[45] which tests whether a model with K classes fits the data better than a model with one fewer class (the K-1 class model). When this VLMR is significant, the K-class model is considered to be a better fit for the data. When models are evaluated sequentially, with each new model having one more class than the previous model, if a model is identified for which the VLMR is not significant, then too many classes were extracted and the K-1 class model is considered to fit the data better than the current K-class model. Furthermore, well-fitting models produce entropy values of ≥0.80.[46] In addition, the optimal fitting model should "make sense"

conceptually and its classes should differ as might be expected on variables not used in the generation of the model.

K-Modes Analysis.   K-modes is a centroid method that is optimized for use with categorical variables.[21] It defines clusters based on the number of matching categories between data points and not on their Euclidean distance (a common similarity index in agglomerative clustering methods). Although its performance is comparable to K-means,[27] the K-modes distance measurement approach is theoretically a more appropriate approach to use to cluster the categorical variable of symptom occurrence.[21,47] The

K-modes analysis was implemented with PyCharm Professional Edition 4.5 and the Scikit-Learn library.[48]

The optimal number of clusters for the K-modes analysis was assessed using the Silhouette Coefficient (SC).[49] The SC represents how well each case (i.e., patient) lies within its cluster and how appropriate each case's assignment is inside a specific cluster. The average SC, called the Silhouette Index (SI), allows one to evaluate the overall quality of the separation between the clusters. The SC is calculated using its intra-cluster distance and its nearest cluster distance.[27] The SC is bounded between -1 for inappropriate clustering and $+1$ for highly compact clustering. An SC around zero indicates that a case is assigned inside

web 4C/FPO
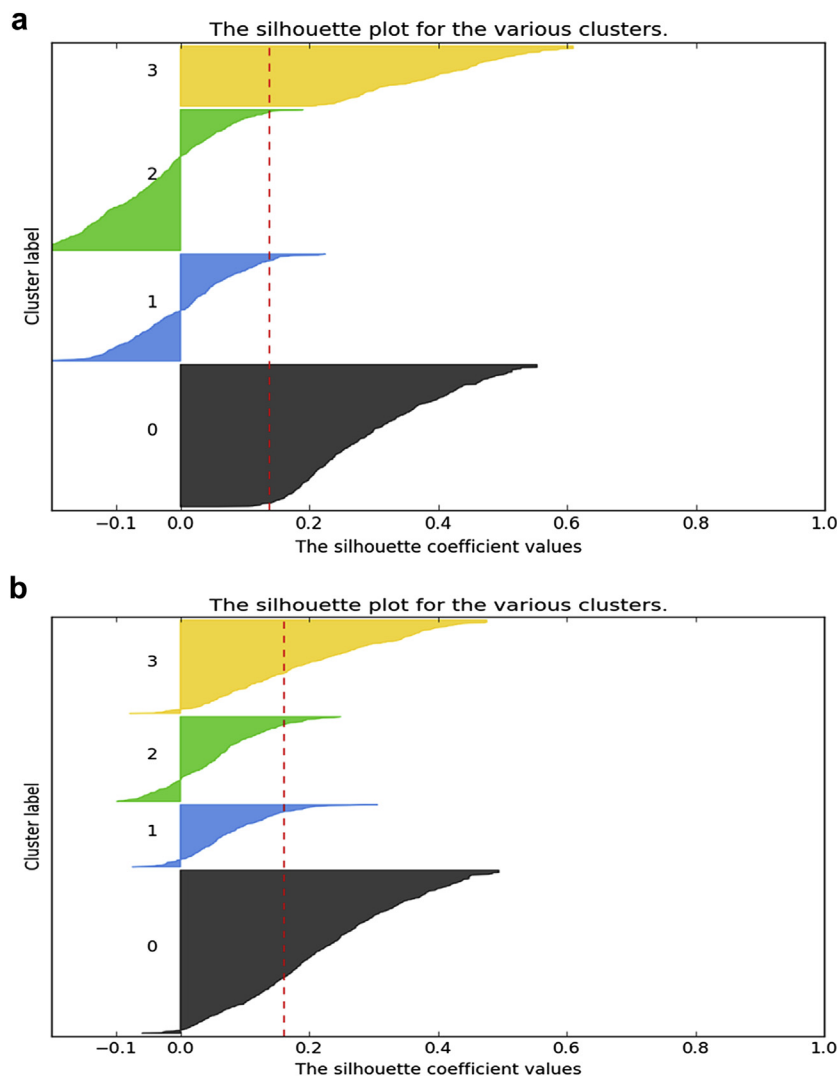
Fig. 1. a) Silhouette coefficient diagram for the four-class solution using latent class analysis. The sizes of the clusters in the diagram are proportional to their size inside the total sample of patients ($n = 1329$). The labels represent the following clusters: 0 (all low [$n = 419$, 31.5%]), 1 (moderate physical and lower psychological [$n = 316$, 23.8%]), 2 (moderate physical and higher psychological [$n = 416$, 31.3%]), and 3 (all high [$n = 178$, 13.4%]). b) Silhouette coefficient diagram for the four-cluster solution using the K-modes analysis. The sizes of the clusters in the diagram are proportional to their size inside the total sample of patients ($n = 1329$). The labels represent the following clusters: 0 (all low [$n = 536$, 40.3%]), 1 (moderate physical and lower psychological [$n = 205$, 15.4%]), 2 (moderate physical and higher psychological [$n = 280$, 21.1%]), and 3 (all high [$n = 308$, 23.2%]).

Model	LL	AIC	BIC	Entropy	VLMR
Three class	−10998.00	22150.00	22505.64	0.85	413.57[a]
Four class[b]	−10835.22	21876.44	22352.17	0.82	325.55[a]
Five class	−10765.09	21788.17	22383.99	0.81	140.27[NS]

AIC = Akaike's information criterion; BIC = Bayesian information criterion; LL = log-likelihood; NS = not significant; VLMR = Vuong-Lo-Mendell-Rubin likelihood ratio test for the K vs. K-1 model.
[a]$P < 0.001$.
[b]The four-class solution was selected because the BIC for that solution was lower than the BIC for both the three- and five-class solutions. In addition, the VLMR for the four-class solution indicates that it fits better than the three-class solution and the VLMR for the five-class solution does not fit better than the four-class solution.

overlapping clusters. In general, the average SI is high when clusters are dense and well separated.

Evaluation of Congruence. In order to evaluate the congruence between the LCA and K-modes solutions (i.e., number of subgroups identified), we compared the solutions using SC diagrams (see Figs. 1a and 1b, respectively).[49] When the SC for a case is > 0, its assignment to this cluster is considered appropriate. When the SC for a case is ≤ 0, this case may have equal similarities with cases in another, overlapping cluster, and its assignment inside a specific cluster may not be an appropriate fit. In addition, Cohen's kappa coefficient was used to evaluate the agreement between the two analytic approaches.

Differences in Demographic, Clinical, and Symptom Characteristics and QOL Outcomes. Descriptive statistics and frequency distributions were calculated for demographic and clinical characteristics using SPSS, version 23 (IBM, Armonk, NY). For each analytic approach, differences in demographic and clinical characteristics and QOL outcomes, among the groups, were evaluated using analyses of variance, Kruskal-Wallis, and chi-squared analyses. Post hoc contrasts were calculated using the Bonferroni corrected alpha of 0.008 (0.05/6 pairwise comparisons).

Results
Number of Subgroups Identified Using LCA and K-Modes Approaches
For the LCA, the fit indices for the candidate models are shown in Table 1. The four-class solution was selected because its BIC was lower than for the three- and five-class solutions. In addition, the VLMR indicated that a four-class solution was better than a three-class solution. However, the VLMR for the five-class solution was not better than the four-class solution indicating that too many classes were extracted.

Using K-modes, although the average SI for the three-class solution was slightly larger than the average SI for the four-class solution (Table 2), given this trivial difference and in order to compare the differences in demographic, clinical, and symptom characteristics and QOL outcomes between the two methods, we used the four-class solution from the K-modes analysis.

As shown in Figures 2 and 3, for the LCA and K-modes analyses, respectively, the four subgroups were named based on the probability of occurrence of the 25 MSAS symptoms that occurred in ≥30% of the patients. The all high and all low groups included patients who reported relatively high and low occurrence rates for most of the 25 MSAS symptoms, respectively. The moderate physical and higher psychological and moderate physical and lower psychological groups included patients who reported relatively moderate occurrence rates for the majority of the physical symptoms and relatively higher or lower occurrence rates, respectively, for the five psychological symptoms (i.e., worrying, feeling irritable, feeling sad, feeling nervous, and I do not look like myself).

The SC diagrams for all the patient cases within each of the four clusters for the LCA and K-modes analyses (Figs. 1a and 1b) showed that their inefficient assignments were mostly within two specific groups (i.e. moderate physical and higher psychological, moderate physical, and lower psychological). Both well-clustered (SC > 0) and inappropriately (SC ≤ 0) clustered cases were included within these clusters. As illustrated in the SC diagrams, K-modes assigned a larger proportion of cases to these two groups (SC > 0). Of note, the two other groups (all low and all high) were well defined and separated using both the LCA and K-modes approaches (SC > 0.4).

Table 2
K-Modes Solutions and Silhouette Indices for Three- Through Five-Class Solutions

Model	Silhouette Index
Three cluster[a]	0.159
Four cluster	0.156
Five cluster	0.129

[a]Based on the Silhouette Index, the three-cluster solution performed higher than both the four- and five-cluster solutions.

Fig. 2. Symptom occurrence for each of the subgroups identified using latent class analysis for the 25 symptoms on the Memorial Symptom Assessment Scale that occurred in ≥30% of the total sample (*n* = 1329) at time 1 (i.e., prior to next dose of chemotherapy).
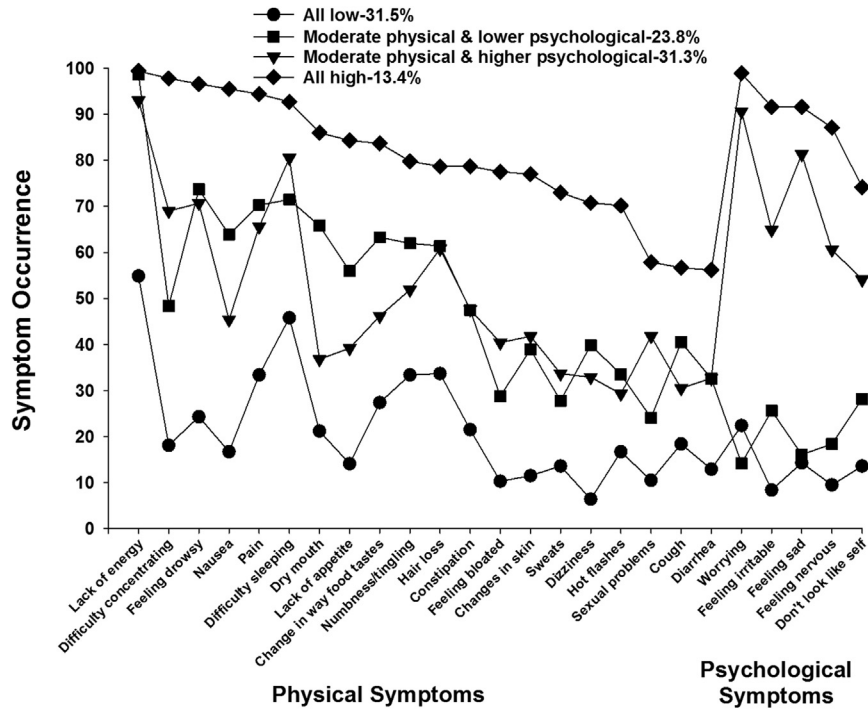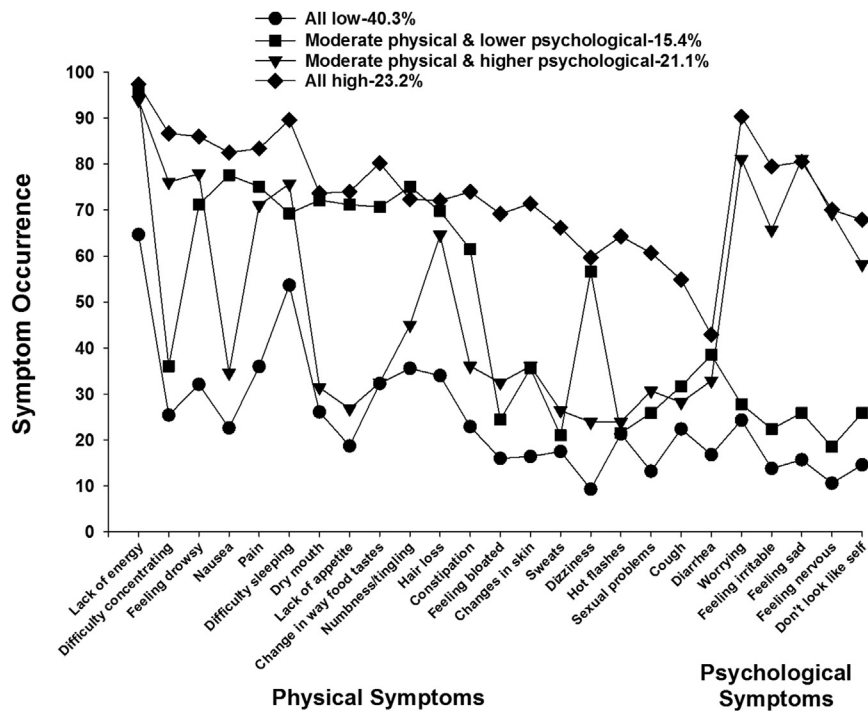

Fig. 3. Symptom occurrence for each of the subgroups identified using K-modes analysis for the 25 symptoms on the Memorial Symptom Assessment Scale that occurred in ≥30% of the total sample (*n* = 1329) at time 1 (i.e., prior to next dose of chemotherapy).

Pairwise Agreement Between the LCA and K-modes Approaches

As shown in Table 3, the observed agreement among the four groups was 75.32%, and the expected agreement was 26.08%. The two analyses separated patients into four distinct groups with substantial agreement beyond chance (range 0.6–0.7) as measured by the Cohen's coefficient (kappa = 0.666).[50] The biggest disagreements between the LCA and K-modes approaches were between: 1) the moderate physical and lower psychological (LCA) and all low (K-modes), and 2) the moderate physical and higher psychological (LCA) and all high (K-modes) groups, with 92 and 101 divergent classifications, respectively.

Group Characteristics Identified with LCA and K-modes Approaches

The all low group consisted of 31.5% (*n* = 419) of the sample using LCA and 40.3% (*n* = 536) using K-modes. The probability of occurrence of the MSAS symptoms for this group ranged from 0.064 to 0.549 for LCA and 0.093 to 0.647 for K-modes.

The second largest group identified using LCA was named moderate physical and higher psychological and consisted of 31.3% (*n* = 416) of the sample. Using K-modes, this group consisted of 21.1% (*n* = 280) of the patients. The occurrence rates for the majority of the physical symptoms ranged from 0.293 to 0.930 for LCA and from 0.236 to 0.939 for K-modes. For the psychological symptoms, the occurrence rates were relatively high. They ranged from 0.541 to 0.906 for LCA and from 0.582 to 0.811 for K-modes.

The third largest group identified using LCA (23.8%, *n* = 316) was named the moderate physical and lower psychological group. Using K-modes, this

group was the smallest one identified (15.4%, *n* = 205). The probability of occurrence for the physical symptoms ranged from 0.241 to 0.987 for LCA and from 0.210 to 0.956 for K-modes. For the psychological symptoms, the range was from 0.142 to 0.282 for LCA and from 0.185 to 0.278 for K-modes.

The all high group was the smallest one for LCA (13.4%, *n* = 178) and the second largest for the K-modes analysis (23.2%, *n* = 308). The probability of occurrence of the MSAS symptoms for this group ranged from 0.562 to 0.994 for LCA and from 0.429 to 0.974 for K-modes.

Differences in Patient Characteristics Among the Groups Identified with LCA and K-modes Approaches

Tables 4 and 5 summarize the differences in demographic and clinical characteristics among the four groups of patients identified using LCA and K-modes, respectively. For both analyses, compared to the "all low" group, patients in the "moderate physical and higher psychological" and the "all high" groups were significantly younger, had a lower KPS score, had a higher SCQ score, were more likely to have breast cancer, and were more likely to report depression and back pain. In addition, for both analyses, compared to the "moderate physical and lower psychological" group and the "moderate physical and higher psychological" group, patients in the "all high" group had a lower KPS score and a higher SCQ score.

Differences in Symptom Occurrence Rates Among the Groups Identified with LCA and K-modes

Supplemental Table 1 summarizes differences in symptom occurrence rates among the four groups of patients identified using LCA and K-modes.

Table 3
Pairwise Agreement Among the Patient Groups Using Latent Class Analysis and K-modes Analysis

Pairwise Agreement Among the Patient Groups	All Low[a] *n*[b] (%[c])	Moderate Physical and Lower Psychological *n*[b] (%[c])	Moderate Physical and Higher Psychological *n*[b] (%[c])	All High *n*[b] (%[c])	Total *n* (%)
All low[d]	406 (30.6)	4 (0.3)	9 (0.7)	0 (0.0)	419 (31.5)
Moderate physical and lower psychological	92 (6.9)	171 (12.9)	23 (1.7)	30 (2.3)	316 (23.8)
Moderate physical and higher psychological	38 (2.9)	30 (2.3)	247 (18.6)	101 (7.6)	416 (31.3)
All high	0 (0.0)	0 (0.0)	1 (0.1)	177 (13.3)	178 (13.4)
Total	536 (40.3)	205 (15.4)	280 (21.1)	308 (23.2)	1329 (100.0)

Agreement	Expected Agreement	Kappa	Standard Error	Z	P value
Cohen's kappa coefficient					
75.32%	26.08%	0.666	.016	42.64	<0.001

LCA = latent class analysis.
[a]For K-modes analysis: all low (*n* = 536, 40.3%), moderate physical and lower psychological (*n* = 205, 15.4%), moderate physical and higher psychological (*n* = 280, 21.1%), and all high (*n* = 308, 23.2%).
[b]Number of the patients who were included in both classes.
[c]Percentage of patients from the total sample of 1329 patients.
[d]For LCA: all low (*n* = 419, 31.5%), moderate physical and lower psychological (*n* = 316, 23.8%), moderate physical and higher psychological (*n* = 416, 31.3%), and all high (*n* = 178, 13.4%).

Table 4
Differences in Demographic and Clinical Characteristics Among the Patient Subgroups Using Latent Class Analysis

Characteristic	All Low $n = 419$ (31.5%) (0)	Moderate Physical and Lower Psychological $n = 316$ (23.8%) (1)	Moderate Physical and Higher Psychological $n = 416$ (31.3%) (2)	All High $n = 178$ (13.4%) (3)	Statistics
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
Age (y)	60.0 (11.2)	57.9 (12.3)	55.3 (12.9)	54.4 (12.0)	$F(3,1325) = 14.66$, $P < 0.001$ 0 and 1 > 2 and 3
Education (y)	16.3 (3.1)	16.2 (2.9)	16.4 (3.1)	15.5 (2.9)	$F(3,1298) = 4.28$, $P = 0.005$ 0 and 2 > 3
Body mass index (kg/m^2)	26.2 (5.5)	26.2 (5.9)	26.0 (5.4)	26.9 (6.4)	$F(3,1307) = 1.08$, $P = 0.358$
Karnofsky Performance Status score	85.8 (11.1)	79.4 (12.2)	78.0 (11.9)	72.3 (11.2)	$F(3,1271) = 62.75$, $P < 0.001$ 0 > 1, 2, and 3 1 and 2 > 3
Number of comorbidities	2.1 (1.3)	2.5 (1.4)	2.5 (1.4)	3.0 (1.6)	$F(3,1325) = 19.32$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
SCQ score	4.5 (2.6)	5.6 (3.2)	5.7 (3.1)	7.1 (4.0)	$F(3,1325) = 29.60$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
AUDIT score	3.1 (2.4)	2.5 (1.9)	3.1 (2.7)	3.1 (3.1)	$F(3,856) = 2.61$, $P = .05$
Time since cancer diagnosis (y)	1.8 (3.1)	2.1 (4.2)	2.1 (4.4)	1.9 (3.7)	KW = 2.64, $P = 0.478$
Time since cancer diagnosis (median)	0.42	0.41	0.44	0.45	
Number of prior cancer treatments	1.5 (1.5)	1.5 (1.5)	1.7 (1.5)	1.8 (1.5)	$F(3,1312) = 1.25$, $P = 0.290$
Number of metastatic sites including lymph node involvement	1.3 (1.2)	1.3 (1.3)	1.3 (1.3)	1.0 (1.1)	$F(3,1325) = 2.31$, $P = 0.075$
Number of metastatic sites excluding lymph node involvement	0.8 (1.0)	0.8 (1.1)	0.8 (1.1)	0.6 (0.9)	$F(3,1325) = 1.85$, $P = 0.136$
	% (n)	% (n)	% (n)	% (n)	
Gender					$\chi^2 = 48.63$, $P < 0.001$
Female[a]	67.8 (284)	76.3 (241)	83.7 (348)	89.9 (160)	0 < 2
Male	32.2 (135)	23.7 (75)	16.1 (67)	10.1 (18)	1 < 3
Transgender[b]	0.0 (0)	0.0 (0)	0.2 (1)	0.0 (0)	
Ethnicity					$\chi^2 = 22.96$, $P = 0.006$
White	70.6 (291)	66.6 (207)	75.1 (310)	61.9 (109)	2 < 3
Black	13.1 (54)	14.8 (46)	8.0 (33)	15.9 (28)	NS
Asian or Pacific Islander	7.5 (31)	9.3 (29)	5.6 (23)	6.8 (12)	1 and 3 < 2
Hispanic mixed or other	8.7 (36)	9.3 (29)	11.4 (47)	15.3 (27)	NS
Married or partnered (% yes)	67.7 (279)	64.3 (202)	64.0 (261)	57.4 (101)	$\chi^2 = 5.78$, $P = 0.123$
Lives alone (% yes)	20.9 (86)	20.4 (64)	21.0 (86)	26.6 (47)	$\chi^2 = 3.03$, $P = 0.387$
Childcare responsibilities (% yes)	18.5 (76)	21.3 (65)	22.2 (91)	31.0 (54)	$\chi^2 = 11.32$, $P = 0.010$ 0 < 3
Care of adult responsibilities (% yes)	5.2 (20)	8.8 (25)	9.6 (36)	8.9 (14)	$\chi^2 = 5.97$, $P = 0.113$
Currently employed (% yes)	40.0 (165)	34.4 (108)	35.9 (148)	23.3 (41)	$\chi^2 = 15.23$, $P = 0.002$ 0 and 2 > 3
Income					KW, $P < 0.001$ 0, 1, and 2 < 3
< $30,000+	14.4 (52)	18.3 (52)	15.9 (61)	33.1 (54)	
$30,000 to <$70,000	19.7 (71)	21.5 (61)	23.2 (89)	19.0 (31)	
$70,000 to < $100,000	18.9 (68)	16.2 (46)	15.4 (59)	16.0 (26)	
≥ $100,000	46.9 (169)	44.0 (125)	45.4 (174)	31.9 (52)	
Specific comorbidities (% yes)					
Heart disease	5.5 (23)	7.6 (24)	4.1 (17)	7.3 (13)	$\chi^2 = 4.91$, $P = 0.178$
High blood pressure	30.5 (128)	33.5 (106)	26.9 (112)	33.1 (59)	$\chi^2 = 4.48$, $P = 0.214$
Lung disease	9.8 (41)	12.7 (40)	10.6 (44)	14.0 (25)	$\chi^2 = 3.10$, $P = 0.377$
Diabetes	9.5 (40)	11.4 (36)	6.7 (28)	9.6 (17)	$\chi^2 = 4.97$, $P = 0.174$
Ulcer or stomach disease	2.9 (12)	4.4 (14)	5.3 (22)	9.0 (16)	$\chi^2 = 10.55$, $P = 0.014$ 0 < 3

(*Continued*)

Table 4
Continued

	% (n)	% (n)	% (n)	% (n)	
Kidney disease	0.7 (3)	1.6 (5)	2.2 (9)	1.1 (2)	$\chi^2 = 3.27$, $P = 0.351$
Liver disease	7.2 (30)	6.0 (19)	5.8 (24)	7.3 (13)	$\chi^2 = 0.98$, $P = 0.806$
Anemia or blood disease	7.2 (30)	13.6 (43)	14.4 (60)	17.4 (31)	$\chi^2 = 16.77$, $P = 0.001$ $0 < 1, 2,$ and 3
Depression	7.2 (30)	11.7 (37)	28.4 (118)	39.3 (70)	$\chi^2 = 119.64$, $P < 0.001$ 0 and $1 < 2$ and 3
Osteoarthritis	10.5 (44)	11.4 (36)	13.0 (54)	16.3 (29)	$\chi^2 = 4.32$, $P = 0.229$
Back pain	15.3 (64)	26.6 (84)	27.6 (115)	44.9 (80)	$\chi^2 = 59.15$, $P < 0.001$ $0 < 1, 2,$ and 3 1 and $2 < 3$
Rheumatoid arthritis	2.6 (11)	4.7 (15)	2.6 (11)	3.4 (6)	$\chi^2 = 3.28$, $P = 0.351$
Exercise on a regular basis (% yes)	73.2 (303)	68.8 (212)	74.9 (305)	59.6 (102)	$\chi^2 = 15.41$, $P = 0.002$ 0 and $2 > 3$
Smoking, current or history of (% yes)	34.2 (142)	37.1 (114)	36.3 (149)	32.6 (57)	$\chi^2 = 1.40$, $P = 0.706$
Cancer diagnosis					$\chi^2 = 34.25$, $P < 0.001$
Breast	32.9 (138)	39.9 (126)	45.7 (190)	44.9 (80)	$0 < 2$ and 3
Gastrointestinal	37.2 (156)	33.5 (106)	23.8 (99)	25.8 (46)	$0 > 2$ and 3; $1 > 2$
Gynecological	16.7 (70)	13.3 (42)	21.2 (88)	18.5 (33)	$1 < 2$
Lung	13.1 (55)	13.3 (42)	9.4 (39)	10.7 (19)	NS
Type of prior cancer treatment					KW, $P = 0.063$
No prior treatment	26.5 (108)	29.0 (89)	22.6 (91)	19.9 (35)	
Only surgery, CTX, or RT	41.0 (167)	41.7 (128)	42.5 (171)	43.8 (77)	
Surgery and CTX, or surgery and RT, or CTX and RT	20.6 (84)	17.3 (53)	21.9 (88)	18.2 (32)	
Surgery and CTX and RT	11.8 (48)	12.1 (37)	12.9 (52)	18.2 (31)	

AUDIT = Alcohol Use Disorders Identification Test; CTX = chemotherapy; kg = kilograms; KW = Kruskal-Wallis; m^2 = meter squared; NS = not significant; RT = radiation therapy; SCQ = Self-Administered Comorbidity Questionnaire.
[a]Reference group for the post hoc comparisons.
[b]Chi-squared analysis and post hoc contrasts done without the transgender patient include in the analyses.

Both analyses identified two groups of oncology patients who reported moderate levels of physical symptoms but differentiated on the occurrence of five psychological symptoms (i.e., worrying, feeling irritable, feeling sad, feeling nervous, and I do not look like myself). For patients in the moderate physical and higher psychological group, worrying (LCA: 0.906, K-modes: 0.811), feeling sad (LCA: 0.813, K-modes: 0.811), and feeling irritable (LCA: 0.649, K-modes: 0.657) were among the top symptoms. In contrast, in the moderate physical and lower psychological group, worrying (LCA: 0.142, K-modes: 0.278), feeling sad (LCA: 0.161, K-modes: 0.259), and feeling irritable (LCA: 0.256, K-modes: 0.224) were among the symptoms with the lowest probability of occurrences. The remaining psychological symptoms, namely: "feeling nervous" (moderate physical and higher psychological group: LCA: 0.606, K-modes: 0.693; moderate physical and lower psychological group: LCA: 0.184, K-modes: 0.185) and "I do not look like myself" (moderate physical and higher psychological group: LCA: 0.541, K-modes: 0.582; moderate physical and lower psychological group: LCA: 0.282, K-modes: 0.259) had significant differences between the aforementioned groups for both analyses.

Across all four groups, lack of energy was the most common symptom. While the probability of its occurrence for the total sample was 0.832, values ranged from 0.549 to 0.994 for LCA and from 0.647 to 0.974 for K-modes. In addition, pain (LCA: 0.944−0.334, K-modes: 0.834−0.360), difficulty in sleeping (LCA: 0.927−0.458, K-modes: 0.896−0.537), numbness/tingling in hands/feet (LCA: 0.798−0.334, K-modes: 0.724−0.356), change in the way food tastes (LCA: 0.837−0.274, K-modes: 0.802−0.323), and feeling drowsy (LCA: 0.966−0.243, K-modes: 0.860−0.321) occurred in the top 10 symptoms across all four groups for both analyses.

Differences in MSAS Summary Scores Among the Groups Identified with LCA and K-modes
Table 6 summarizes differences in the MSAS summary scores among the four groups of patients identified using LCA and K-modes. For the physical subscale, the psychological subscale, and the Global Distress Index, the differences among the four groups followed the same pattern for both analyses. For the MSAS total score, as well as for the total number of MSAS symptoms, the pattern observed using the LCA was in the expected direction (i.e., all low < moderate physical and lower psychological < moderate physical

Table 5
Differences in Demographic and Clinical Characteristics Among the Patient Subgroups Using K-Modes Analysis

Characteristic	All Low $n = 536$ (40.3%) (0) Mean (SD)	Moderate Physical & Lower Psychological $n = 205$ (15.4%) (1) Mean (SD)	Moderate Physical & Higher Psychological $n = 280$ (21.1%) (2) Mean (SD)	All High $n = 308$ (23.2%) (3) Mean (SD)	Statistics
Age (y)	59.6 (11.7)	58.1 (12.1)	55.3 (13.1)	54.4 (12.1)	$F(3,1325) = 15.10, P < 0.001$ 0 > 2 and 3 1 > 3
Education (y)	16.3 (3.1)	16.0 (2.9)	16.7 (3.0)	15.6 (2.9)	$F(3,1298) = 6.44, P < 0.001$ 0 > 2 and 3
Body mass index (kg/m^2)	26.2 (5.5)	26.3 (5.8)	25.8 (5.2)	26.7 (6.3)	$F(3,1307) = 1.26, P = 0.287$
Karnofsky Performance Status Score	85.0 (11.3)	77.8 (12.2)	78.6 (11.9)	74.2 (11.7)	$F(3,1271) = 59.38, P < 0.001$ 0 >1, 2, and 3 1 and 2 > 3
Number of comorbidities	2.1 (1.3)	2.6 (1.4)	2.4 (1.4)	2.9 (1.6)	$F(3,1325) = 20.27, P < 0.001$ 0 < 1 and 3 2 < 3
SCQ score	4.7 (2.7)	5.9 (3.1)	5.5 (3.0)	6.6 (3.8)	$F(3,1325) = 28.30, P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
AUDIT score	3.1 (2.2)	2.3 (1.9)	3.1 (2.7)	3.1 (2.9)	$F(3,856) = 3.92, P = .009$ 1 < 0, 2 and 3
Time since cancer diagnosis (y)	2.0 (3.8)	2.2 (4.0)	2.1 (4.3)	1.7 (3.6)	KW, $P = 0.831$
Time since cancer diagnosis (median)	0.42	0.40	0.45	0.42	
Number of prior cancer treatments	1.6 (1.5)	1.6 (1.5)	1.7 (1.5)	1.6 (1.5)	$F(3,1312) = 0.41, P = 0.748$
Number of metastatic sites including lymph node involvement	1.3 (1.2)	1.4 (1.3)	1.2 (1.2)	1.1 (1.2)	$F(3,1325) = 2.33, P = 0.073$
Number of metastatic sites excluding lymph node involvement	0.8 (1.0)	0.9 (1.1)	0.8 (1.1)	0.7 (1.0)	$F(3,1325) = 1.83, P = 0.140$
	% (*n*)	% (*n*)	% (*n*)	% (*n*)	
Gender					$\chi^2 = 50.10, P < 0.001$
Female[a]	69.6 (373)	74.1 (152)	83.9 (235)	88.6 (273)	0 < 2 and 3
Male	30.4 (163)	25.9 (53)	15.7 (44)	11.4 (35)	1 < 3
Transgender[b]	0.0 (0)	0.0 (0)	0.4 (1)	0.0 (0)	
Ethnicity					$\chi^2 = 24.93, P = 0.003$
White	71.2 (375)	60.7 (122)	77.8 (217)	66.6 (203)	1 and 3 < 2
Black	12.7 (67)	16.9 (34)	8.6 (24)	11.8 (36)	1 > 2
Asian or Pacific Islander	7.2 (38)	10.9 (22)	4.3 (12)	7.5 (23)	1 > 2
Hispanic mixed or other	8.9 (47)	11.4 (23)	9.3 (26)	14.1 (43)	NS
Married or partnered (% yes)	66.9 (354)	64.4 (130)	60.9 (167)	63.0 (192)	$\chi^2 = 3.16, P = 0.367$
Lives alone (% yes)	20.7 (109)	20.2 (41)	23.3 (64)	22.5 (69)	$\chi^2 = 1.12, P = 0.773$
Childcare responsibilities (% yes)	19.4 (102)	17.4 (34)	20.4 (57)	31.0 (93)	$\chi^2 = 19.01, P = 0.000$ 0, 1, and 2 < 3
Care of adult responsibilities (% yes)	6.1 (30)	9.9 (18)	8.3 (21)	9.4 (26)	$\chi^2 = 4.15, P = 0.246$
Currently employed (% yes)	38.9 (206)	36.0 (73)	37.5 (104)	25.9 (79)	$\chi^2 = 15.42, P = 0.001$ 0 and 2 > 3
Income					KW, $P = 0.001$ 0 and 2 < 3
< $30,000+	15.1 (70)	20.4 (38)	15.1 (39)	25.6 (72)	
$30,000 to <$70,000	19.8 (92)	21.0 (39)	22.8 (59)	22.1 (62)	
$70,000 to < $100,000	18.8 (87)	17.7 (33)	13.1 (34)	16.0 (45)	
≥ $100,000	46.3 (215)	40.9 (76)	49.0 (127)	36.3 (102)	
Specific comorbidities (% yes)					
Heart disease	6.3 (34)	7.3 (15)	4.6 (13)	4.9 (15)	$\chi^2 = 2.33, P = 0.507$
High blood pressure	30.4 (163)	36.1 (74)	25.7 (72)	31.2 (96)	$\chi^2 = 6.13, P = 0.106$
Lung disease	11.2 (60)	9.3 (19)	12.1 (34)	12.0 (37)	$\chi^2 = 1.21, P = 0.752$
Diabetes	8.8 (47)	15.1 (31)	5.7 (16)	8.8 (27)	$\chi^2 = 12.97, P = 0.005$ 1 > 2
Ulcer or stomach disease	3.4 (18)	4.9 (10)	3.9 (11)	8.1 (25)	$\chi^2 = 10.29, P = 0.016$ 0 < 3
Kidney disease	0.9 (5)	1.5 (3)	1.4 (4)	2.3 (7)	$\chi^2 = 2.49, P = 0.476$

(*Continued*)

Table 5
Continued

	% (n)	% (n)	% (n)	% (n)	
Liver disease	6.2 (33)	8.3 (17)	5.7 (16)	6.5 (20)	$\chi^2 = 1.48, P = 0.688$
Anemia or blood disease	8.6 (46)	15.1 (31)	9.3 (26)	19.8 (61)	$\chi^2 = 26.75, P < 0.001$ 0 and 2 < 3
Depression	7.5 (40)	13.7 (28)	28.6 (80)	34.7 (107)	$\chi^2 = 115.51, P < 0.001$ 0 and 1 < 2 and 3
Osteoarthritis	9.9 (53)	12.2 (25)	13.2 (37)	15.6 (48)	$\chi^2 = 6.20, P = 0.102$
Back pain	16.0 (86)	29.3 (60)	26.4 (74)	39.9 (123)	$X^2 = 60.12, P < 0.001$ 0 < 1, 2, and 3 2 < 3
Rheumatoid arthritis	3.2 (17)	3.9 (8)	1.8 (5)	4.2 (13)	$\chi^2 = 3.13, P = 0.372$
Exercise on a regular basis (% yes)	73.6 (388)	69.0 (138)	74.2 (204)	64.4 (192)	$\chi^2 = 9.73, P = 0.021$ 0 > 3
Smoking, current or history of (% yes)	35.5 (188)	34.0 (68)	37.9 (105)	33.6 (101)	$\chi^2 = 1.38, P = 0.710$
Cancer diagnosis					$\chi^2 = 43.25, P < 0.001$
Breast	34.9 (187)	37.1 (76)	47.1 (132)	45.1 (139)	0 < 2 and 3
Gastrointestinal	34.9 (187)	40.5 (83)	20.4 (57)	26.0 (80)	0 and 1 > 2 and 3
Gynecological	16.8 (90)	10.7 (22)	22.1 (62)	19.2 (59)	1 < 2
Lung	13.4 (72)	11.7 (24)	10.4 (29)	9.7 (30)	NS
Type of prior cancer treatment					
No prior treatment	25.9 (135)	30.8 (61)	20.1 (55)	24.0 (72)	KW, P = 0.226
Only surgery, CTX, or RT	41.7 (217)	37.9 (75)	44.3 (121)	43.3 (130)	
Surgery and CTX, or surgery and RT, or CTX and RT	20.0 (104)	19.7 (39)	22.7 (62)	17.3 (52)	
Surgery and CTX and RT	12.5 (65)	11.6 (23)	12.8 (35)	15.3 (46)	

AUDIT = Alcohol Use Disorders Identification Test; CTX = chemotherapy; kg = kilograms; KW = Kruskal-Wallis; m^2 = meter squared; NS = not significant; RT = radiation therapy; SCQ = Self-Administered Comorbidity Questionnaire.
[a]Reference group for the post hoc comparisons.
[b]Chi-squared analysis and post hoc contrasts done without the transgender patient include in the analyses.

and higher psychological < all high). For the MSAS total score, as well as for the total number of MSAS symptoms, the pattern observed using K-modes was as follows: all low < moderate physical and lower psychological, moderate physical and higher psychological and all high (i.e., 0 < 1, 2, and 3), as well as moderate physical and lower psychological and moderate physical and higher psychological < all high (i.e., 1 and 2 < 3).

Differences in QOL Scores Among the Groups Identified With LCA and K-Modes

Table 7 summarizes differences in MQOLS-CA subscale and total scores among the four groups of patients identified using LCA and K-modes. For the MQOLS psychological and social well-being subscales, and total QOL scores, the differences among the four groups followed the same pattern for both analyses (i.e., all low > moderate physical and lower psychological > moderate physical and higher psychological > all high). In addition, for the physical well-being subscale scores, the differences among the four groups followed the same pattern for both analyses (i.e., all low > moderate physical and lower psychological, moderate physical and higher psychological, and all high (i.e., 0 > 1, 2, and 3) and moderate physical and lower psychological and moderate physical and higher psychological > all high (i.e., 1 and 2 > 3).

For the SF12, for both analyses, the MCS scores followed a similar pattern (i.e., all low > moderate physical

and lower psychological > moderate physical and higher psychological > all high). For the PCS scores, the post hoc contrasts were different depending on the method of analysis. For LCA, the pattern was all low > moderate physical and higher psychological > moderate physical and lower psychological > all high. For the K-modes analysis, the pattern was as follows: all low > moderate physical and lower psychological, moderate physical and higher psychological and all high (i.e., 0 > 1, 2, and 3), as well as moderate physical and higher psychological > moderate physical and lower psychological and all high (i.e., 2 > 1 and 3).

Discussion

This study is the first to evaluate for congruence between the ability of two different analytic approaches to identify subgroups of oncology patients with distinct symptom profiles. Using both LCA and K-modes, four groups of patients with distinct symptom profiles were identified. The Cohen's kappa coefficient of 0.666 represents a moderate level of agreement between the two approaches.[51–53] Potential reasons for only a moderate level of agreement may be related to differences in the underlying assumptions of each of the methods. LCA is a model-based approach where "clusters" (i.e. classes) are defined by parametric probability distributions that can be interpreted to generate homogenous points, whereas the whole data set is modeled by a mixture

Table 6
Differences in Memorial Symptom Assessment Scale Scores Among the Patient Subgroups Using Latent Class Analysis or K-Modes Analysis

MSAS Scores	All Low[a,b] (0) Mean (SD)	Moderate Physical and Lower Psychological (1) Mean (SD)	Moderate Physical and Higher Psychological (2) Mean (SD)	All High (3) Mean (SD)	Statistics
Patient subgroups using latent class analysis					
Physical subscale	0.3 (0.2)	1.0 (0.4)	0.8 (0.4)	1.6 (0.5)	$F(3,1325) = 578.78, P < 0.001$ $0 < 2 < 1 < 3$
Psychological subscale	0.3 (0.3)	0.6 (0.4)	1.3 (0.5)	1.9 (0.6)	$F(3,1325) = 717.30, P < 0.001$ $0 < 1 < 2 < 3$
Global Distress Index	0.4 (0.3)	0.9 (0.4)	1.3 (0.5)	2.1 (0.6)	$F(3,1305) = 770.22, P < 0.001$ $0 < 1 < 2 < 3$
Total score	0.3 (0.2)	0.7 (0.3)	0.8 (0.3)	1.6 (0.4)	$F(3,1325) = 11,037.63, P < 0.001$ $0 < 1 < 2 < 3$
Total number of MSAS symptoms (out of 32)	5.6 (2.5)	12.9 (3.2)	14.6 (3.0)	23.0 (3.3)	$F(3,1325) = 1601.27, P = 0.000$ $0 < 1 < 2 < 3$
Total number of MSAS symptoms (out of 38)	6.3 (2.9)	14.4 (3.5)	16.1 (3.5)	26.1 (4.4)	$F(3,1325) = 1474.65, P < 0.001$ $0 < 1 < 2 < 3$
Patient subgroups using K-modes analysis					
Physical subscale	0.4 (0.3)	1.1 (0.4)	0.7 (0.4)	1.4 (0.6)	$F(3,1325) = 578.28, P < 0.001$ $0 < 2 < 1 < 3$
Psychological subscale	0.4 (0.3)	0.6 (0.4)	1.3 (0.5)	1.6 (0.7)	$F(3,1325) = 553.73, P < 0.001$ $0 < 1 < 2 < 3$
Global Distress Index	0.4 (0.3)	1.0 (0.4)	1.3 (0.5)	1.8 (0.6)	$F(3,1305) = 588.21, P < 0.001$ $0 < 1 < 2 < 3$
Total score	0.3 (0.2)	0.8 (0.3)	0.8 (0.3)	1.3 (0.4)	$F(3,1325) = 765.76, P < 0.001$ $0 < 1, 2,$ and 3 1 and $2 < 3$
Total number of MSAS symptoms (out of 32)	6.7 (3.2)	13.9 (2.8)	13.7 (2.8)	20.6 (4.1)	$F(3,1325) = 1187.40, P < 0.001$ $0 < 1, 2,$ and 3 1 and $2 < 3$
Total number of MSAS symptoms (out of 38)	7.6 (3.6)	15.2 (3.1)	15.0 (3.3)	23.2 (5.1)	$F(3,1325) = 1068.59, P < 0.001$ $0 < 1, 2,$ and 3 1 and $2 < 3$

LCA = latent class analysis; MSAS = Memorial Symptom Assessment Scale.
[a]For LCA: all low ($n = 419$, 31.5%), moderate physical and lower psychological ($n = 316$, 23.8%), moderate physical and higher psychological ($n = 416$, 31.3%), and all high ($n = 178$, 13.4%).
[b]For K-modes analysis: all low ($n = 536$, 40.3%), moderate physical and lower psychological ($n = 205$, 15.4%), moderate physical and higher psychological ($n = 280$, 21.13%), and all high ($n = 303$, 23.24%).

of such distributions.[54] Its key assumption is the conditional independence of the observed variables given the latent class. Inside the same class, the presence or the absence of one symptom is viewed as unrelated to the presence or absence of all of the others. On the other hand, K-modes is a distance-based clustering method that separates clusters as data subsets that have small within-cluster distances and large separation from other clusters. K-modes tries to find clusters that bring similar observations together without making an assumption about their distribution or attempt to fit a mixture distribution. Our findings, as well as others,[54–56] suggest that further research is needed, using both approaches, to determine the most sensitive and specific method(s) to risk profile oncology patients based on symptom occurrence rates.

While the absolute percentages of patients in the four groups differed depending on the analytic approach, the specific symptom profiles within each of the four groups were very similar. In addition, previous work in heterogeneous samples of oncology patients, using a different numbers of MSAS symptoms,[9,57] found the same four phenotypic profiles identified in the current study. Across these three studies, the percentage of patients in the all low group ranged from 28.0%[9] to 40.3% (using K-modes in the current study), and the percentage of patients in the all high class ranged from 13.4% (using LCA in the current study) to 27.8%.[57] Across these three studies, these relatively wide ranges may be related to differences in the number and types of symptoms evaluated, the timing of the symptom assessments in relationship to cancer diagnosis and treatments, and/or the specific cancer diagnoses of the patients in each of the studies. That said, these two extreme phenotypes were identified in previous studies that used only four symptoms[6,7,10,11] or identified only two or three groups.[15–17]

Across the two previous studies[9,57] and with the two analytic methods used in the current study, the consistent phenotypic characteristics associated with membership in the all high group were younger age and poorer functional status. The association between

Differences in Quality of Life Scores Among the Patient Subgroups Using Latent Class Analysis or K-Modes Analysis

QOL Scores	All Low[a,b] (0) Mean (SD)	Moderate Physical and Lower Psychological (1) Mean (SD)	Moderate Physical and Higher Psychological (2) Mean (SD)	All High (3) Mean (SD)	Statistics
Patient subgroups using latent class analysis					
MQOLS-PV—physical well-being	7.8 (1.4)	6.5 (1.5)	6.3 (1.5)	4.7 (1.6)	$F(3,1292) = 179.64$, $P < 0.001$ 0 > 1, 2, and 3 1 and 2 > 3
MQOLS-PV—psychological well-being	6.5 (1.6)	6.0 (1.6)	4.7 (1.6)	4.0 (1.5)	$F(3,1281) = 154.85$, $P < 0.001$ 0 > 1 > 2 > 3
MQOLS-PV—social well-being	6.9 (1.7)	6.0 (1.8)	5.1 (1.8)	4.1 (1.8)	$F(3,1274) = 123.13$, $P < 0.001$ 0 > 1 > 2 > 3
MQOLS-PV—spiritual well-being	5.5 (2.2)	5.5 (2.1)	5.3 (2.0)	5.6 (2.0)	$F(3,1286) = 0.61$, $P = 0.611$
MQOLS-PV—total QOL score	6.7 (1.2)	6.0 (1.2)	5.2 (1.2)	4.4 (1.2)	$F(3,1276) = 177.88$, $P < 0.001$ 0 > 1 > 2 > 3
SF12—PCS score	45.6 (9.6)	39.0 (10.1)	41.1 (10.5)	35.7 (9.7)	$F(3,1225) = 45.76$, $P < 0.001$ 0 > 2 > 1 > 3
SF12—MCS score	54.0 (8.4)	51.9 (8.5)	45.4 (9.8)	40.5 (11.1)	$F(3,1225) = 113.49$, $P < 0.001$ 0 > 1 > 2 > 3
Patient subgroups using k-modes analysis					
MQOLS-PV—physical well-being	7.6 (1.5)	6.3 (1.5)	6.5 (1.5)	5.2 (1.7)	$F(3,1292) = 153.99$, $P < 0.001$ 0 > 1, 2, and 3 1 and 2 > 3
MQOLS-PV—psychological well-being	6.4 (1.6)	5.9 (1.6)	4.7 (1.6)	4.3 (1.6)	$F(3,1281) = 128.41$, $P < 0.001$ 0 > 1 > 2 > 3
MQOLS-PV—social well-being	6.7 (1.8)	5.9 (1.8)	5.2 (1.8)	4.4 (1.8)	$F(3,1274) = 115.73$, $P < 0.001$ 0 > 1 > 2 > 3
MQOLS-PV—spiritual well-being	5.5 (2.1)	5.5 (2.1)	5.3 (2.0)	5.5 (2.0)	$F(3,1286) = 0.71$, $P = 0.547$
MQOLS-PV—total QOL score	6.5 (1.2)	5.9 (1.3)	5.3 (1.2)	4.7 (1.3)	$F(3,1276) = 152.38$, $P < .001$ 0 > 1 > 2 > 3
SF12—PCS score	44.8 (9.9)	38.1 (9.3)	41.6 (10.3)	37.0 (10.5)	$F(3,1225) = 43.78$, $P < 0.001$ 0 > 1, 2, and 3 2 > 1 and 3
SF12—MCS score	53.7 (8.3)	51.2 (9.0)	45.3 (10.3)	42.9 (10.5)	$F(3,1225) = 98.06$, $P < 0.001$ 0 > 1 > 2 > 3

MCS = Mental Component Summary; MQOLS-PV = Multidimensional Quality of Life Scale—Patient Version; PCS = Physical Component Summary; QOL = Quality of life; SF12 = Medical Outcomes Study-Short Form 12.
[a]For LCA: all low ($n = 419$, 31.5%), moderate physical and lower psychological ($n = 316$, 23.8%), moderate physical and higher psychological ($n = 416$, 31.3%), and all high ($n = 178$, 13.4%).
[b]For K-modes analysis: all low ($n = 536$, 40.3%), moderate physical and lower psychological ($n = 205$, 15.4%), moderate physical and higher psychological ($n = 280$, 21.13%), and all high ($n = 303$, 23.24%).

younger age and a higher symptom burden is consistent with previous studies.[6,7] While younger patients may receive more aggressive cancer treatments,[58] equally plausible hypotheses for this association include that older adults experience a "response shift" in their perception of symptoms[59]; that chronological age may not be an accurate representation of the biological age of oncology patients[60]; and/or that accelerated aging occurs with cancer and its treatment.[61–63]

Similar to age, the association between a higher symptom burden and poorer functional status was reported previously.[11,16,18] In the current study and in the one conducted in Norway,[57] that both used the KPS scale, compared to patients in the all low group who had KPS scores between 85 and 95, patients in the all high group reported KPS scores in the mid-70s. This difference represents a clinically meaningful change in functional status on this scale. Given that patients typically report lower KPS scores than their clinicians,[64,65] patients should be interviewed not only about the number and severity of their symptoms but also about changes in functional status during and following cancer treatment.

An equally important finding in this study and in the two previous studies[9,57] is the identification of two groups of patients who differentiated based on the occurrence of psychological symptoms. While our phenotypic data suggest that these two groups have lower KPS scores and a higher comorbidity profile than the all low group and better scores for both characteristics than the All High group, the demographic and clinical characteristics that distinguish between these two "moderate" groups are not readily apparent. These findings are similar to previous reports[9,57] and warrant investigation in future studies.

An evaluation of additional psychosocial characteristics (e.g., coping styles, personality, and social support) may improve the phenotypic characterization of these two "moderate" groups.

In terms of the QOL outcomes, regardless of whether a generic (i.e., SF12) or disease-specific (i.e., Multidimensional Quality of Life Scale—Patient Version) measure was used, the pattern of the differences in scores was in the expected direction, namely that as the symptom phenotype worsened, QOL decreased. The one interesting finding on Table 7 relates to the PCS scores from the SF12. While none of the groups had PCS scores of ≥50 (i.e., the normative value for the general population in the U.S.), patients in the moderate physical and lower psychological group had worse scores than patients in the moderate physical and higher psychological group. This finding is consistent with the report by Astrup et al.[57] Additional research is warranted to explain this finding and to determine the specific phenotypic characteristics that distinguish between these two moderate groups.

In terms of study limitations, patients were recruited at various points in their CTX treatment. In addition, the types of CTX were not homogeneous. While we cannot rule out the potential contributions of clinical characteristics to patients' symptom experiences, the relatively similar percentages of cancer diagnoses, reasons for current treatment, time since cancer diagnosis, and evidence of metastatic disease across the four groups, suggest that the patients were relatively similar in terms of disease and treatment characteristics. Although it is possible that patients in the "all low" group were receiving more aggressive symptom management interventions, the occurrence rates for the five most common symptoms were relatively similar across the four classes for both analyses. It is possible that using ratings of frequency, severity, or distress to create patient groups would provide additional information on interindividual differences in the symptom experience of these patients.

Additional research is warranted using different analytic methods to optimize the identification of oncology patients with a higher symptom burden. Future studies can evaluate different machine learning approaches, as well as real-time collection of different dimensions of a patient's symptom experience (i.e., occurrence, severity, and distress) to determine the most sensitive and specific methods to use to risk profile patients and design and test more effective symptom management interventions.

Acknowledgments

References

1. Esther Kim JE, Dodd MJ, Aouizerat BE, Jahan T, Miaskowski C. A review of the prevalence and impact of multiple symptoms in oncology patients. J Pain Symptom Manage 2009;37:715−736.

2. Reilly CM, Bruner DW, Mitchell SA, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer 2013;21:1525−1550.

3. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015;372:793−795.

4. Miaskowski C. Future directions in symptom cluster research. Semin Oncol Nurs 2016;32:405−415.

5. Miaskowski C, Barsevick A, Berger A, et al. Advancing symptom science through symptom cluster research: Expert panel proceedings and recommendations. J Natl Cancer Inst 2017;109.

6. Miaskowski C, Cooper BA, Paul SM, et al. Subgroups of patients with cancer with different symptom experiences and quality-of-life outcomes: a cluster analysis. Oncol Nurs Forum 2006;33:E79−E89.

7. Pud D, Ben Ami S, Cooper BA, et al. The symptom experience of oncology outpatients has a different impact on quality-of-life outcomes. J Pain Symptom Manage 2008;35: 162−170.

8. Illi J, Miaskowski C, Cooper B, et al. Association between pro- and anti-inflammatory cytokine genes and a symptom cluster of pain, fatigue, sleep disturbance, and depression. Cytokine 2012;58:437−447.

9. Miaskowski C, Dunn L, Ritchie C, et al. Latent class analysis reveals distinct subgroups of patients based on symptom occurrence and demographic and clinical characteristics. J Pain Symptom Manage 2015;50:28−37.

10. Langford DJ, Paul SM, Cooper B, et al. Comparison of subgroups of breast cancer patients on pain and co-occurring symptoms following chemotherapy. Support Care Cancer 2016;24:605−614.

11. Dodd MJ, Cho MH, Cooper BA, et al. Identification of latent classes in patients who are receiving biotherapy based on symptom experience and its effect on functional status and quality of life. Oncol Nurs Forum 2011;38:33−42.

12. 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 2015;17:237−247.

13. Miaskowski C, Cooper BA, Aouizerat B, et al. The symptom phenotype of oncology outpatients remains relatively stable from prior to through 1 week following chemotherapy. Eur J Cancer Care (Engl) 2016;26:e12437.

14. Miaskowski C, Cooper BA, Melisko M, et al. Disease and treatment characteristics do not predict symptom occurrence profiles in oncology outpatients receiving chemotherapy. Cancer 2014;120:2371−2378.

15. Ferreira KA, Kimura M, Teixeira MJ, et al. Impact of cancer-related symptom synergisms on health-related quality of life and performance status. J Pain Symptom Manage 2008;35:604−616.

16. Gwede CK, Small BJ, Munster PN, Andrykowski MA, Jacobsen PB. Exploring the differential experience of breast cancer treatment-related symptoms: a cluster analytic approach. Support Care Cancer 2008;16:925−933.

17. Reese JB, Blackford A, Sussman J, et al. Cancer patients' function, symptoms and supportive care needs: a latent class analysis across cultures. Qual Life Res 2015;24:135−146.

18. Snyder CF, Garrett-Mayer E, Blackford AL, et al. Concordance of cancer patients' function, symptoms, and supportive care needs. Qual Life Res 2009;18:991−998.

19. Arthur D, Vassilvitskii S. k-means++: the advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete Algorithms 20071027−1035.

20. Cao F, Liang J, Bai L. A new initialization method for categorical data clustering. Expert Syst Appl 2009;36: 10223−10228.

21. Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min Knowl Discov 1998;2:283−304.

22. Ng AY, Jordan MI, Weiss Y. On spectral clustering: analysis and an algorithm. Adv Neural Inf Process Syst 2002;2: 849−856.

23. Zhang T, Ramakrishnan R, Livny M. BIRCH: an efficient data clustering method for very large databases. ACM SIGMOD Rec 1996;25:103−114.

24. Sasirekha K, Baby P. Agglomerative hierarchical clustering algorithm−a review. Int J Sci Res Pub 2013;3:1−3.

25. Zhao Y, Karypis G, Fayyad U. Hierarchical clustering algorithms for document datasets. Data Min Knowl Discov 2005;10:141−168.

26. Dhillon IS, Guan Y, Kulis B. A unified view of kernel k-means, spectral clustering and graph cuts. Austin, TX: University of Austin. UTCS Technical Report TR-04−25 2005.

27. Papachristou N, Miaskowski C, Barnaghi P, et-al. Comparing machine learning clustering with latent class analysis on cancer symptoms' data. New York: IEEE. Proceedings of the IEEE Healthcare Innovation Point-of-Care Technologies Conference 2016.

28. Wright F, Hammer M, Paul SM, et al. Inflammatory pathway genes associated with inter- individual variability in the trajectories of morning and evening fatigue in patients receiving chemotherapy. Cytokine 2017;91:187−210.

29. Kober KM, Cooper BA, Paul SM, et al. Subgroups of chemotherapy patients with distinct morning and evening fatigue trajectories. Support Care Cancer 2016;24: 1473−1485.

30. Karnofsky D, Abelmann WH, Craver LV, Burchenal JH. The use of nitrogen mustard in the palliative treatment of cancer. Cancer 1948;1:634−656.

31. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003;49:156−163.

32. Portenoy RK, Thaler HT, Kornblith AB, et al. Symptom prevalence, characteristics and distress in a cancer population. Qual Life Res 1994;3:183−189.

33. Portenoy RK, Thaler HT, Kornblith AB, et al. The Memorial Symptom Assessment Scale−an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer 1994;30a:1326−1336.

34. Ferrell BR, Wisdom C, Wenzl C. Quality of life as an outcome variable in the management of cancer pain. Cancer 1989;63:2321−2327.

35. Padilla GV, Grant MM. Quality of life as a cancer nursing outcome variable. Adv Nurs Sci 1985;8:45−60.

36. Padilla GV, Presant C, Grant MM, et al. Quality of life index for patients with cancer. Res Nurs Health 1983;6: 117−126.

37. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220−233.

38. Padilla GV, Ferrell B, Grant MM, Rhiner M. Defining the content domain of quality of life for cancer patients with pain. Cancer Nurs 1990;13:108−115.

39. Ferrell BR, Dow KH, Grant M. Measurement of the quality of life in cancer survivors. Qual Life Res 1995;4:523−531.

40. Ferrell BR. The impact of pain on quality of life. A decade of research. Nurs Clin North Am 1995;30:609−624.

41. Collins LM, Lanza ST. Latent class and latent transition analysis: With applications in the social, behavioral, and health science. Hoboken, NJ: John Wiley & Sons, 2010.

42. Nylund K, Bellmore A, Nishina A, Graham S. Subtypes, severity, and structural stability of peer victimization: what does latent class analysis say? Child Dev 2007;78:1706−1722.

43. Muthen LK, Muthen BO. Mplus (Version 7.4). Los Angeles, CA: Muthen & Muthen, 2015.

44. Muthen B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 1999;55:463−469.

45. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Modeling 2007;14:535−569.

46. Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif 1996;13:195−212.

47. Ordonez C. Clustering binary data streams with K-means. San Diego, CA: AGM SIGMOD. Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. 2003:12-19.

48. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825−2830.

49. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53−65.

50. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159−174.

51. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Social Adm Pharm 2013;9:330−338.

52. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22:276−282.

53. Steinijans VW, Diletti E, Bomches B, Greis C, Solleder P. Interobserver agreement: Cohen's kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. Int J Clin Pharmacol Ther 1997; 35:93−95.

54. Anderlucci L, Hennig C. The clustering of categorical data: a comparison of a model- based and a distance-based approach. Commun Stat Theory Methods 2014;43:704−721.

55. Hennig C, Liao TF. How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. J R Stat Soc Ser C 2013;62:309−369.

56. Oberski DL. Beyond the number of classes: separating substantive from non-substantive dependence in latent class analysis. Adv Data Anal Classif 2016;10:171−182.

57. Astrup GL, Hofso K, Bjordal K, et al. Patient factors and quality of life outcomes differ among four subgroups of oncology patients based on symptom occurrence. Acta Oncol 2017;56:462−470.

58. Klepin HD, Rodin M, Hurria A. Treating older adults with cancer: geriatric perspectives. Am Soc Clin Oncol Educ Book 2015;35:e544−e552.

59. Sprangers MA, Schwartz CE. The challenge of response shift for quality-of-life-based clinical oncology research. Ann Oncol 1999;10:747−749.

60. Bae CY, Kang YG, Piao MH, et al. Models for estimating the biological age of five organs using clinical biomarkers that are commonly measured in clinical practice settings. Maturitas 2013;75:253−260.

61. Henderson TO, Ness KK, Cohen HJ. Accelerated aging among cancer survivors: from pediatrics to geriatrics. Am Soc Clin Oncol Educ Book 2014;e423−e430.

62. Hurria A, Jones L, Muss HB. Cancer treatment as an accelerated aging process: assessment, biomarkers, and interventions. Am Soc Clin Oncol Educ Book 2016;35: e516−e522.

63. Ness KK, Krull KR, Jones KE, et al. Physiologic frailty as a sign of accelerated aging among adult survivors of childhood cancer: a report from the St Jude Lifetime cohort study. J Clin Oncol 2013;31:4496−4503.

64. Schnadig ID, Fromme EK, Loprinzi CL, et al. Patient-physician disagreement regarding performance status is associated with worse survivorship in patients with advanced cancer. Cancer 2008;113:2205−2214.

65. Ando M, Ando Y, Hasegawa Y, et al. Prognostic value of performance status assessed by patients themselves, nurses, and oncologists in advanced non-small cell lung cancer. Br J Cancer 2001;85:1634−1639.

Appendix

Supplemental Table 1
Differences in Symptom Occurrence Rates Among the Patient Subgroups Using Latent Class Analysis or K-Modes Analysis

Symptom	Method	All low[a,b] (0)	Moderate physical and lower psychological (1)	Moderate physical and higher psychological (2)	All high (3)	Statistics
Physical symptoms						
Lack of energy	LCA	54.9	98.7	93.0	99.4	$\chi^2 = 357.44, P < 0.001$ 0 < 1, 2, and 3 1 > 2; 2 < 3
	K-modes	64.7	95.6	93.9	97.4	$\chi^2 = 221.00, P < 0.001$ 0 < 1, 2, and 3
Difficulty concentrating	LCA	18.1	48.4	69.0	97.8	$\chi^2 = 391.45, P < 0.001$ 0 < 1 < 2 < 3
	K-modes	25.4	36.1	76.1	86.7	$\chi^2 = 386.45, P < 0.001$ 0 < 1 < 2 < 3
Feeling drowsy	LCA	24.3	73.7	70.7	96.6	$\chi^2 = 366.85, P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	32.1	71.2	77.9	86.0	$\chi^2 = 309.61, P < 0.001$ 0 < 1, 2, and 3 1 < 3
Nausea	LCA	16.7	63.9	45.4	95.5	$\chi^2 = 358.73, P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	22.6	77.6	34.6	82.5	$\chi^2 = 377.42, P < 0.001$ 0 < 1, 2, and 3 1 > 2; 2 < 3
Pain	LCA	33.4	70.3	65.6	94.4	$\chi^2 = 231.14, P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	36.0	75.1	71.1	83.4	$\chi^2 = 233.65, P < 0.001$ 0 < 1, 2, and 3 2 < 3
Difficulty sleeping	LCA	45.8	71.5	80.5	92.7	$\chi^2 = 178.97, P < 0.001$ 0 < 1 < 2 < 3
	K-modes	53.7	69.3	75.7	89.6	$\chi^2 = 125.66, P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
Dry mouth	LCA	21.2	65.8	36.8	86.0	$\chi^2 = 282.43, P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	26.1	72.2	31.4	73.7	$\chi^2 = 261.36, P < 0.001$ 0 < 1 and 3 1 > 2; 2 < 3
Lack of appetite	LCA	14.1	56.0	39.2	84.3	$\chi^2 = 292.58, P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	18.7	71.2	26.8	74.0	$\chi^2 = 349.43, P < 0.001$ 0 < 1 and 3 1 > 2; 2 < 3
Change in the way food tastes	LCA	27.4	63.3	46.2	83.7	$\chi^2 = 190.76, P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	32.3	70.7	32.5	80.2	$\chi^2 = 249.04, P < 0.001$ 0 < 1 and 3 1 > 2; 2 < 3
Numbness/tingling in hands/feet	LCA	33.4	62.0	51.9	79.8	$\chi^2 = 125.76, P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	35.6	75.1	45.0	72.4	$\chi^2 = 158.32, P < 0.001$ 0 < 1 and 3 1 > 2; 2 < 3
Hair loss	LCA	33.7	61.4	60.8	78.7	$\chi^2 = 128.15, P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	34.0	69.8	64.6	72.1	$\chi^2 = 160.60, P < 0.001$ 0 < 1, 2, and 3

(*Continued*)

Supplemental Table 1
Continued

Symptom	Method	All low[a,b] (0)	Moderate physical and lower psychological (1)	Moderate physical and higher psychological (2)	All high (3)	Statistics
Constipation	LCA	21.5	47.5	47.6	78.7	$\chi^2 = 177.03$, $P < 0.001$ 0 < 1, 2, and 3 1 < 2 and 3
	K-modes	22.9	61.5	36.1	74.0	$\chi^2 = 242.11$, $P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
Feeling bloated	LCA	10.3	28.8	40.4	77.5	$\chi^2 = 269.93$, $P < 0.001$ 0 < 1 < 2 < 3
	K-modes	16.0	24.4	32.5	69.2	$\chi^2 = 258.27$, $P < 0.001$ 0 < 2 and 3 1 and 2 < 3
Changes in skin	LCA	11.5	38.9	41.8	77.0	$\chi^2 = 245.68$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	16.4	35.6	36.1	71.4	$\chi^2 = 256.15$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
Sweats	LCA	13.6	27.8	33.7	73.0	$\chi^2 = 208.28$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	17.5	21.0	26.4	66.2	$\chi^2 = 235.56$, $P < 0.001$ 0 < 2 and 3 1 and 2 < 3
Dizziness	LCA	6.4	39.9	32.9	70.8	$\chi^2 = 260.76$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	9.3	56.6	23.6	59.7	$\chi^2 = 304.91$, $P < 0.001$ 0 < 1, 2, and 3 1 > 2; 2 < 3
Hot flashes	LCA	16.7	33.5	29.3	70.2	$\chi^2 = 166.73$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	21.3	21.5	23.9	64.3	$\chi^2 = 195.29$, $P < 0.001$ 0, 1, and 2 < 3
Problems with sexual interest or activity	LCA	10.5	24.1	41.8	57.9	$X^2 = 175.13$, $P < 0.001$ 0 < 1 < 2 < 3
	K-modes	13.2	25.9	30.7	60.7	$\chi^2 = 212.26$, $P < 0.001$ 0 < 1 and 3 1 and 2 < 3
Cough	LCA	18.4	40.5	30.5	56.7	$\chi^2 = 95.62$, $P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	22.4	31.7	28.2	54.9	$\chi^2 = 97.52$, $P < 0.001$ 0, 1, and 2 < 3
Diarrhea	LCA	12.9	32.6	32.7	56.2	$\chi^2 = 119.84$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	16.8	38.5	32.9	42.9	$\chi^2 = 77.50$, $P < 0.001$ 0 < 1, 2, and 3
Psychological symptoms Worrying	LCA	22.4	14.2	90.6	98.9	$\chi^2 = 732.68$, $P < 0.001$ 0 > 1 0 and 1 < 2 and 3; 2 < 3
	K-modes	24.3	27.8	81.1	90.3	$\chi^2 = 488.89$, $P < 0.001$ 0 and 1 < 2 and 3 2 < 3
Feeling irritable	LCA	8.4	25.6	64.9	91.6	$\chi^2 = 500.74$, $P < 0.001$ 0 < 1 < 2 < 3
	K-modes	13.8	22.4	65.7	79.5	$\chi^2 = 451.86$, $P < 0.001$ 0 < 2 < 3 1 < 2 and 3
Feeling sad	LCA	14.3	16.1	81.3	91.6	$\chi^2 = 639.54$, $P < 0.001$ 0 and 1 < 2 and 3 2 < 3
	K-modes	15.7	25.9	81.1	80.5	$\chi^2 = 518.29$, $P < 0.001$ 0 < 1, 2, and 3 1 < 2 and 3

Supplemental Table 1
Continued

Symptom	Method	All low[a,b] (0)	Moderate physical and lower psychological (1)	Moderate physical and higher psychological (2)	All high (3)	Statistics
Feeling nervous	LCA	9.5	18.4	60.6	87.1	$\chi^2 = 467.74$, $P < 0.001$ $0 < 1 < 2 < 3$
	K-modes	10.6	18.5	69.3	70.1	$X^2 = 454.62$, $P < 0.001$ 0 < 1, 2, and 3 1 < 2 and 3
I do not look like myself	LCA	13.6	28.2	54.1	74.2	$\chi^2 = 263.69$, $P < 0.001$ $0 < 1 < 2 < 3$
	K-modes	14.6	25.9	58.2	67.9	$X^2 = 303.48$, $P < 0.001$ 0 < 1, 2, and 3 1 < 2 and 3
Symptoms not included formation of patient groups						
Difficulty swallowing	LCA	4.1	18.7	10.6	35.4	$\chi^2 = 113.35$, $P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	4.7	21.5	8.2	29.5	$\chi^2 = 119.48$, $P < 0.001$ 0 < 1 and 3 1 > 2; 2 < 3
Abdominal cramps	LCA	8.1	21.5	21.9	59.6	$\chi^2 = 190.13$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	10.3	22.4	20.7	45.5	$\chi^2 = 139.63$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
Shortness of breath	LCA	10.3	30.1	28.1	57.3	$\chi^2 = 144.71$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	14.4	30.7	28.2	44.8	$\chi^2 = 94.90$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
Difficulty breathing	LCA	7.4	22.8	17.5	50.0	$\chi^2 = 145.13$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	10.4	21.0	18.9	36.7	$\chi^2 = 84.69$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
Chest tightness	LCA	5.5	16.8	18.0	48.3	$\chi^2 = 156.69$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	6.9	16.6	18.2	37.3	$\chi^2 = 123.91$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
Itching	LCA	12.2	29.1	24.8	47.2	$\chi^2 = 86.76$, $P < 0.001$ 0 < 1, 2, and 3 1 < 3
	K-modes	14.6	27.3	24.3	41.6	$\chi^2 = 77.24$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
Weight loss	LCA	13.1	30.4	24.3	46.6	$\chi^2 = 80.44$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	14.2	40.0	17.1	41.9	$\chi^2 = 113.46$, $P < 0.001$ 0 < 1 and 3 1 > 2; 2 < 3
Increased appetite	LCA	16.0	23.7	28.8	46.1	$\chi^2 = 61.84$, $P < 0.001$ 0 < 2 and 3 1 and 2 < 3
	K-modes	18.3	22.4	25.4	41.9	$\chi^2 = 58.54$, $P < 0.001$ 0, 1 and 2 < 3
Mouth sores	LCA	9.1	21.8	24.3	39.3	$\chi^2 = 75.02$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	10.3	23.9	22.5	36.0	$\chi^2 = 80.90$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3

(Continued)

Supplemental Table 1
Continued

Symptom	Method	All low[a,b] (0)	Moderate physical and lower psychological (1)	Moderate physical and higher psychological (2)	All high (3)	Statistics
Problems with urination	LCA	5.7	12.3	13.2	38.8	$\chi^2 = 114.92$, $P < 0.001$ 0 < 1, 2, and 3 1 and 2 < 3
	K-modes	6.9	12.2	11.8	29.9	$\chi^2 = 88.17$, $P < 0.001$ 0, 1, and 2 < 3
Weight gain	LCA	16.2	25.9	28.8	37.6	$\chi^2 = 35.37$, $P < 0.001$ 0 < 1, 2, and 3
	K-modes	18.1	23.9	28.2	36.4	$\chi^2 = 36.08$, $P < 0.001$ 0 < 2 and 3 1 < 3
Vomiting	LCA	4.1	16.5	9.6	30.9	$\chi^2 = 91.05$, $P < 0.001$ 0 < 1, 2, and 3 1 > 2; 1 and 2 < 3
	K-modes	5.8	19.0	8.9	22.4	$\chi^2 = 61.61$, $P < 0.001$ 0 < 1 and 3 1 > 2; 2 < 3
Swelling in the arms and/or legs	LCA	7.6	17.1	14.7	26.4	$\chi^2 = 37.76$, $P < 0.001$ 0 < 1, 2, and 3 2 < 3
	K-modes	8.8	16.1	15.0	23.4	$\chi^2 = 34.06$, $P < 0.001$ 0 < 1 and 3

LCA = latent class analysis.

[a]For LCA: all low ($n = 419$, 31.5%), moderate physical and lower psychological ($n = 316$, 23.8%), moderate physical and higher psychological ($n = 416$, 31.3%), and all high ($n = 178$, 13.4%).

[b]For K-modes analysis: all low ($n = 536$, 40.3%), moderate physical and lower psychological ($n = 205$, 15.4%), moderate physical and higher psychological ($n = 280$, 21.13%), and all high ($n = 303$, 23.24%).