Symptom Cluster Research with Biomarkers and Genetics Using Latent Class Analysis

Samantha Conley, PhD, FNP-BC
Post-Doctorate Fellow, Yale University School of Nursing, Yale University School of Nursing, PO Box 27399, West Haven, CT 06516, Phone: 203-737-5129, Fax: 203.737.4480

Abstract

The purpose of this article is to provide an overview of latent class analysis (LCA) and examples from symptom cluster research that includes biomarkers and genetics. A review of LCA with genetics and biomarkers was conducted using Medline, Embase, PubMed, and Google Scholar. LCA is a robust latent variable model used to cluster categorical data and allows for the determination of empirically determined symptom clusters. Researchers should consider using LCA to link empirically determined symptom clusters to biomarkers and genetics to better understand the underlying etiology of symptom clusters. The full potential of LCA in symptom cluster research has not yet been realized because it has been used in limited populations and researchers have explored limited biologic pathways.

Keywords

Latent class analysis; symptom clusters; biomarkers; genetics

---

Symptoms rarely occur in isolation, and may occur in clusters [1]. Symptom clusters are two or more related symptoms that occur together. They may or may not share the same etiology [2]. Several statistical methods have been used to determine symptom clusters including correlations, factor analysis, principal components analysis, cluster analysis, and latent class analysis [3]. There is no single ideal statistical method for identifying symptom clusters, the choice of a statistical method must be carefully considered and based on the research question and the theoretical foundation of the study [3,4].

Latent class analysis (LCA) is a robust categorical statistical method that is well suited to answer many questions pertinent to nursing and symptom cluster research [3]. LCA allows researchers to empirically determine symptom clusters. LCA has recently gained attention and use in social science research [5]. However, LCA has been largely overlooked by nurse researchers for more familiar clustering techniques, such as factor analysis, cluster analysis, and structural equation modeling, even when LCA is better suited to answer the proposed research questions [6,7].

The author reports no conflicts of interest

Medline, Embase, PubMed, and Google Scholar were searched to find publications that provide guidance for using LCA and examples of LCA being used in symptom cluster research. The reference lists from the relevant publication also were reviewed to uncover additional publications. An overview of LCA is presented and examples from symptom cluster research that includes biomarkers and genetics are presented as well.

Conceptual Views of Symptom Clusters

There are two to conceptual views of symptom clusters: the grouping of variables (variable-oriented) and the grouping of people (person-oriented) [4]. A variable-oriented approach focuses on identifying relationships between variables and it is assumed that these relationships are stable across the population. Traditional factor analysis is an example of a variable-oriented clustering approach, and it is useful to determine relationships in a homogenous population [8,9]. In contrast, a person-oriented approach identifies subgroups of people who exhibit similar patterns of characteristics. LCA is an example of a person-oriented approach [9]. A person-oriented approach is particularly useful to uncover subgroups in heterogeneous populations [10]. The use of variable-oriented versus person-oriented approaches allow for different statements to be made, thus, the selection of which approach to use is determined by the research questions and the theoretical foundation [8,11]. The rest of this article will focus on the use of a person-oriented approach, latent class analysis.

LCA is a categorical statistical technique that is used to identify subgroups or classes of individuals based on response patterns in a set of categorical data [9,12,13]. These subgroups or classes are latent variables. Latent variables are unmeasured and unobserved variables, which are interfered from observed variables (indicator variables) using statistical methods [14]. The main assumption of latent variables is that of local independence, where it is assumed that observed variables are only connected through the latent variable [14]. This assumption refers only to conditioning on the latent variable and does not imply that the indicator variables in the dataset are independent. It expected that the indicators would be correlated in the overall sample [9]. In symptom cluster research the latent variables are symptom clusters, which are inferred from observed symptoms that are measured using validated symptom measures.

LCA discerns meaningful latent classes against background noise and provides a way to arrange complex data in a parsimonious manner [9]. For example, when using LCA for symptom clusters, LCA identifies groups of people who have a similar symptom experience. LCA has been successfully used to determine symptom cluster membership in various patient populations including myocardial infarction [15], cancer [16,17], and the menopause transition [18]

Selecting Indicator Variables

Symptoms that could be used in LCA are diverse but may include sleep disturbance pain, fatigue, depression, anxiety, abdominal bloating, itching, etc. The selection of indicators in an LCA model should be based on the empiric literature and guided by theory. (See Table 1 for examples of indicator variables used in symptom cluster research.) When selecting indicator variables, it has been suggested that using no less than five indicator variables may

help with model convergence [19]. However, some researchers also have suggested that using a limited number of indicator variables may assist with interpretability, help with classification, and increase accuracy of parameter estimates [20].

In LCA, there are no assumptions regarding the normal distribution of the indicator variables [21]. Data suitable for LCA includes binary, categorical, Likert-scale, or nominal data [22]. LCA cannot be used with purely ordinal data [9,23]. If indicators are continuous, they need to be reduced into meaningful categorical data for LCA or latent profile analysis, a variation of LCA that allows for the use of continuous indicator variables, can be used [24]. Due to the need for categorical variables, using symptom measures with established cut-off scores is helpful so that the presence/absence of symptom is a cluster is clear and clinically meaningful.

Model Selection

The number of latent classes in the final model is determined by a combination of statistical criteria, parsimony, and interpretability [25]. To determine the final model, the researcher runs models with different numbers of classes and the best fit is determined by relative fit statistics, where fit is compared between models available [26]. Common statistics used to determine what number of latent classes better represent the data are the Akaike Information Criterion (AIC) [27], and the Bayesian Information Criteria (BIC) [12]. For both of these statistics, both lower numbers indicate better fit.

Parameter Estimates

In LCA, two sets of parameters are estimated: (1) prevalence of each latent class and (2) conditional response probabilities or the probability that each indicator is present/absent for a member of the latent class. In LCA, subjects are classified into groups that are mutually exclusive and exhaustive, meaning that each individual is assigned to one group, but only one group, thus, latent class prevalences sum to 100 [9]. Individuals are classified into latent classes based on probability. Item-response probabilities range from 0 to 1 with 1 meaning that conditional membership in a latent class is certain and 0 meaning that there is independence between the indicator variable and the class [23]. The interpretation and labeling of latent classes are done by the researcher and are based on the item-response probabilities [24].

Sample Size Requirements

LCA required a large sample size. In general, larger sample sizes provide better model estimation; thus, researchers have suggested that a minimum sample size of 100 to 300 is optimal to ensure an optimal model [9,19]. For additional guidance on power and sample size, Dziak, Lanza, and Tan (2014) published power tables to guide sample size selection in LCA.

Statistical Software

Currently, LCA is not included in standard statistical packages. Free downloadable add-ons for SAS, Stata, and R are available from the Methodology Center at Pennsylvania State University (http://methodology.psu.edu/downloads). Other examples of packages that can be used to perform LCA include Latent GOLD [28], MPlus [29], and PANMARK3 [30]. Each

software package uses slightly different language and fit statistics, so it is important to become familiar the package that you are planning to use.

Extensions of LCA

Many research questions regarding symptom clusters extend beyond just determining symptom clusters at one time point. Many extensions of LCA allow LCA to be useful in answering diverse research questions about how symptom cluster groups differ and how symptom cluster membership changes over time. Multiple-group LCA allows the researcher to explore if there are group differences between the latent group prevalences and item-response probabilities where the groups are observed groups (e.g., gender, anemia) [9]. LCA with covariates uses a logistic link function to identify characteristics, such as age, gender or biomarkers, to predict symptom cluster membership [9,23]. Latent transition analysis is a longitudinal extension of LCA that addresses the factor of time and allows for the modeling of changes in symptom cluster membership over two or more time periods [31]. Using the extensions of LCA greatly increases one's ability to answer pertinent nursing questions but can make the interpretation of the results complex. As such, support from a statistician familiar with latent class modeling for more complex analysis is needed. See Table 1 for examples of latent class analysis with biomarkers and genetics.

Examples of Latent Class Analysis in the Symptom Cluster Literature

In addition to identifying empirically determined symptom clusters, the linking of symptom clusters to biomarkers, genetics, and epigenetics is essential to understanding the underlying etiology of symptom clusters [32]. The use of biomarkers in symptom cluster research has the potential to assist in identifying who is at risk for experiencing a high symptom burden and treatment responders and nonresponses [33]. In LCA, biomarkers and genetic information could be used as covariates, indicator variables, or as a grouping variable to assist with uncovering the etiology of symptom clusters.

Biomarkers and Genetics as Indicator Variables

Researchers can use LCA to cluster biomarkers and genetic variables, along with the symptoms, as indicator variables. There is limited research that includes biological variables as an indicator variable. While not specific to symptom clusters, one study in asthma used symptoms, demographic, clinical and biomarker variables as indicator variables and found interpretable and clinically relevant latent clusters [34]. Other authors have suggested that including signs, symptoms and biomarkers as indicator variables in LCA may provide improved case detection [35]. However, a lack of clear clinical cut-off points may limit the use of adding biomarkers as indictor variables in LCA, as LCA requires the used use of categorical indicator variables. Latent profile analysis, a variation of LCA that allows for the use of continuous indicator variables [6], may be more appropriate to use when adding biomarkers as an indicator variable.

Biomarkers and Genetics as Multiple-Group Variable

Biomarker and genetic variables can be included as a grouping variable or covariate in multiple-group LCA. Biomarkers with established cut-off scores lend themselves to this type

of analysis because distinct existing groups are needed to conduct multiple-group LCA. However, many novel biomarkers do not have clear cut-off points of what is considered "normal/abnormal", which limits the use of multiple-group LCA. An example of a biomarker that could be used in multiple-group LCA is hemoglobin/hematocrit with groups of anemic and non-anemic to compare differences in latent class symptom cluster membership across the two groups. Another limitation to using a gene or a biomarker as a grouping variable in multiple-group LCA is that the researcher can only use one polymorphism or biomarker to compare classes by groups.

Biomarkers and Genetics as Covariates

In LCA with covariates, a logistic or multinomial regression is performed to explore associations between biomarkers or genetic variables with symptom cluster membership. Both categorical and continuous covariates can be used in LCA with covariates. LCA with covariates has been used to identify associations between biomarker or genetic variables and symptom cluster membership in diverse populations including cancer [16,17], the menopause transition [36], and depressive symptoms (Woods et al., 2014).

Two studies explored how pro-inflammatory genetic polymorphisms are associated with symptom cluster membership in cancer [16,17]. The researchers found that there were genetic polymorphisms in the following pro-inflammatory cytokine genes were associated with membership in the all-high symptom class compared with the all-low symptom class, interleukin 6 (IL6), IL13, tumor necrosis factor-alpha [17], and IL4 [16].

Another study explored hypothalamic-pituitary-ovarian axis (HPO), hypothalamic-pituitary-adrenal axis (HPA), and autonomic nervous system biomarkers in menopause transition and early menopause [18]. The researchers found associations between low estrogen, high follicle-stimulating hormone, low epinephrine, and low norepinephrine with membership in the low symptom severity group compared with the high symptom severity group.

A fourth study explored metabolic, inflammatory, and HPA axis biomarkers association with depressive symptom cluster membership [36]. The researchers suggest that atypical and melancholic depression have distinct symptom and biomarkers as atypical depression is associated with increased metabolic and inflammatory biomarkers and melancholic depression is associated with HPA-axis biomarkers.

As demonstrated in the above studies, a benefit of using LCA with covariates with genetic and biomarker data is that it allows for the use of multiple genetic polymorphisms and biomarkers to determine the association with symptom cluster membership. The identification of multiple genetic polymorphisms and biomarkers may allow for the discovery of a common etiology in symptom clusters, as common pathways may be implicated in symptom cluster membership. However, to date research is limited in connecting symptom clusters to genetics and biomarkers and the further research is needed to better understand the etiology of symptom clusters.

Discussion

LCA allows researchers to connect biomarkers and genetics to empirically validated symptom clusters. The results of these few studies that have used LCA with genetics and biomarkers are encouraging the idea that symptom clusters have a common etiology. However, the full potential of LCA in symptom cluster research has not been realized, as it has been used in limited populations and explored limited pathways. Other common mechanisms that researchers should consider exploring with symptom clusters determined by LCA include inflammatory cytokines [37], genetic polymorphisms and genetic expression of inflammatory pathways [38], and the microbiota-gut-brain axis [39,40]. Research linking biologic pathways and symptom clusters may eventually provide knowledge about personalized symptom risk profiles needed to create tailored symptom management interventions and improve patient outcomes [38].

There are several strengths of using LCA in symptom cluster research. Using LCA with biomarkers and genetics data is that it allows researchers to connect biomarkers and genetics to empirically determined symptoms clusters. Also, LCA is data-driven, and; thus, no hypotheses about the interaction of the indicators are needed to perform latent class analysis [9]. Because no hypothesis is needed, LCA is particularly useful in situations where little is known about the phenomena, which is frequently the case in symptom cluster research. In addition, LCA provides robust model fit statistics, which provides confidence in model selection and allows for the stability of the findings in subgroups to be tested [6,9].

Another strength of LCA is that it does not have any assumptions of normality. The lack of normality assumptions in makes LCA well-suited for use with a variety of data where assumptions of multivariate normality and that the underlying latent variable is continuous may not be able to be met, including administrative databases and previously collected data [41]. However, LCA, as with any statistical method, cannot overcome measurement or collection errors in symptoms, genetics of biomarker data, which may be an issue in pre-existing data.

The major limitation of LCA is the need for a large sample size. This requirement can be difficult to meet in research that uses biomarkers and genetic information due to the cost of data collection processing for researchers [42]. Another limitation in LCA is that complex latent class models require statistical expertise to run and interpret. Researchers without expertise in this method should ensure they have statistical support before starting a study. Additional studies are needed that explore symptom clusters and biomarker and genetic variables to determine the full strengths and limitations of LCA in this research.

The use of LCA in conjunction with biomarkers, genetics, and epigenetic has the potential to expand our knowledge about the underlying etiology of symptom clusters and to allow for early detection, diagnosis, and treatment of symptoms. Potentially, people at the greatest risk for high symptom burdens can be identified and treated early, thus preventing long-term sequela of untreated symptoms [33].

Acknowledgments

References

1. Aktas A, Walsh D, Rybicki L. Symptom clusters: Myth or reality? Palliat Med. 2010; 24(4):373–385. Accessed 1/31/2014 3:54:50 PM. 10.1177/0269216310367842. DOI: 10.1177/0269216310367842 [PubMed: 20507866]

2. Fan G, Filipczak L, Chow E. Symptom clusters in cancer patients: A review of the literature. Curr Oncol. 2007; 14(5):173–179. Accessed 11/16/2013 11:54:31 AM. [PubMed: 17938700]

3. Kim HJ, Abraham I, Malone PS. Analytical methods and issues for symptom cluster research in oncology. Curr Opin Support Palliat Care. 2013; 7(1):45–53. Accessed 4/24/2014 9:49:45 AM; 4/24/2014 9:49:45 AM. [doi]. DOI: 10.1097/SPC.0b013e32835bf28b [PubMed: 23196378]

4. Maliski SL, Kwan L, Elashoff D, Litwin MS. Symptom clusters related to treatment for prostate cancer. Oncol Nurs Forum. 2008; 35(5):786–793. Accessed 1/31/2014 3:50:08 PM. 10.1188/08.ONF.786-793. DOI: 10.1188/08.ONF.786-793 [PubMed: 18765324]

5. Lanza ST, Rhoades BL. Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. Prev Sci. 2013; 14(2):157–168. Accessed 4/21/2015 2:38:57 PM. [doi]. DOI: 10.1007/s11121-011-0201-1 [PubMed: 21318625]

6. Berlin KS, Williams NA, Parra GR. An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. J Pediatr Psychol. 2013; Accessed 1/4/2014 12:32:51 PM. doi: 10.1093/jpepsy/jst084

7. Hagenaars, JA., McCutcheon, AL. Preface. In: Hagenaars, JA., McCutcheon, AL., editors. Applied latent class analysis. New York, NY: Cambridge University Press; 2002. p. xi-xxii.

8. Bergman LR, Trost K. The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? Merrill-Palmer Quarterly. 2006; 52(3):601–631.

9. Collins, LM., Lanza, ST. Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. Hoboken, NJ: Wiley; 2010.

10. Magnusson D. The person approach: Concepts, measurement models, and research strategy. New Dir Child Adolesc Dev. 2003; (101):3–23. (101). Accessed 4/20/2015 1:50:12 PM. [doi]. DOI: 10.1002/cd.79 [PubMed: 15460974]

11. von Eye A, Bogat GA, Rhodes JE. Variable-oriented and person-oriented perspectives of analysis: The example of alcohol consumption in adolescence. J Adolesc. 2006; 29(6):981–1004. Accessed 12/10/2015 10:00:05 AM. doi:S0140-1971(06)00085-6. [pii]. [PubMed: 17045640]

12. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika. 1974; 61:215–231.

13. Lazarsfeld, PF., Henry, NW. Latent structure analysis. Boston, MA: Houghton-Mifflin; 1968.

14. Bollen KA. Latent variables in psychology and the social sciences. Annu Rev Psychol. 2002; 53:605–634. Accessed 9/6/2016 10:25:05 AM; 9/6/2016 10:25:05 AM. [doi]. DOI: 10.1146/annurev.psych.53.100901.135239 [PubMed: 11752498]

15. Ryan CJ, DeVon HA, Horne R, et al. Symptom clusters in acute myocardial infarction: A secondary data analysis. Nurs Res. 2007; 56(2):72–81. Accessed 1/4/2014 12:26:39 PM. DOI: 10.1097/01.NNR.0000263968.01254.d6 [PubMed: 17356437]

16. 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(3):437–447. Accessed 4/15/2015 10:06:35 AM. [doi]. DOI: 10.1016/j.cyto.2012.02.015 [PubMed: 22450224]

17. 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. 2014 Accessed 4/15/2015 10:03:37 AM. doi:1099800414550394. [pii].

18. Woods NF, Cray L, Mitchell ES, Herting JR. Endocrine biomarkers and symptom clusters during the menopausal transition and early postmenopause: Observations from the seattle midlife

women's health study. Menopause. 2014; 21(6):646–652. Accessed 4/15/2015 10:09:06 AM. [doi]. DOI: 10.1097/GME.0000000000000122 [PubMed: 24781854]

19. Wurpts IC, Geiser C. Is adding more indicators to a latent class analysis beneficial or detrimental? results of a monte-carlo study. Front Psychol. 2014; 5:920. Accessed 4/15/2015 9:09:00 AM. [doi]. doi: 10.3389/fpsyg.2014.00920 [PubMed: 25191298]

20. Dean N, Raftery AE. Latent class analysis variable selection. Ann Inst Stat Math. 2010; 62(1):11–35. Accessed 4/24/2014 12:45:38 PM. [doi]. DOI: 10.1007/s10463-009-0258-9 [PubMed: 20827439]

21. McCutcheon, AL. Sage university paper series on quantitative applications in the social sciences no 07-064. Newbury Park, CA: Sage; 1987. Latent class analysis.

22. Vermunt, JK., Magidson, J. Latent class analysis. In: Lewis-Beck, MS.Bryman, A., Liao, TF., editors. The sage encyclopedia of social sciences research methods. Thousand Oakes, CA: Sage Publications; 2004. p. 549-553.

23. Lanza ST, Collins LM, Lemmon DR, Schafer JL. PROC LCA: A SAS procedure for latent class analysis. Struct Equ Modeling. 2007; 14(4):671–694. Accessed 1/31/2014 3:44:53 PM; 1/31/2014 3:44:53 PM. [PubMed: 19953201]

24. Vermunt, JK., Magidson, J. Applied latent class analysis. Cambridge: Cambridge University Press; 2002. Latent class cluster analysis; p. 89-106.

25. Lanza ST, Collins LM. A new SAS procedure for latent transition analysis: Transitions in dating and sexual risk behavior. Dev Psychol. 2008; 44(2):446–456. Accessed 4/24/2014 9:52:55 AM. [doi]. DOI: 10.1037/0012-1649.44.2.446 [PubMed: 18331135]

26. Karen L, Nylund K, Asparouhov T, Muthén B. Deciding on the number of classes in latent class analysis: A monte carlo simulation study. Structural equation modeling : a multidisciplinary journal. 2007; 14:535–569.

27. Akaike H. Factor analysis and AIC. Psychometrika. 1987; 52(2):317–332.

28. Vermunt, JK., Magidson, J. Latent GOLD user's guide. Boston, MA: Statistical Innovations, Inc; 2000.

29. Muthé, LK., Muthén, BO. Mplus user's guide. Los Angeles, CA: Muthén & Muthén; p. 1998-2010.

30. van de Pol, F., Langeheine, R., de Jong, W. PANMARK 3 user's manual. Voorburg, The Netherland: Netherlands Central Bureau of Statistic; 1998.

31. Roberts TJ, Ward SE. Using latent transition analysis in nursing research to explore change over time. Nurs Res. 2011; 60(1):73–79. Accessed 4/15/2015 9:15:10 AM. [doi]. DOI: 10.1097/NNR. 0b013e3182001c63 [PubMed: 21127448]

32. 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. 2003; 97(11):2919–2925. Accessed 1/31/2014 1:56:33 PM. DOI: 10.1002/cncr.11382 [PubMed: 12767108]

33. Miaskowski C, Aouizerat BE. Biomarkers: Symptoms, survivorship, and quality of life. Semin Oncol Nurs. 2012; 28(2):129–138. Accessed 4/9/2015 2:39:07 PM. [doi]. DOI: 10.1016/j.soncn. 2012.03.008 [PubMed: 22542321]

34. Siroux V, Basagana X, Boudier A, et al. Identifying adult asthma phenotypes using a clustering approach. Eur Respir J. 2011; 38(2):310–317. Accessed 12/7/2015 6:22:32 PM. [doi]. DOI: 10.1183/09031936.00120810 [PubMed: 21233270]

35. Sood R, Ford AC. Use of biomarkers in irritable bowel syndrome: To predict the future, look at the past. Clin Transl Gastroenterol. 2015; 6:e116. Accessed 12/9/2015 10:34:56 AM; 12/9/2015 10:34:56 AM. [doi]. doi: 10.1038/ctg.2015.41 [PubMed: 26448457]

36. Lamers F, Vogelzangs N, Merikangas KR, de Jonge P, Beekman AT, Penninx BW. Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. Mol Psychiatry. 2013; 18(6):692–699. Accessed 12/7/2015 6:28:34 PM. [doi]. DOI: 10.1038/mp.2012.144 [PubMed: 23089630]

37. Cleeland CS. Symptom burden: Multiple symptoms and their impact as patient-reported outcomes. J Natl Cancer Inst Monogr. 2007; (37):16–21. (37). Accessed 9/3/2014 5:27:15 PM. doi: 2007/37/16. [pii].

38. Starkweather AR, Lyon DE, Elswick RK Jr, Montpetit AJ, Conley Y, McCain NL. A conceptual model of psychoneurological symptom cluster variation in women with breast cancer: Bringing nursing research to personalized medicine. Curr Pharmacogenomics Person Med. 2013; 11(3): 224–230. Accessed 4/9/2016 12:20:55 PM. [PubMed: 24497894]

39. Mayer EA, Tillisch K, Gupta A. Gut/brain axis and the microbiota. J Clin Invest. 2015; 125(3): 926–938. Accessed 3/28/2016 12:28:49 PM. [doi]. DOI: 10.1172/JCI76304 [PubMed: 25689247]

40. Montiel-Castro AJ, Gonzalez-Cervantes RM, Bravo-Ruiseco G, Pacheco-Lopez G. The microbiota-gut-brain axis: Neurobehavioral correlates, health and sociality. Front Integr Neurosci. 2013; 7:70. Accessed 3/29/2016 12:14:16 PM. [doi]. doi: 10.3389/fnint.2013.00070 [PubMed: 24109440]

41. Magidson, J., Vermunt, JK. Comparing latent class factor analysis with the traditional approach in data mining. In: Bozdogan, H., editor. Statistical data mining & knowledge discovery. New York, NY: Chapman & Hall/CRC; 2003. p. 373-383.

42. Mayeux R. Biomarkers: Potential uses and limitations. NeuroRx. 2004; 1(2):182–188. Accessed 12/8/2015 12:45:51 PM. [doi]. DOI: 10.1602/neurorx.1.2.182 [PubMed: 15717018]

43. Lamers F, Rhebergen D, Merikangas KR, de Jonge P, Beekman AT, Penninx BW. Stability and transitions of depressive subtypes over a 2-year follow-up. Psychol Med. 2012; 42(10):2083–2093. Accessed 9/19/2016 12:04:33 PM; 9/19/2016 12:04:33 PM. [doi]. DOI: 10.1017/S0033291712000141 [PubMed: 22340131]

Table 1

Examples of Using Latent Class Analysis to Determine Symptom Clusters

First author (year) Sample size	Study aims	Indicator variables	Method	Latent Class Results	Associations with biomarkers or genetics
Doong (2014) N=398	"To determine whether distinct latent classes of patients with breast cancer could be identified, and whether patients in these latent classes differed on demographic and clinical characteristics, and whether variations in cytokine genes were associated with latent class membership."	Pain, sleep disturbance, and depression	Latent class profile analysis using Mplus [18]	Three classes were identified: all-low (61.0%), low pain and high fatigue (31.6%), and all-high (7.1%).	Significant associations were found between interleukin 6 (IL6), rs2069845, IL13, and tumor necrosis factor alpha rs18800610 and latent class membership in the all class compared to the all-low class.
Illi (2012) N=253	"[T]o determine: if distinct classes of individuals could be identified based on their experience with pain, fatigue, sleep disturbance, and depression; if these classes differed on demographic and clinical characteristics; and if variations in pro- and anti-inflammatory cytokine genes were associated with latent class membership."	Pain, fatigue, sleep disturbance, and depression	Latent class profile analysis using Mplus	Three classes were identified: low depression and low pain (83%), high depression and low pain (4.7%), all high symptoms (12.3%).	The minor allele of IL4 rs2243248 was associated with membership in the "all-high" class.
Lamers (2013) N=776	"The aim of the current study was to compare different pathophysiological indicators of HPA-axis function, the inflammatory response system and metabolic syndrome across these two subtypes of MDD and healthy controls."	16 depressive symptoms including depressed mood, loss of interest, weight, appetite, sleep, psychomotor, fatigue/energy loss, guilt/worthlessness, lack of concentration/ indecisiveness, suicidal, lack of responsiveness, leaden paralysis, interpersonal sensitivity, quality of mood, worse in morning, early morning awakening	Latent transition analysis using MPlus with multivariable analysis using SPSS	Previous research found * three depressive symptom clusters: severe melancholic/typical (27.2%) characterized by loss of appetite and weight loss; severe atypical subgroup (32.2%) characterized by weight gain, and overeating; and a moderate symptom subgroup (40.5%) characterized by overall lower symptom severity.	People with atypical depression compared to melancholic depression and controls had higher metabolic markers (increased waist circumference and BMI) and inflammatory markers (increased CRP, IL-6, TNF-a), and decreased cortisol diurnal slope. People with melancholic depression had increased triglycerides and cortisol awakening repose area under the curve with respect to the ground. And the control group and increase systolic blood pressure.
Ryan (2007) N=1,073	"The objective of this study was to identify clusters of symptoms that represent AMI and to relate the clusters to specific demographic groups."	Chest pain or discomfort, shoulder, arm or hand pain or discomfort, sweating, general weakness, shortness of breath, fatigue, nausea or vomiting, dizziness or lightheadedness, indigestion, neck or jaw pain or discomfort, abdomen pain or discomfort	Latent class analysis with covariates using Latent Gold	5 clusters were found Cluster 1 (43%) high probability of chest discomfort, shoulder, arm or hand pain or discomfort, weakness; Cluster 2 (65.2%) high probability of chest discomfort, shoulder, arm or hand pain or discomfort; Cluster 3 (17%) high probability of chest discomfort, shoulder, arm or hand pain or discomfort, nausea or vomiting, shortness of breath, sweating, dizziness or lightheadedness, weakness, fatigue;	N/A

First author (year) Sample size	Study aims	Indicator variables	Method	Latent Class Results	Associations with biomarkers or genetics
				Cluster 4 (8%) high probability of shoulder, arm or hand pain or discomfort, abdominal pain, indigestion Cluster 5 (6%) high probability of no high probability symptoms	
Woods (2014) N=292	"[T]ested models of the differential effects of hypothalamic-pituitary-ovarian axis, hypothalamic-pituitary-adrenal axis, and autonomic nervous system biomarkers on the three symptom severity classes."	Sleep, pain, mood, cognitive, tension & hot flashes	Multilevel latent class analysis	Three symptom classes were identified: high-severity hot flash group (13%) characterized by severe hot flashes with moderate sleep, mood, cognitive, and pain symptoms; moderate severity (17%) characterized by moderate levels of all symptoms but low hot flashes; and low severity group (70%) characterized by low levels for all symptoms.	Lower estrogen and higher FSH levels and having lower epinephrine and high norepinephrine levels increased the likelihood of belonging to the high-severity hot flash class versus the low-severity hot flash class. Having lower epinephrine levels was associated with belonging to the moderate-severity class versus the low-severity class. Cortisol and testosterone were unrelated to symptom severity class membership.

Note:

*43