

Identifying Subgroups of Complex Patients With Cluster Analysis

Sophia R. Newcomer, MPH; John F. Steiner, MD, MPH;
and Elizabeth A. Bayliss, MD, MSPH

Objective: To illustrate the use of cluster analysis for identifying sub-populations of complex patients who may benefit from targeted care management strategies.

Study Design: Retrospective cohort analysis.

Methods: We identified a cohort of adult members of an integrated health maintenance organization who had 2 or more of 17 common chronic medical conditions and were categorized in the top 20% of total cost of care for 2 consecutive years (n = 15,480). We used agglomerative hierarchical clustering methods to identify clinically relevant subgroups based on groupings of coexisting conditions. Ward's minimum variance algorithm provided the most parsimonious solution.

Results: Ward's algorithm identified 10 clinically relevant clusters grouped around single or multiple "anchoring conditions." The clusters revealed distinct groups of patients including: coexisting chronic pain and mental illness, obesity and mental illness, frail elderly, cancer, specific surgical procedures, cardiac disease, chronic lung disease, gastrointestinal bleeding, diabetes, and renal disease. These conditions co-occurred with multiple other chronic conditions. Mental health diagnoses were prevalent (range 28% to 100%) in all clusters.

Conclusions: Data mining procedures such as cluster analysis can be used to identify discrete groups of patients with specific combinations of comorbid conditions. These clusters suggest the need for a range of care management strategies. Although several of our clusters lend themselves to existing care and disease management protocols, care management for other subgroups is less well-defined. Cluster analysis methods can be leveraged to develop targeted care management interventions designed to improve health outcomes.

(Am J Manag Care. 2011;17(8):e324-e332)

For author information and disclosures,
see end of text.

By 2020, over 81 million persons in the United States will have 2 or more chronic conditions.¹ Multimorbidity results in adverse health outcomes and higher healthcare costs, and challenges current models of care delivery.^{2,3} Care management has the potential to improve health outcomes for persons with multimorbidities. However, most disease and care management strategies have been developed to improve specific health outcomes for populations defined by single diseases or specific circumstances (such as hospital discharge).⁴⁻¹¹ There is a need for strategies that can identify sub-populations with multiple, interacting diseases, in order to provide them with appropriate and relevant care management support.

Investigations to identify these populations of complex patients have traditionally relied upon multivariable regression analyses to identify patient-level characteristics (such as demographics and diseases) that predict the outcome of interest (such as hospitalization).¹²⁻¹⁴ As compared with investigations that use multivariable regression analyses to identify individual disease predictors of specific outcomes, data mining techniques provide an opportunity to empirically identify groups of patients with similar patterns of multimorbidities. One such technique, cluster analysis, refers to classification methods that are used for discovering groups or "clusters" of "highly similar entities" within data sets.¹⁵ Cluster analyses are common in psychology, sociology, and marketing research, and the methods have been used to a limited extent in health services research.¹⁶⁻¹⁸ While cluster analyses previously have been used to discover patterns of multimorbidities,¹⁹⁻²¹ in this study we demonstrate the application of such methods for identifying clusters of patients with high utilization that may suggest opportunities for enhanced care management in a managed care setting.

We used cluster analysis to explore a large, 2-year cohort of health maintenance organization members with 2 or more chronic conditions. We hypothesized that within a large, complex patient population, cluster analysis would reveal groups of patients with distinct patterns of comorbid conditions. Although some of these subgroups would be characterized by well-known patterns of co-occurring medical conditions with established care management strategies, other subgroups would reveal combi-

In this article
Take-Away Points / e325
Published as a Web exclusive
www.ajmc.com

nations of comorbidities that might benefit from new, proactive, and targeted care management.

METHODS

Setting

Kaiser Permanente Colorado (KPCO) is an integrated, not-for-profit health maintenance organization. During the years studied (2006 and 2007), KPCO had approximately 430,000 members. This study was approved by KPCO's Institutional Review Board.

Study Population

The study population consisted of KPCO members 21 years or older on January 1, 2006, categorized in the top 20% of total cost of care in both 2006 and 2007, each with 2 or more of 17 common chronic medical conditions. Annual cost estimates combined general ledger costs with direct and indirect utilization-related costs to provide cost-of-care estimates for KPCO members.²² We excluded members with a long-term care facility stay, chronic kidney dialysis, or an inpatient visit of greater than 30 days during the 2 years based on the premise that their unique and significant care management needs are likely to already be well defined. Six extremely high cost outliers were also removed from the cohort.

We compiled a list of 17 chronic medical conditions based on prevalence in the general population, prevalence in our specific cohort, and a literature search of conditions likely to predict hospitalization or adverse health outcomes in complex patients.^{13,23-34} The selected conditions were diabetes, chronic obstructive pulmonary disease (COPD), chronic kidney disease, stroke, obesity, dementia, fall, hip fracture, chronic pain, skin ulcer, orthopedic surgery, back surgery, abdominal surgery, gastrointestinal bleeding, cancer (excluding non-melanoma skin cancer), cardiac disease (which included coronary artery disease and congestive heart failure), and mental health conditions—primarily depression, but also including generalized anxiety and bipolar disorders. Determinations of whether cohort members had a given condition were based on inpatient and outpatient *International Classification of Diseases, Ninth Revision (ICD-9)* diagnosis and procedure codes in 2006 and 2007. In addition, we used KPCO's cancer registry to determine cancer diagnoses in 2005, 2006, and 2007. We considered a cohort member to have obesity if they had an ICD-9 diagnosis code for obesity or a median body mass index (BMI) greater than or equal to 30 in 2006 and 2007. BMI data were available for 98.6% of cohort members; if a cohort member

Take-Away Points

This study illustrates the use of cluster analysis to identify sub-populations of complex patients for potential targeted care management within an integrated health maintenance organization.

- Among a cohort of adults with multimorbidity and high healthcare utilization, we identified 10 clinically relevant clusters of complex patients.
- While care management protocols may already exist in many healthcare settings for some common clusters, other clusters identified present opportunities for new or enhanced care management.
- Data mining methods such as cluster analysis can be applied in other settings where electronic diagnosis data are readily available.

did not have a BMI value or an obesity diagnosis in 2006 or 2007 then we did not consider them to be obese.

Statistical Analysis

SAS version 9.2 (SAS Institute, Cary, North Carolina) was used for all analyses. We described demographic attributes, healthcare utilization, comorbidity score (using the Quan adaptation of the Elixhauser comorbidity index),³⁵ and prevalence of clinical conditions within the cohort using frequencies and medians with 25th and 75th percentiles.

Agglomerative Hierarchical Clustering. We used agglomerative hierarchical clustering to identify clinically relevant groups of cohort members with similar multimorbid conditions. With this method of cluster analysis, each cohort member starts as its own cluster. The 2 most similar clusters are merged and this new cluster replaces the 2 former clusters. The process continues until there is only 1 cluster containing all observations.^{15,36,37} After the clustering algorithm is run, the user must select the appropriate cutoff point for the number of clusters desired based on clinical importance or other pre-specified criteria.

Algorithms. Various algorithms are available for cluster analysis. For this study, we used Ward's minimum variance method as the primary algorithm. With this algorithm, every possible cluster combination is considered at each step of agglomerative hierarchical clustering, and the combination that results in the smallest addition to the error sum of squares is selected.^{15,37} Ward's method is a widely used algorithm which minimizes the variance within clusters, and is also known to produce clusters of similar sizes.^{15,17,18,38-43} We compared results from Ward's method to results using the flexible beta algorithm, where the user sets different levels of beta, and beta values less than zero optimize the dissimilarity between clusters.^{19,20,44}

Analytic Process. In the analytic data set, the presence or absence of each of the 17 conditions was represented with a 1 or 0 for each cohort member. We first randomly split the full analytic data set into 2 equally sized data sets. We then converted each split data set into a dissimilarity matrix using

■ **Table 1.** Description of Study Cohort (n = 15,480)

Characteristic	No. (%)
Female	9145 (59.1%)
	Median (25th percentile, 75th percentile)
Age as of January 1, 2006	65 (54, 74)
Quan score (number of comorbidities), 2006-2007 ³⁵	5 (3, 7)
Inpatient hospitalizations in 2006-2007	1 (0, 2)
Emergency department visits in 2006-2007	1 (0, 2)
In-network primary care visits in 2006-2007	8 (5, 12)
In-network specialty care visits in 2006-2007	11 (7, 18)
Prevalence of clinical conditions used in cluster analysis	No. (%)
Obesity	9084 (58.7%)
Mental health conditions ^a	7351 (47.5%)
Diabetes	6204 (40.1%)
Cardiac disease ^b	4401 (28.4%)
Chronic obstructive pulmonary disease	3634 (23.5%)
Kidney disease (not requiring dialysis)	3565 (23.0%)
Cancer	1543 (10.0%)
Gastrointestinal bleeding	1272 (8.2%)
Chronic pain	1217 (7.9%)
Stroke	1174 (7.6%)
Skin ulcer	834 (5.4%)
Dementia	809 (5.2%)
Fall	688 (4.4%)
Abdominal surgery	677 (4.4%)
Orthopedic surgery	497 (3.2%)
Back surgery	158 (1.0%)
Hip fracture	145 (0.9%)

^aPrimarily depression, but also includes generalized anxiety and bipolar disorder.
^bIncludes coronary artery disease and congestive heart failure.

Jaccard's coefficient. This is an appropriate distance measure for clinical conditions, as it considers the number of conditions that 2 people have in common and ignores conditions that neither person has.¹⁹

We ran our primary algorithm, Ward's minimum variance method, on both split data sets. The pseudo F, pseudo T, and r^2 statistics were examined for different numbers of clusters to identify possible clustering solutions.³⁷ These statistics pointed to several desirable numbers of clusters, and membership in these clusters was described by examining the prevalence of each condition in the cluster. Cluster membership was com-

pared between the 2 split data sets to assess the consistency of the clustering process. Since cluster membership was similar between the 2 split data sets, thus reinforcing the stability of the algorithm in this population, Ward's algorithm was run on the entire data set. We subjectively determined that a 10-cluster solution produced the most clinically relevant clusters. For comparison, we then produced 10 cluster solutions using the flexible beta method, with beta set at -0.25 and -0.5, and compared these results with Ward's method. This confirmed that Ward's algorithm resulting in 10 clusters appeared to be the most parsimonious solution and provided the most clinically relevant groups. We then described the 10 clusters by the number of cohort members in the cluster, median age of cluster members, and percentage of cluster members with the most prevalent conditions in that cluster. We also described relative cost of care ratios for each cluster.

RESULTS

Table 1 provides descriptive demographic and disease characteristics of the study cohort (n = 15,480). The median age of cohort members was 65 years, and 59.1% of cohort members were women. Cohort members had a median of 5 chronic medical conditions (including, but not limited to, the 17 conditions included in the cluster analysis).

In this study we tried 3 clustering algorithms—Ward's, flexible beta with beta set to -0.25, and flexible beta with beta set to -0.5. Full results from the 3 methods are in the [Appendix](#). All 3 methods identified some of the same distinct clusters, including clusters of patients with chronic pain and mental health conditions and gastrointestinal bleeding and mental health conditions. However, we preferred Ward's method because the flexible beta methods tended to produce many small homogeneous clusters and several large heterogeneous clusters, while Ward's method produced more similarly sized and more homogeneous clusters. Eight of the 10 clusters produced by Ward's method had what we referred to as “anchoring” conditions, or chronic conditions shared by almost all (>98%) of cluster members. We considered such anchoring conditions potentially useful given the ultimate goal of proactively identifying characteristics of persons who may benefit from targeted care management. With the flexible beta results, only 5 and 7 of the clusters from the method with beta set at -0.25 and -0.5, respectively, had anchoring conditions.

Cluster Analysis for Targeted Care Management

Table 2. Description of Subgroups of Complex Patients Identified Through Cluster Analysis Using Ward's Minimum Variance Method

Cluster Number	Anchoring Conditions	Number of Members	% of Total Cohort	Median Age (25th Percentile, 75th Percentile)	Most Prevalent Medical Conditions in Cluster ^a
1	Chronic pain with mental health conditions	1017	6.6%	55 (46, 65)	99.8% Chronic pain 69.2% Mental health conditions 47.2% Obesity
2	Diabetes with obesity and mental health conditions	1855	12.0%	59 (51, 67)	100% Diabetes 86.0% Obesity 44.0% Mental health conditions 0% Have any of the other 14 conditions
3	Kidney disease with diabetes and obesity	2164	14.0%	72 (64, 78)	99.9% Kidney disease 51.2% Diabetes 50.9% Obesity
4	Mental health conditions and obesity in younger adults	1676	10.8%	50 (41, 59)	100% Mental health conditions 100% Obesity 0% Have any of the other 15 conditions
5	Frailty in the elderly	2788	18.0%	73 (63, 80)	45.4% Mental health conditions 39.7% Diabetes 37.7% Obesity 35.2% Stroke 30.7% Cardiac disease 26.7% Kidney disease 26.4% Skin ulcers 25.8% Dementia
6	Cardiac disease and obesity	1776	11.5%	68 (58, 75)	100% Cardiac disease 54.2% Obesity 39.4% Diabetes
7	Chronic obstructive pulmonary disease (COPD) with obesity and mental health conditions	1140	7.4%	67 (59, 74)	100% COPD 60.4% Obesity 55.4% Mental health conditions
8	Gastrointestinal bleeding with obesity and mental health conditions	889	5.7%	69 (56, 78)	100% Gastrointestinal bleeding 42.1% Obesity 34.9% Mental health conditions
9	Abdominal and orthopedic surgeries with obesity	909	5.9%	61 (51, 70)	66.7% Abdominal surgery 60.8% Obesity 48.0% Orthopedic surgery
10	Cancer with obesity and mental health conditions	1266	8.2%	67 (58, 74)	100% Cancer 47.7% Obesity 33.9% Mental health conditions

^aPrevalence of all 17 conditions in each cluster is available in the Appendix.

Table 2 provides a description of cluster membership using Ward's algorithm. Of the 10 clusters, several described groups of patients with well-known and highly prevalent comorbidities such as diabetes and obesity, cardiac disease and obesity, renal disease and diabetes, and the multiple diseases and conditions common in the frail elderly. Other clusters were defined by combinations of comorbidities that are less frequently described including abdominal and orthopedic surgeries with obesity. Two clusters included individuals who were substan-

tially younger than the others: mental health conditions and chronic pain, and mental health conditions and obesity. Two single conditions were highly prevalent across all groups: mental health conditions (primarily depression) and obesity. These conditions were present in all clusters with prevalence rates that ranged from 28% to 100% (mental health conditions) and from 38% to 100% (obesity). Prevalence of all 17 conditions in each cluster is available in the Appendix, as are relative cost ratios.

DISCUSSION

In this investigation, we demonstrated the use of cluster analysis to identify distinct subgroups of patients with specific combinations of co-occurring conditions in a managed care population. Exploratory by nature, cluster analysis provided a unique method for investigating the co-occurrence of multiple conditions. When working with large data sets, simple tabulations of the number of chronic conditions can be difficult to interpret. In our population in which cohort members had at least 2 of 17 conditions of interest, there are theoretically over 100,000 different possible combinations of coexisting conditions. In actuality, our cohort only had 1507 different combinations of the 17 conditions of interest. Even this number of naturally occurring clusters is much too large to allow identification of targeted care management opportunities.

Compared with people with only 1 chronic condition, persons with multiple chronic conditions experience worse health outcomes, including, but not limited to, lower quality of life, poorer functional status, and excess morbidity and mortality.^{3,45-49} Chronic conditions do not occur in isolation, yet many existing care management strategies are directed toward single conditions. Cluster analysis provided a data-driven approach to identifying 10 clinically relevant groups of patients with patterns of comorbidities that could be targeted with new, enhanced care management strategies.

In our analysis, the largest cluster was characterized by older age coupled with conditions associated with frailty, such as stroke and dementia, along with chronic conditions including diabetes, obesity, and cardiac disease. Along with this group, our analysis produced several other clusters of well-known, high-cost co-occurring conditions for which chronic disease management programs are widespread, such as cancer with obesity and/or mental health conditions (primarily depression) and cardiac disease co-occurring with obesity and diabetes. Identification of these expected groups in our analysis provided additional assurance of the validity of our data mining method.

Our analysis produced several clusters that suggest potential care management opportunities in our managed care setting, especially in relatively younger adults. Two clusters from this cohort of individuals had median ages of 50 years (cluster 4) and 55 years (cluster 1). All 1676 members of cluster 4 had both mental health conditions and obesity, but none of the other 15 conditions of interest. This highly homogeneous cluster represented 10.8% of the entire study cohort. Individuals in cluster 1 were characterized by chronic pain co-occurring with mental health conditions and/or obesity. Chronic pain is associated with depression and anxiety, and has been linked to increased utilization of health resources.^{50,51}

Coordinated care strategies to relieve depression and improve chronic pain outcomes have been evaluated to a limited extent.^{52,53} Cohort members in cluster 1 also had the third highest median total cost of care of all clusters, after the clusters of cancer and surgical patients. Identification of these 2 clusters supports the notion that multimorbidity and patient complexity are not limited to the elderly.⁵⁴ It is likely that the relatively younger adults in these 2 groups would benefit from care management that emphasizes integrated depression care.^{55,56} Finally, we noted a high prevalence of mental health conditions and obesity across all clusters, suggesting that these 2 conditions should be addressed in most, if not all, comprehensive care management programs.

Our study has several limitations. Cluster analysis is an exploratory classification method that is supported by a relatively small body of statistical evidence, and different clustering algorithms produce different results.¹⁵ Knowing this inherent problem, we tried several algorithms. We chose the ultimate clustering solution based on subjective review and clinical relevance of cluster membership. In addition, agglomerative hierarchical clustering forces everyone into a cluster. There are likely small groups of patients within some of our clusters who have unique combinations of comorbid conditions that differ from the majority of people in that cluster. To assess the consistency of the clustering process within our patient population, we compared results from a randomly split data set of cohort members in 2006 and 2007. To further assess the stability of these clusters over time, analyses should be conducted on cohorts from different years. As with any investigation, the characteristics of our clusters are limited to our data and setting. Replicating these analyses in other settings and other patient populations may potentially reveal different clusters. However, these differences would and should inform management strategies specific to populations in those settings.

An additional limitation is that we identified the 17 conditions of interest based on electronic diagnosis and procedure data in 2006 and 2007, with the exception of cancer, for which we also included 2005 data and used a cancer registry, and obesity, for which we used BMI data. We anticipate that for most chronic conditions, such as COPD or diabetes, the member would have had at least 1 diagnosis in these years; however, the sensitivity and specificity of diagnostic data are imperfect, leading to potential misclassification of conditions. Lastly, it is likely that individuals in our cohort carried multiple other diagnoses not specifically assessed. Based on the Quan adaptation of the Elixhauser comorbidity index, cohort members had a median of 5 chronic conditions.³⁵ We selected the 17 conditions examined in this study based on prevalence in the general population and in our specific cohort, and on prior knowledge of or association with increased healthcare

utilization; however, it is likely that including other conditions in the analysis would result in different clusters or subgroups.

In this investigation, we demonstrated how cluster analysis can be used to identify homogeneous groups of complex patients from a large heterogeneous population. Such data mining methods can be applied in other settings where electronic diagnosis data are readily available. Alternatively, it is possible to use the conceptual results from this investigation to increase awareness of the need for a diverse array of care management services for individuals with high levels of healthcare utilization. However, further understanding of the care management needs of clusters of patients with similar comorbidities is warranted before designing specific tailored interventions. For example, the specific needs of persons with mental health conditions and obesity should be explored in detail in order to develop relevant care management strategies for that group.

Cluster analysis methods can be leveraged for targeted care management interventions designed to improve health outcomes and potentially lower healthcare costs. This cluster analysis of a large cohort of individuals with multiple morbidities suggests that complex patients with high healthcare utilization represent a highly diverse group of individuals. While some subgroups may respond well to existing approaches to care management (such as those designed for populations with diabetes, cardiac disease, and frailty), others will likely require new and/or individualized care management strategies to achieve favorable health outcomes.

Author Affiliations: From Institute for Health Research (SRN, JFS, EAB), Kaiser Permanente Colorado, Denver; Department of Family Medicine (EAB), University of Colorado, Aurora.

Funding Source: Supported by the Agency for Healthcare Research and Quality: K08 HS015476.

Author Disclosures: Preliminary findings of this study were presented in poster format at the American Public Health Association Annual Meeting 2010; November 8, 2010; Denver, CO. The authors (SRN, JFS, EAB) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (SRN, JFS, EAB); acquisition of data (SRN); analysis and interpretation of data (SRN, JFS, EAB); drafting of the manuscript (SRN, JFS, EAB); critical revision of the manuscript for important intellectual content (SRN, JFS, EAB); statistical analysis (SRN); obtaining funding (EAB); and supervision (EAB).

Address correspondence to: Sophia Raff Newcomer, MPH, Institute for Health Research, Kaiser Permanente Colorado, 10065 E Harvard Ave, Ste 300, Denver, CO 80231. E-mail: Sophia.r.newcomer@kp.org.

REFERENCES

- Anderson GF. Physician, public, and policymaker perspectives on chronic conditions. *Arch Intern Med.* 2003;163:437-442.
- Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med.* 2002;162(20):2269.
- Parekh AK, Barton MB. The challenge of multiple comorbidity for the US health care system. *JAMA.* 2010;303(13):1303.
- Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. *JAMA.* 2002;288(14):1775-1779.
- Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. *JAMA.* 2002;288(15):1909-1914.
- Rundall TG, Shortell SM, Wang MC, et al. As good as it gets? chronic care management in nine leading US physician organizations. *BMJ.* 2002;325(7370):958-961.
- Tsai AC, Morton SC, Mangione CM, Keeler EB. A meta-analysis of interventions to improve care for chronic illnesses. *Am J Manag Care.* 2005;11(8):478-488.
- GESICA Investigators. Randomised trial of telephone intervention in chronic heart failure: DIAL trial. *BMJ.* 2005;331(7514):425.
- Griffiths C, Foster G, Barnes N, et al. Specialist nurse intervention to reduce unscheduled asthma care in a deprived multiethnic area: the east London randomised controlled trial for high risk asthma (ELECTRA). *BMJ.* 2004;328(7432):144.
- Sochalski J, Jaarsma T, Krumholz HM, et al. What works in chronic care management: the case of heart failure. *Health Aff (Millwood).* 2009;28(1):179-189.
- Katz BP, Holmes AM, Stump TE, et al. The Indiana Chronic Disease Management Program's impact on Medicaid claims: a longitudinal, statewide evaluation. *Med Care.* 2009;47(2):154-160.
- Boult C, Dowd B, McCaffrey D, Boult L, Hernandez R, Krulwich H. Screening elders for risk of hospital admission. *J Am Geriatr Soc.* 1993;41(8):811.
- Dorr DA, Jones SS, Burns L, et al. Use of health related, quality of life metrics to predict mortality and hospitalizations in community dwelling seniors. *J Am Geriatr Soc.* 2006;54(4):667-673.
- Inouye SK, Zhang Y, Jones RN, et al. Risk factors for hospitalization among community-dwelling primary care older patients: development and validation of a predictive model. *Med Care.* 2008;46(7):726.
- Aldenderfer MS, Blashfield RK. *Cluster Analysis: Quantitative Applications in the Social Sciences.* Beverly Hills, CA: Sage Publications; 1984.
- Punj G, Stewart DW. Cluster analysis in marketing research: review and suggestions for application. *J Marketing Res.* 1983;20(2):134-148.
- Braet C, Beyers W. Subtyping children and adolescents who are overweight: different symptomatology and treatment outcomes. *J Consult Clin Psychol.* 2009;77(5):814-824.
- Gerlinger C, Wessel J, Kallschnigg G, Endrikat J. Pattern recognition in menstrual bleeding diaries by statistical cluster analysis. *BMC Womens Health.* 2009;9(1):21.
- Cornell JE, Pugh JA, Williams JW Jr, et al. Multimorbidity clusters: clustering binary data from a large administrative medical database. *Appl Multivariate Res.* 2009;12(3):163.
- Goldstein G, Luther JF, Jacoby AM, Haas GL, Gordon AJ. A taxonomy of medical comorbidity for veterans who are homeless. *J Health Care Poor Underserved.* 2008;19(3):991-1005.
- Marengoni A, Rizzuto D, Wang HX, Winblad B, Fratiglioni L. Patterns of chronic multimorbidity in the elderly population. *J Am Geriatr Soc.* 2009;57(2):225-230.
- Estabrooks PA, Shetterly S. The prevalence and health care use of overweight children in an integrated health care system. *Arch Pediatr Adolesc Med.* 2007;161(3):222.
- Mudge AM, Kasper K, Clair A, et al. Recurrent readmissions in medical patients: a prospective study. *J Hosp Med.* 2011;6(2):61-67.
- de Boer AGEM, Wijker W, de Haes HCJM. Predictors of health care utilization in the chronically ill: a review of the literature. *Health Policy.* 1997;42(2):101-115.
- Forrest CB, Lemke KW, Bodycombe DP, Weiner JP. Medication, diagnostic, and cost information as predictors of high-risk patients in need of care management. *Am J Manag Care.* 2009;15(1):41.
- Cohen-Mansfield J, Pawlson G, Lipson S, Volpato S. The measurement of health: a comparison of indices of disease severity. *J Clin Epidemiol.* 2001;54(11):1094-1102.
- Lee SJ, Lindquist K, Segal MR, Covinsky KE. Development and validation of a prognostic index for 4-year mortality in older adults. *JAMA.* 2006;295(7):801.
- Lyon D, Lancaster GA, Taylor S, Dowrick C, Chellaswamy H. Predicting the likelihood of emergency admission to hospital of older people: development and validation of the Emergency Admission Risk Likelihood Index (EARLI). *Fam Pract.* 2007;24(2):158.
- Fortinsky RH, Madigan EA, Sheehan TJ, Tullai-McGuinness S, Fenster JR. Risk factors for hospitalization among Medicare home care patients. *West J Nurs Res.* 2006;28(8):902.

30. **Donnan PT, Dorward DWT, Mutch B, Morris AD.** Development and validation of a model for predicting emergency admissions over the next year (PEONY): a UK historical cohort study. *Arch Intern Med.* 2008;168(13):1416-1422.
31. **Desai MM, Bogardus ST Jr, Williams CS, Vitagliano G, Inouye SK.** Development and validation of a risk adjustment index for older patients: the high risk diagnoses for the elderly scale. *J Am Geriatr Soc.* 2002;50(3):474-481.
32. **Piccirillo JF, Tierney RM, Costas I, Grove L, Spitznagel EL.** Prognostic importance of comorbidity in a hospital-based cancer registry. *JAMA.* 2004;291(20):2441.
33. **Satish S, Winograd CH, Chavez C, Bloch DA.** Geriatric targeting criteria as predictors of survival and health care utilization. *J Am Geriatr Soc.* 1996;44(8):914-921.
34. **Fried LF, Shlipak MG, Crump C, et al.** Renal insufficiency as a predictor of cardiovascular outcomes and mortality in elderly individuals. *J Am Coll Cardiol.* 2003;41(8):1364-1372.
35. **Quan H, Sundararajan V, Halfon P, et al.** Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care.* 2005;43(11):1130.
36. **Johnson DE.** *Applied Multivariate Methods for Data Analysts.* Pacific Grove, CA: Duxbury Press; 1998.
37. **SAS Institute Inc.** SAS/STAT 9.1 User's Guide. Cary, NC: SAS Institute Inc; 2004.
38. **Smit ES, Hoving C, De Vries H.** Does a typical contemplator exist? three clusters of smokers in contemplation. *Health Educ Res.* 2010; 25(1):61.
39. **Walker LO.** Low-income women's reproductive weight patterns: empirically based clusters of prepregnant, gestational and postpartum weights. *Womens Health Issues.* 2011;19:398-405.
40. **Sanchez-Orturo MM, Edinger JD.** A penny for your thoughts: patterns of sleep-related beliefs, insomnia symptoms and treatment outcome. *Behav Res Ther.* 2010;48(2):125.
41. **Rhee H, Holditch-Davis D, Miles MS.** Patterns of physical symptoms and relationships with psychosocial factors in adolescents. *Psychosom Med.* 2005;67(6):1006.
42. **Peretti-Watel P, Garelik D, Baron G, Spire B, Ravaud P, Duval X.** Smoking motivations and quitting motivations among HIV-infected smokers. *Antivir Ther.* 2009;14:781-787.
43. **Ott CH, Lueger RJ, Kelber ST, Prigerson HG.** Spousal bereavement in older adults: common, resilient, and chronic grief with defining characteristics. *J Nerv Ment Dis.* 2007;195(4):332.
44. **Lance GN, Williams WT.** A general theory of classificatory sorting strategies: 1. Hierarchical systems. *The Computer Journal.* 1967;9(4):373.
45. **Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B.** Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet.* 2007;370(9590):851-858.
46. **Egede LE.** Major depression in individuals with chronic medical disorders: prevalence, correlates and association with health resource utilization, lost productivity and functional disability. *Gen Hosp Psychiatry.* 2007;29(5):409-416.
47. **Felker B, Katon W, Hedrick SC, et al.** The association between depressive symptoms and health status in patients with chronic pulmonary disease. *Gen Hosp Psychiatry.* 2001;23(2):56-61.
48. **Katon WJ, Lin EHB, Williams LH, et al.** Comorbid depression is associated with an increased risk of dementia diagnosis in patients with diabetes: a prospective cohort study. *J Gen Intern Med.* 2010;25(5): 423-429.
49. **Gadalla TM.** Association of obesity with mood and anxiety disorders in the adult general population. *Chronic Dis Can.* 2009;30(1): 29-36.
50. **Roy-Byrne PP, Davidson KW, Kessler RC, et al.** Anxiety disorders and comorbid medical illness. *Gen Hosp Psychiatry.* 2008;30(3): 208-225.
51. **Arnow BA, Blasey CM, Lee J, et al.** Relationships among depression, chronic pain, chronic disabling pain, and medical costs. *Psychiatr Serv.* 2009;60(3):344.
52. **Pols RG, Battersby MW.** Coordinated care in the management of patients with unexplained physical symptoms: depression is a key issue. *Med J Aust.* 2008;188(12):133.
53. **Dobscha SK, Corson K, Leibowitz RQ, Sullivan MD, Gerrity MS.** Rationale, design, and baseline findings from a randomized trial of collaborative care for chronic musculoskeletal pain in primary care. *Pain Med.* 2008;9(8):1050-1064.
54. **Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L.** Prevalence of multimorbidity among adults seen in family practice. *Ann Fam Med.* 2005;3(3):223-228.
55. **Simon G.** Collaborative care for mood disorders. *Curr Opin Psychiatry.* 2009;22(1):37.
56. **Kroenke K, Theobald D, Wu J, et al.** Effect of telecare management on pain and depression in patients with cancer: a randomized trial. *JAMA.* 2010;304(2):163. ■

Cluster Analysis for Targeted Care Management

■ Appendix. Description and Prevalence of 17 Common Conditions by Cluster and by Clustering Method Used

1a. Ward's Minimum Variance Method

Cluster Number and Description	(1) Chronic pain with mental health conditions	(2) Diabetes with obesity and mental health conditions	(3) Kidney disease with diabetes and obesity	(4) Mental health conditions and obesity in younger adults	(5) Frailty in the elderly	(6) Cardiac disease and obesity	(7) Chronic obstructive pulmonary disease with obesity and mental health conditions	(8) Gastrointestinal bleeding with obesity and mental health conditions	(9) Abdominal and orthopedic surgeries with obesity	(10) Cancer with obesity and mental health conditions
Number of members	1017	1855	2164	1676	2788	1776	1140	889	909	1266
Percent of total cohort	6.6%	12.0%	14.0%	10.8%	18.0%	11.5%	7.4%	5.7%	5.9%	8.2%
Percent of cluster members with the following conditions:										
Obesity	47.2	86.0	50.9	100.0	37.7	54.2	60.4	42.1	60.8	47.7
Mental health conditions	69.2	44.0	28.1	100.0	45.4	34.4	55.4	34.9	33.0	33.9
Diabetes	20.3	100.0	51.2	0.0	39.7	39.4	41.1	28.9	20.0	25.4
Cardiac disease	12.6	0.0	42.4	0.0	30.7	100.0	7.8	30.2	14.2	18.8
Chronic obstructive pulmonary disease	14.3	0.0	24.5	0.0	24.1	31.7	100.0	25.3	9.9	21.3
Kidney disease (not requiring dialysis)	11.0	0.0	99.9	0.0	26.7	0.0	4.7	21.8	8.4	17.7
Cancer	4.0	0.0	<1.0	0.0	2.6	0.0	<1.0	2.10	15.5	100.0
Gastrointestinal bleeding	5.9	0.0	3.1	0.0	3.8	1.3	0.0	100.0	5.2	6.3
Chronic Pain	99.8	0.0	<1.0	0.0	4.0	1.0	3.7	<1.0	1.1	<1.0
Stroke	5.0	0.0	<1.0	0.0	35.2	<1.0	<1.0	6.9	1.7	4.4
Skin Ulcer	6.1	0.0	<1.0	0.0	26.4	<1.0	0.0	1.7	<1.0	<1.0
Fall	<1.0	0.0	<1.0	0.0	22.4	<1.0	<1.0	<1.0	<1.0	3.0
Dementia	<1.0	0.0	<1.0	0.0	25.8	<1.0	<1.0	6.2	<1.0	1.8
Abdominal surgery	3.6	0.0	0.0	0.0	<1.0	0.0	<1.0	1.0	66.7	<1.0
Orthopedic surgery	2.6	0.0	0.0	0.0	1.0	0.0	0.0	<1.0	48.0	<1.0
Back surgery	<1.0	0.0	<1.0	0.0	5.4	<1.0	0.0	<1.0	<1.0	0.0
Hip fracture	<1.0	0.0	<1.0	0.0	4.7	0.0	0.0	<1.0	<1.0	<1.0
Cost Ratio^a	1.62	1.00	1.33	1.03	1.55	1.46	1.18	1.55	2.17	2.18

^aThe total cost of care in 2006 and 2007 for each cohort member was used to calculate a median total cost of care for each cluster. The cost ratio for the cluster with the lowest median cost was set to 1.00 (cluster 2). The cost ratios for the other clusters were calculated as the median total cost of care for the cluster divided by the median total cost of care of cluster 2. The cost ratios allow a relative comparison of median cost of care between clusters.

1b. Flexible Beta Method With Beta Set to -0.25

Cluster Number and Description	(1) Chronic pain with mental health conditions	(2) Mental health conditions, diabetes and obesity, some falls	(3) Diabetes, kidney disease and cardiac disease, some skin ulcers	(4) Obesity, mental health conditions, and cancer	(5) Stroke with diabetes, mental health conditions, and obesity	(6) Dementia with mental health and other conditions	(7) Large non-specific group including obesity, cardiac disease, chronic obstructive pulmonary disease, mental health conditions, diabetes, and kidney disease	(8) Gastrointestinal bleeding with mental health and other conditions	(9) Obesity, diabetes, and other conditions	(10) Back surgery with obesity and mental health conditions
Number of members	935	1243	1643	3157	851	613	4249	733	1912	144
Percent of total cohort	6.0%	8.0%	10.6%	20.4%	5.5%	4.0%	27.4%	4.7%	12.4%	0.9%
Percent of cluster members with the following conditions:										
Obesity	50.3	56.5	25.4	78.1	37.5	27.4	63.2	29.7	82.1	46.5
Mental health conditions	72.4	82.1	28.1	68.5	38.2	54.8	39.1	43.0	17.1	45.8
Diabetes	23.2	76.1	66.4	10.3	40.3	30.5	36.7	36.7	64.1	26.4
Cardiac disease	15.0	7.5	54.8	7.1	30.7	35.1	50.3	40.2	6.3	9.0
Chronic obstructive pulmonary disease	20.7	7.7	16.4	8.2	26.8	31.3	49.3	30.0	3.4	11.1
Kidney disease (not requiring dialysis)	3.0	7.4	61.5	6.3	31.1	29.4	35.0	28.8	4.3	5.6
Cancer	1.8	3.1	<1.0	41.8	8.0	1.3	<1.0	<1.0	2.6	3.5
Gastrointestinal bleeding	5.8	1.2	2.5	8.3	1.9	10.6	1.1	100.0	1.9	2.8
Chronic Pain	99.8	2.1	1.4	<1.0	<1.0	3.4	2.3	2.6	3.1	8.3
Stroke	5.9	4.9	<1.0	<1.0	98.7	9.0	1.2	8.3	1.0	2.8
Skin Ulcer	5.1	1.4	37.8	1.1	5.1	3.1	<1.0	2.0	1.2	<1.0
Fall	1.3	34.4	3.0	<1.0	<1.0	15.5	1.1	4.1	<1.0	2.1
Dementia	<1.0	1.8	3.2	<1.0	7.5	99.7	<1.0	0.0	<1.0	<1.0
Abdominal surgery	0.0	<1.0	<1.0	2.8	<1.0	<1.0	<1.0	<1.0	28.7	2.08
Orthopedic surgery	<1.0	<1.0	<1.0	<1.0	<1.0	<1.0	<1.0	<1.0	24.0	1.39
Back surgery	<1.0	0.0	0.0	0.0	<1.0	0.0	<1.0	<1.0	<1.0	100.0
Hip Fracture	<1.0	<1.0	<1.0	<1.0	<1.0	1.6	2.4	0.0	<1.0	0.0

(Continued)

■ **Appendix.** Description and Prevalence of 17 Common Conditions by Cluster and by Clustering Method Used
(Continued)

1c. Flexible Beta Method With Beta Set to -0.5

Cluster Number and Description	(1) Chronic pain with mental health conditions	(2) Obesity with mental health conditions, diabetes and other conditions	(3) Cardiac disease, obesity, kidney disease, diabetes, and mental health conditions	(4) Dementia with mental health and other conditions	(5) Gastrointestinal bleeding with obesity and mental health conditions	(6) Abdominal and orthopedic surgeries with obesity	(7) Cancer with obesity and other conditions	(8) Stroke with chronic obstructive pulmonary disease, mental health conditions, and other conditions	(9) Fall with mental health and other conditions	(10) Skin ulcer with diabetes, obesity, and other conditions
Number of members	760	6799	3387	500	703	768	987	483	456	637
Percent of total cohort	4.9%	43.9%	21.9%	3.2%	4.5%	5.0%	6.4%	3.1%	2.9%	4.1%
Percent of cluster members with the following conditions:										
Obesity	33.6	74.1	56.2	17.2	46.1	60.2	43.3	27.5	37.5	44.7
Mental health conditions	70.3	56.2	35.7	51.6	44.2	35.2	31.6	44.1	50.4	29.7
Diabetes	32.8	47.2	47.5	30.6	20.6	16.0	9.1	27.7	35.1	52.4
Cardiac disease	18.0	13.1	64.5	34.0	28.3	16.3	19.0	35.4	30.7	30.8
Chronic obstructive pulmonary disease	22.1	32.5	3.8	30.8	23.5	9.4	25.2	49.5	24.6	21.4
Kidney disease (not requiring dialysis)	8.7	11.3	53.1	30.8	25.3	10.2	14.7	21.5	22.1	27.5
Cancer	5.5	4.4	<1.0	2.4	<1.0	2.7	100.0	13.5	9.6	7.7
Gastrointestinal bleeding	9.6	1.3	1.1	12.0	100.0	5.9	10.3	17.2	8.8	6.8
Chronic Pain	98.2	4.0	1.1	4.0	<1.0	2.7	<1.0	10.8	7.5	4.9
Stroke	1.7	<1.0	16.5	3.4	0.0	2.5	<1.0	99.6	2.4	3.3
Skin Ulcer	6.8	<1.0	1.9	5.8	<1.0	<1.0	0.0	1.2	2.9	99.1
Fall	1.1	<1.0	<1.0	7.6	<1.0	2.0	<1.0	13.0	99.1	6.4
Dementia	<1.0	<1.0	<1.0	100.0	0.0	1.0	<1.0	22.6	14.7	6.1
Abdominal surgery	3.3	<1.0	0.0	<1.0	0.0	65.9	9.2	<1.0	<1.0	1.7
Orthopedic surgery	3.3	<1.0	<1.0	0.0	0.0	50.4	1.8	<1.0	<1.0	1.7
Back surgery	<1.0	2.1	<1.0	0.0	0.0	<1.0	0.0	<1.0	0.0	<1.0
Hip fracture	<1.0	<1.0	0.0	<1.0	0.0	<1.0	<1.0	<1.0	1.5	<1.0