HHS Public Access

Author manuscript

Med Care. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Med Care. 2017 March ; 55(3): 276–284. doi:10.1097/MLR.0000000000000660.

Identifying Specific Combinations of Multimorbidity that Contribute to Health Care Resource Utilization: an Analytic Approach

Nicholas K. Schiltz, PhD[1], **David F. Warner, PhD**[3], **Jiayang Sun, PhD**[1], **Paul M. Bakaki, MBChB, PhD**[1], **Avi Dor, PhD**[4], **Charles W. Given, PhD**[5], **Kurt C. Stange, MD, PhD**[1,2], and **Siran M. Koroukian, PhD**[1]

[1]Department of Epidemiology & Biostatistics, Case Western Reserve University School of Medicine, Cleveland, Ohio

[2]Department of Family Medicine and Community Health, Case Western Reserve University School of Medicine, Cleveland, Ohio

[3]Department of Sociology, University of Nebraska-Lincoln, Lincoln, Nebraska

[4]Department of Health Policy and Management, George Washington University Milken Institute School of Public Health, Washington, D.C

[5]Department of Family Medicine, Michigan State University, East Lansing, Michigan

Abstract

Background—Multimorbidity affects the majority of elderly adults and is associated with higher health costs and utilization, but how specific patterns of morbidity influence resource use is less understood.

Objective—To identify specific combinations of chronic conditions, functional limitations, and geriatric syndromes associated with direct medical costs and inpatient utilization.

Corresponding Author (and reprints): Nicholas K. Schiltz, PhD, Instructor, Case Western Reserve University, Department of Epidemiology & Biostatistics, 10900 Euclid Avenue, WG-49, Cleveland, Ohio 44106-4945, Phone: 216-368-5626, nks8@case.edu.
David F. Warner, PhD, Assistant Professor, Department of Sociology, University of Nebraska-Lincoln, 726 Oldfather Hall, Lincoln, NE 68588-0324, Phone: (402) 472-3421, dwarner3@unl.edu
Jiayang Sun, PhD, Professor, Department of Epidemiology & Biostatistics, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, WG-43, Cleveland, Ohio 44106-4945, Phone: 216-368-2000, jsun@case.edu
Paul M. Bakaki, MBChB, PhD, Instructor, Department of Epidemiology & Biostatistics, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, WG-49, Cleveland, Ohio 44106-4945, Phone: 216-368-3925, pmb18@case.edu
Avi Dor, PhD, Professor, Department of Health Policy and Management, George Washington University Milken Institute School of Public Health, 950 New Hampshire Ave NW, Washington, DC, 20052, Phone: 202-994-4202, avidor@gwu.edu
Charles W. Given, PhD, Professor, Department of Family Medicine Clinical Center, Michigan State University, 788 Service Rd., #B120, East Lansing, MI 48824-7052, Phone: 517-884-0420, bill.given@hc.msu.edu
Kurt C. Stange, MD, PhD, Professor, Department of Family Medicine & Community Health, Case Western Reserve University, 11000 Cedar Ave., Ste. 402, Cleveland, OH 44106, Phone: (216) 368-6297, kcs@case.edu
Siran M. Koroukian, PhD, Associate Professor, Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, WG-49, Cleveland, OH 44106-4945, Phone: 216-368-5816, skoroukian@case.edu

An earlier version of the analysis was presented in a symposium at the Gerontological Society of America's Annual Scientific Meeting in Orlando, Florida on November 20, 2015.

Conflicts of Interest: The authors of this report have no conflicts of interest to disclose.

Design—Retrospective cohort study using the Health and Retirement Study (2008–2010) linked to Medicare claims. Analysis used machine learning techniques: classification and regression trees (CART) and random forest.

Subjects—A population-based sample of 5,771 Medicare-enrolled adults age 65 and older in the United States.

Measures—Main covariates: self-reported chronic conditions (measured as none, mild, or severe), geriatric syndromes, and functional limitations. Secondary covariates: demographic, social, economic, behavioral, and health status measures. Outcomes: Medicare expenditures in the top quartile and inpatient utilization.

Results—Median annual expenditures were $4,354, and 41% were hospitalized within two-years. The tree model shows some notable combinations: 64% of those with self-rated poor health plus ADL and IADL disabilities had expenditures in the top quartile. Inpatient utilization was highest (70%) in those age 77 – 83 with mild to severe heart disease plus mild to severe diabetes. Functional limitations were more important than many chronic diseases in explaining resource use.

Conclusions—The multimorbid population is heterogeneous and there is considerable variation in how specific combinations of morbidity influence resource use. Modeling the conjoint effects of chronic conditions, functional limitations, and geriatric syndromes can advance understanding of groups at greatest risk and inform targeted tailored interventions aimed at cost-containment.

Keywords

Comorbidity; Health care costs; Utilization; Chronic Disease; Functional Status

INTRODUCTION

More than 75% of adults over age 65 have multiple (two or more) concurrent chronic conditions, accounting for 93% of total Medicare expenditures.[1] The positive association between the number of conditions and health expenditures and utilization is well established.[2–4] Less known is the relationship between specific combinations of conditions and resource use.[5] It is highly likely, however, that different combinations of conditions affect expenditures differently.[6]

Many studies measure multimorbidity as a count of the chronic conditions.[7–10] The limitation of this approach is that it gives each condition equal weight, even though that may not reflect the actual effect of each condition on different outcomes. Other approaches have used multivariable models to model health expenditures and utilization by multiple individual conditions.[11,12] The limitation with this approach is that it assumes an additive relationship and does not account for the effect of possible nonlinear, non-additive co-occurrence of these conditions. Most importantly, given the reliance of most previous studies on administrative claims data alone, multimorbidity has been equated with multiple chronic conditions, without recognizing the simultaneous presence of functional limitations and geriatric syndromes, which are highly prevalent in older adults.[13] Indeed, the fact that chronic conditions often co-occur with functional limitations and geriatric syndromes

reflects highly complex, interdependent, and bidirectional relationships among these conditions, with considerable consequences for health outcomes, health care utilization, and costs.[13,14]

An additional layer of complexity arises with the possible combinations of conditions, the sheer number of which makes traditional analytical approaches (e.g., multivariable regression models) inadequate, especially when the goal is to identify combinations of chronic conditions, functional limitations and/or geriatric syndromes that are associated with high resource use. A possible solution may be found in machine learning methods developed to mine information from large complex datasets, allowing for such combinations of conditions to emerge empirically from the data.[15,16]

Therefore, the aim of this study is to use a data-driven approach to identify specific combinations of chronic conditions, functional limitations, and/or geriatric syndromes, that — in addition to socio-demographic factors and self-rated health status — are most highly associated with Medicare expenditures and inpatient utilization in older adults. The findings will help to identify subgroups of older adults that can be targeted for tailored interventions aimed at cost-containment.

METHODS

This is a retrospective study using existing national panel data linked with Medicare claims data. The Institutional Review Board of Case Western Reserve University approved the study. The Centers for Medicare and Medicaid Services and the University of Michigan granted permission to use the data for this study.

Data Source and Study Population

The Health and Retirement Study (HRS) is a biennial longitudinal panel of a U.S. representative sample of adults 50 years of age or older.[17] In each wave, the HRS collects data on a broad array of health measures including demographics; chronic conditions; self-reported health status; functional status, including limitations in activities of daily living (ADL), and instrumental activities of daily living (IADL); and geriatric syndromes. These comprehensive surveys are conducted both in-person and over the phone, and subjects are compensated for their participation. A total of 37,319 respondents have participated in the survey since its inception.

Medicare-enrolled respondents participating in HRS are asked for permission to link the survey with their Medicare claims records, of which over 80% have agreed.[18] Those that agreed are slightly older than those that did not (mean age 75.6 vs 70.9, $p<.001$), but it is unknown if other selection biases exist. The HRS and Medicare data are linked by the Medicare ID and Social Security number by the Centers for Medicaid and Medicare Services. We used 2008–2012 Medicare claims data to calculate annual Medicare expenditures. The expenditure data include all services covered by Medicare under Part A and B —inpatient, outpatient, office visits and other professional services, hospice care, skilled nursing facilities, and durable medical equipment.

The unit of analysis for this study were people that participated in the HRS survey in 2008 and whom had linked Medicare claims data. The time period of follow-up in the Medicare claims begins at the month following the survey, and ends at the next survey wave or death. On average the follow-up time is 23.9 months (median: 24, interquartile range: 21 – 26). In total there were 15,235 surveys from living respondents in 2008, and of these 9,955 were able to be linked to Medicare claims. We then excluded a) 3,375 who were enrolled in a Medicare Health Maintenance Organization (HMO) plan at any time during the post-survey follow-up period as complete cost data was not available for these respondents; b) an additional 286 who were under age 65 at the time of the HRS survey; c) another 256 that were enrolled in Medicare Part A and B for zero months; and d) 267 that were residing in a nursing home. The final study population was 5,771 people. We also used a testing dataset to validate our models consisting of the 2010 HRS survey linked to 2010 – 2012 Medicare data. The same inclusion and exclusion criteria was applied to this dataset resulting in 5,186 subjects. These were the most recent years of HRS-Medicare linked data available.

Outcomes

The primary outcomes of interest were a) health care expenditures, measured as the per-member-per-month (PMPM) Medicare amount reimbursed for each person-wave, and b) inpatient utilization, measured as a claim for a hospital stay in the person-wave. PMPM was calculated by taking the total Part A and Part B expenditures accrued in the follow-up period and dividing by the number of months a respondent was enrolled in Medicare fee-for-service. We then dichotomized health expenditures into those in the top quartile of PMPM (high cost) and those in the bottom three quartiles (low to medium cost). We also calculated median and average annual expenditures for each person-wave by extrapolating the PMPM cost out to 12 months.

Primary Covariates

Our primary covariates of resource use were self-reported morbidity, broadly classified into three categories: chronic conditions, functional limitations, and geriatric syndromes. These definitions have been used previously.[13] Chronic conditions were defined as a binary indicator for whether the respondent was ever told by a physician that s/he had hypertension, heart disease, lung disease, diabetes, stroke, arthritis, cancer, or psychiatric conditions. Additional questions assessed if a condition caused limitations in usual activities and/or whether the respondent was receiving treatment for a given condition (e.g., medication for hypertension, or oxygen for lung disease). These variables allowed us to categorize each condition as either no disease, mild disease as indicated by self-reported diagnosis alone, or severe disease.[14]

Functional limitations were grouped in the following categories: a) Strength limitations (difficulty pulling/pushing a large object; lifting ten pounds; rising from chair; and sitting for two hours); b) upper body limitations (difficulty picking up a dime; reaching overhead); c) lower body limitations (difficulty walking one or several blocks; going up one or several flights; stooping, kneeling, or crouching); d) Activities of Daily Living (ADL: difficulty bathing; eating; transferring in and out of bed; walking across the room; and dressing); and

Instrumental Activities of Daily Living (IADL: difficulty preparing meals, taking medications, managing money, grocery shopping).

Geriatric syndromes, which are conditions commonly experienced by older individuals[19] included: a) visual impairment (rated by the respondent as fair or poor even when wearing corrective lenses as usual, or legally blind); b) hearing impairment (rated by the respondent as fair or poor even when wearing hearing aid as usual); c) moderate or severe depressive symptoms (4 or more symptoms on the modified 8-item Center for Epidemiologic Studies Depression Scale [CES-D])[20]; d) urinary incontinence; e) low cognitive performance (bottom third of a 35-point scale[21] designed to measure working memory, mental processing speed, knowledge and language, and orientation, or a proxy reporting that the respondent's cognitive performance was fair or poor;[22] and f) severe pain ("often troubled by").[23]

Other covariates of interest

Age was grouped in 5-year increments (65–69, 70–74, 75–79, 80–84, and 85+). Race/Ethnicity included four categories: White non-Hispanic, Black non-Hispanic, Hispanic, and Other. Marital status was identified as Married, Divorced, Widowed, and Never Married. Years of education were grouped in 6 categories: < 9, 9–11, 12, 13–15, 16, and ≥17. Income was adjusted for the household size, and expressed as the ratio of household income to the federal poverty level, as follows: < 100%, 100–199%, 200–299%, and ≥300%. Each of smoking status (never smoked, former smoker, and current smoker) and alcohol use (none, moderate, and heavy) included three categories. We characterized body mass index (BMI, measured as kg/m2) as underweight (BMI ≤18), normal/overweight (BMI of 18.5–30), and obese (BMI ≥ 30); in addition, 2% of respondents had missing values for BMI. Vigorous exercise was measured as a dichotomous variable to reflect engagement in vigorous sports or activities more than once a week. We also used a dichotomous indicator of whether or not a person required a proxy respondent. We also included self-rated poor health status as this has been shown to be a significant indicator of health expenditures.[24]

Statistical analyses

We used classification and regression tree (CART) analysis to identify combinations of covariates associated with the outcome of interest. CART is a nonparametric, machine-learning method that uses repeated binary partitioning of the value spaces of explanatory variables so that each partition corresponds to as homogenous outcome as possible.[25] Each covariate is considered as a potential split, including every value of an interval-level variable. Each node can split and form two child nodes, which can in turn split and create two more child nodes each. Nodes that are not split are called terminal nodes, and each study respondent can only be in one terminal node. This process continues with both tree building and pruning until all possible splits are exhausted or until some stopping criteria is met. In this study we set the following stopping criteria (based on model-tuning described below): a maximum tree depth of 5 splits, a minimum node size of 60 respondents, and required a split to increase R-square by a minimum of 0.001.

To build our model, we used the study data (2008 survey) to train our models, and the 2010 survey to validate and test our models. We also used 10-fold cross-validation repeated three

times on the training dataset testing to build the CART models.[26] We then tested the accuracy of our models on the validation data set. A bootstrap aggregation method, Random Forest, was used to determine if our CART models were capturing the most important variables related to the outcomes. The Random Forest algorithm creates multiple decision trees using random variable selection, a detailed description of which is provided by Breiman.[27] For each random forest model we created 2000 trees and sampled one-third of the explanatory variables at each node split. We compared the performance of the CART models with multivariable logistic regression using the same independent variables. We used R version 3.1.3 and the 'rpart' (CART), 'randomForest' (Random Forest), 'partykit' (graphics), and 'caret' (model tuning and cross-validation) packages for the CART analysis, and SAS version 9.3 for data management and descriptive statistics.[28–33]

RESULTS

The median annual Medicare reimbursed expenditures per person is $4,354, with a mean average cost of $13,413 per year. The large difference between the mean and median indicates that costs are right-skewed. The expenditure distribution by age exhibited a dose-response relationship with highest costs in those age 85 and above. Unadjusted expenditures and inpatient utilization are highest in Black non-Hispanics and Hispanics, those with less than a high school education, those with incomes below the federal poverty line, those with poor self-rated health, and those who needed a proxy respondent. The distribution of median annual costs and percent hospitalized by other individual characteristics is shown in Table 1.

Descriptive statistics of unadjusted annual expenditures by morbid conditions are shown in Table 2. The median annual cost is higher for every condition compared to the overall median, with the exception of mild arthritis. For chronic conditions, those with a history of moderate to severe cancer ($19,792) and those with a history of moderate to severe lung disease ($16,528) have the highest median costs. Among functional limitations, those with activities of daily of living limitations ($20,539) have the highest median cost, followed by those with instrumental ADL limitations ($11,735). Respondents with severe poor cognitive functioning have the highest median cost ($17,309) compared to other geriatric syndromes. Every chronic condition, geriatric syndrome, and functional limitation is associated with an unadjusted higher percentage hospitalized compared to the total population, again with the exception of mild arthritis. Many of the same conditions with the highest median cost are also those with elevated percentage of hospitalization, as evidence that inpatient utilization is an important driver of health care costs.

Figure 1 shows the CART analysis for being in the top quartile of annual expenditures. The tree shows how different “paths” or “rules” down the tree are associated with varying levels of annual expenditures. Those with the combination of self-rated poor health plus IADL limitations plus ADL limitations (n=222) are in the top quartile of medical expenditures 63.5% of the time. On the other hand, those in self-rated good health and under age 75 (n=2,080) are in the top quartile only 12.3% of the time. Other notable rules include that 63.4% of those over age 86 with self-rated poor health and IADL but without ADL limitations (n=83) and 55.7% of those in self-rated good health over age 75, but with mild or

severe heart disease, IADL limitations, and incontinence (n=61) are in the top quartile of annual expenditures.

Figure 2 shows the CART analysis for two-year hospitalization. Certain combinations have hospitalization rates above 50%. These included self-rated poor health with history of mild to severe stroke (n=460), poor self-rated health with mild to severe heart disease and mild to severe diabetes (n=136). People in self-rated good health with no lower body mobility limitations and under age 84 have the lowest percent hospitalized at 17.7% (n=2,368).

The Random Forest shows which variables are the most important in terms of improving the accuracy of the model from a bootstrap sample of 2000 trees for each outcome (Figure 3). The following variables rank in the top seven most important variables for both outcomes: self-rated poor health, IADL, upper mobility limitation, lower mobility limitations, age, and heart disease. As most of these variables appear frequently in the CART models in Figure 1 and 2, it lends further validity to those models. It also demonstrates the importance of functional limitations as a factor explaining high expenditures and utilization.

Measures of CART model performance are shown in the supplemental material. The c-statistic for the expenditure model was 0.698 and for hospitalization was 0.699. For comparison the c-statistic for a logistic regression model was 0.696 (expenditures) and 0.731 (hospitalization).

CONCLUSIONS

In this study we analyzed Medicare reimbursed expenditures and inpatient utilization in a U.S. representative sample population of older adults, and identified specific combinations of chronic conditions, functional limitations, and geriatric syndromes that are most highly associated with high and low costs and utilization. Our analysis identifies *empirically emerging* combinations of conditions that are associated with high costs and inpatient utilization than the average. An important finding is that while chronic conditions are important covariates associated with resource use, functional limitations emerge prominently in our models. This shows the importance of accounting for functional limitations in research on health care resource use; however, these measures are not available in claims data analysis. The unique HRS-Medicare linked database is able to account for functional limitations along with an array of other factors including geriatric syndromes, behavioral factors, and self-reported health status.

These findings have important implications both in clinical practice and in research. First, with regard to clinical practice, our findings highlight the importance of evaluating an older individual’s health not just based on the presence of (multiple) chronic conditions, but also a number of additional factors that affect health status and health care use. Identifying people at highest risk based on multiple domains of risk factors is likely to be more robust than merely counting diseases, in order to target individuals who may benefit from tailored case management interventions aimed at preventing decline and containing cost.

With respect to the implications for research, our findings strongly support the use of data on functional limitations as a way to project health care resource use. Medicare claims data

alone are likely to be deficient when it comes to identifying functional limitations with ICD-9-CM coding.[34] Models relying solely on chronic conditions are likely to be deficient in identifying older adults who will incur high costs to the Medicare program. A previous study showed that that functional limitations and geriatric syndromes are also important predictors of poor health outcomes and mortality.[14] In our study, functional limitations were important, but geriatric syndromes were less so.

To our knowledge this is the first study to identify combinations of specific conditions that are associated with high Medicare expenditures and inpatient utilization using the CART method. A key benefit of using this analytic approach lies in our ability to learn about empirically emerging combinations of conditions instead of needing to rely solely on a priori hypotheses to guide the queries. An important strength of the study is the comprehensive set of nationally-representative data from the HRS, including sociodemographic variables, behavioral factors, and a number of clinically pertinent measures, which we were able to link to Medicare claims data to provide accurate measures of Medicare expenditures and inpatient utilization. Another strength of the study is the use of self- or proxy-reported measures for all of our multimorbidity measures and self-rated health which captures how patients perceive, live, and function with their chronic conditions and disabilities.

We chose to use the CART method because our goal was to empirically identify specific combinations of conditions that influence cost and utilization, without *a priori* knowledge of what those combinations may be or the constraint of linear (additive) relationships. CART is useful here because it is a non-parametric method and good for large data. If sample size is small and/or the relationship being modeled is linear, then parametric methods like generalized linear regression (e.g. logistic regression) may be preferred. Logistic regression can handle some additive interactions or polynomial terms, but these must be specified & validated by the researcher. If prediction is the main goal, ensemble methods like random forest and support vector machines may outperform CART. However, these ensemble models are almost impossible to interpret if making inferences is the goal. Methods like CART and generalized linear regression are easier to interpret. Linear regression may be preferred over CART for hypothesis testing as the user can specify which variables should be in the model, whereas in CART the computer algorithm makes the decision. CART, on the other hand works better in situations with a large number of potential predictors.

This study has several limitations. We chose not to use claims-derived measures of morbidity history, because we have found in previous studies that it did not provide much additional information beyond self-reported measures in terms of altering the CART models.[14] Another limitation is that CART produces a single tree. This is in contrast with the random forest approach, which uses subsets of the data and variables, as well as bootstrapping, to produce many trees. On the other hand, it is not possible to use the random forest approach to identify combinations of conditions that affect an outcome, which was the primary objective of this study. Many of the same covariates that appeared in our model, are also the ones that were identified as the most important by the random forest method, lending further validity to our model. We were not able to take advantage of the longitudinal design of the Health & Retirement Study as mixed effects tree-based models have only recently developed, and the software currently available is unstable and/or cannot handle

binary outcomes.[35] The cross-sectional measurement of our explanatory variables may not capture important information on how longitudinal changes in multimorbidity impact future health care expenditures. We did not have access to geographic information in the HRS or Medicare claims, which may have improved model fit as geographic location has been shown to be an important determinant of Medicare spending.[36] We limited our study to Medicare Part A and Part B reimbursed expenditures. A CART analysis using other cost outcomes like out-of-pocket costs, or private health insurance expenditures may produce a very different tree model.

In conclusion, this study highlights that specific combinations of conditions constituting multimorbidity are associated with markedly higher health care expenditures and utilization in a U.S. representative sample of Medicare beneficiaries. The tree-based approach produces results that can aid in identifying subgroups of patients that are most at-risk of high resource use, and to tailor interventions appropriately. This approach could also be further adopted into population health management systems to provide real time forecasts of future health expenditures for individuals. Future studies may also take a longitudinal approach to investigate how trajectories of multimorbidity influence health and health expenditures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Xiaozhen Han, MS (Case Western Reserve University) for her role in exploratory data analysis and data preprocessing.

Funding Disclosure: This study was funded by the Agency for Healthcare Research and Quality (R21HS023113; PI: SM Koroukian). Dr. Schiltz and Dr. Koroukian were supported, in part, by the Clinical and Translational Science Collaborative of Cleveland, #UL1 TR000439 and #KL2 TR000440 from the National Center for Advancing Translational Sciences (NCATS) component of the National Institutes of Health. Dr. Stange's time is supported as a Scholar of The Institute of Integrative Health and as a Clinical Research Professor of the American Cancer Society. Dr. Koroukian is also supported by Grant # U48 DP005030-01S3 under the Health Promotion and Disease Prevention Research Centers Program, funded by the Centers for Disease Control and Prevention (CDC).

REFERENCES

1. Centers for Medicare and Medicaid Services. Chronic Conditions among Medicare Beneficiaries. 2012 Edition. Baltimore, MD: 2012.
2. Machlin SR, Soni A. Health care expenditures for adults with multiple treated chronic conditions: estimates from the Medical Expenditure Panel Survey, 2009. Prev Chronic Dis. 2013; 10:E63. [PubMed: 23618543]
3. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Arch Intern Med. 2002; 162(20):2269–2276. [PubMed: 12418941]
4. Yoon J, Zulman D, Scott JY, Maciejewski ML. Costs Associated With Multimorbidity Among VA Patients. Med Care. 2014; 52(Suppl 3):S31–S36. [PubMed: 24561756]
5. US Department of Health and Human Services. Multiple Chronic conditions—A Strategic Framework: Optimum Health and Quality of Life for Individuals with Multiple Chronic Conditions. Washington, DC: Department of Health and Human Services; 2010.
6. Vogeli C, Shields AE, Lee TA, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med. 2007; 22(Suppl 3): 391–395. [PubMed: 18026807]

7. Fortin M, Soubhi H, Hudon C, Bayliss EA, van den Akker M. Multimorbidity's many challenges. BMJ. 2007; 334(7602):1016–1017. [PubMed: 17510108]
8. Mercer SW, Smith SM, Wyke S, O'Dowd T, Watt GCM. Multimorbidity in primary care: developing the research agenda. Fam Pract. 2009; 26(2):79–80. [PubMed: 19287000]
9. Noël PH, Parchman ML, Williams JW, et al. The challenges of multimorbidity from the patient perspective. J Gen Intern Med. 2007; 22(Suppl 3):419–424. [PubMed: 18026811]
10. Salive ME. Multimorbidity in older adults. Epidemiol Rev. 2013; 35:75–83. [PubMed: 23372025]
11. Carstensen J, Andersson D, André M, Engström S, Magnusson H, Borgquist LA. How does comorbidity influence healthcare costs? A population-based cross-sectional study of depression, back pain and osteoarthritis. BMJ Open. 2012; 2(2):e000809.
12. Schiltz NK, Kaiboriboon K, Koroukian SM, Singer ME, Love TE. Long-term reduction of health care costs and utilization after epilepsy surgery. Epilepsia. 2016; 57(2):316–324. [PubMed: 26693701]
13. Koroukian SM, Warner DF, Owusu C, Given CW. Multimorbidity redefined: prospective health outcomes and the cumulative effect of co-occurring conditions. Prev Chronic Dis. 2015; 12:E55. [PubMed: 25906436]
14. Koroukian SM, Schiltz N, Warner DF, et al. Combinations of Chronic Conditions, Functional Limitations, and Geriatric Syndromes that Predict Health Outcomes. J Gen Intern Med. 2016 Feb.
15. Schiltz, NK., Warner, DF., Sun, J., Han, Xiaozhen, Koroukian, Siran M. Symposium: Combination of Chronic Conditions Determining Clinical Relevance and Resource Use. Vol. 55. Orlando, FL: Gerontologist; 2015. Novel use of data mining methods in multimobidity research; p. 357-357.
16. Parikh RB, Kakad M, Bates DW. Integrating predictive analytics into high-value care: The dawn of precision delivery. JAMA. 2016; 315(7):651–652. [PubMed: 26881365]
17. National Institute of Aging. Growing Older in America: The Health and Retirement Study. Washington, DC: National Institutes of Health; 2007.
18. Acumen LLC. MedRIC Documentation for HRS Data Requestors.; Date unknown. [Accessed January 12, 2015] http://hrsonline.isr.umich.edu/sitedocs/rda/cmsdocs/MedRICdocumentation.pdf.
19. Inouye SK, Studenski S, Tinetti ME, Kuchel GA. Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. J Am Geriatr Soc. 2007; 55(5):780–791. [PubMed: 17493201]
20. Cigolle CT, Ofstedal MB, Tian Z, Blaum CS. Comparing models of frailty: the Health and Retirement Study. J Am Geriatr Soc. 2009; 57(5):830–839. [PubMed: 19453306]
21. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. Neuropsychiatry Neuropsychol Behav Neurol. 1988; 1:111–117.
22. Langa KM, Larson EB, Karlawish JH, et al. Trends in the prevalence and mortality of cognitive impairment in the United States: is there evidence of a compression of cognitive morbidity? Alzheimers Dement J Alzheimers Assoc. 2008; 4(2):134–144.
23. Andrews JS, Cenzer IS, Yelin E, Covinsky KE. Pain as a risk factor for disability or death. J Am Geriatr Soc. 2013; 61(4):583–589. [PubMed: 23521614]
24. DeSalvo KB, Jones TM, Peabody J, et al. Health care expenditure prediction with a single item, self-rated health measure. Med Care. 2009; 47(4):440–447. [PubMed: 19238099]
25. Breiman, L., Friedman, J., Stone, J., Olshen, R. Classification and Regression Trees. 1. Boca Raton: Chapman and Hall/CRC; 1984.
26. James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning: With Applications in R. 1st. New York: Springer; 2013. 2013, Corr. 5th printing 2015 edition
27. Breiman L. Random Forests. Mach Learn. 2001; 45(1):5–32.
28. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria: 2014. http://www.R-project.org/
29. Therneau T, Atkinson B, Ripley B. Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-8. 2014 http://CRAN.R-project.org/package=rpart.
30. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002:18–22.

31. Hothorn, T., Zeileis, A. Partykit: A Modular Toolkit for Recursive Partytioning in R. Faculty of Economics and Statistics, University of Innsbruck; 2014. http://econpapers.repec.org/paper/innwpaper/2014-10.htm [Accessed January 13, 2016]

32. Kuhn M. Caret: Classification and Regression Training. R Package Version 6.0-41. 2015http://CRAN.R-project.org/package=caret

33. SAS Institute Inc. SAS/STAT(R) 9.3 User's Guide, Second Edition. Cary, NC, USA: SAS Institute Inc.; 2011.

34. Iezzoni LI. Using administrative data to study persons with disabilities. Milbank Q. 2002; 80(2): 347–379. [PubMed: 12101876]

35. Sela RJ, Simonoff JS. RE-EM trees: a data mining approach for longitudinal and clustered data. Mach Learn. 2011; 86(2):169–207.

36. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder ÉL. The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care. Ann Intern Med. 2003; 138(4):273–287. [PubMed: 12585825]

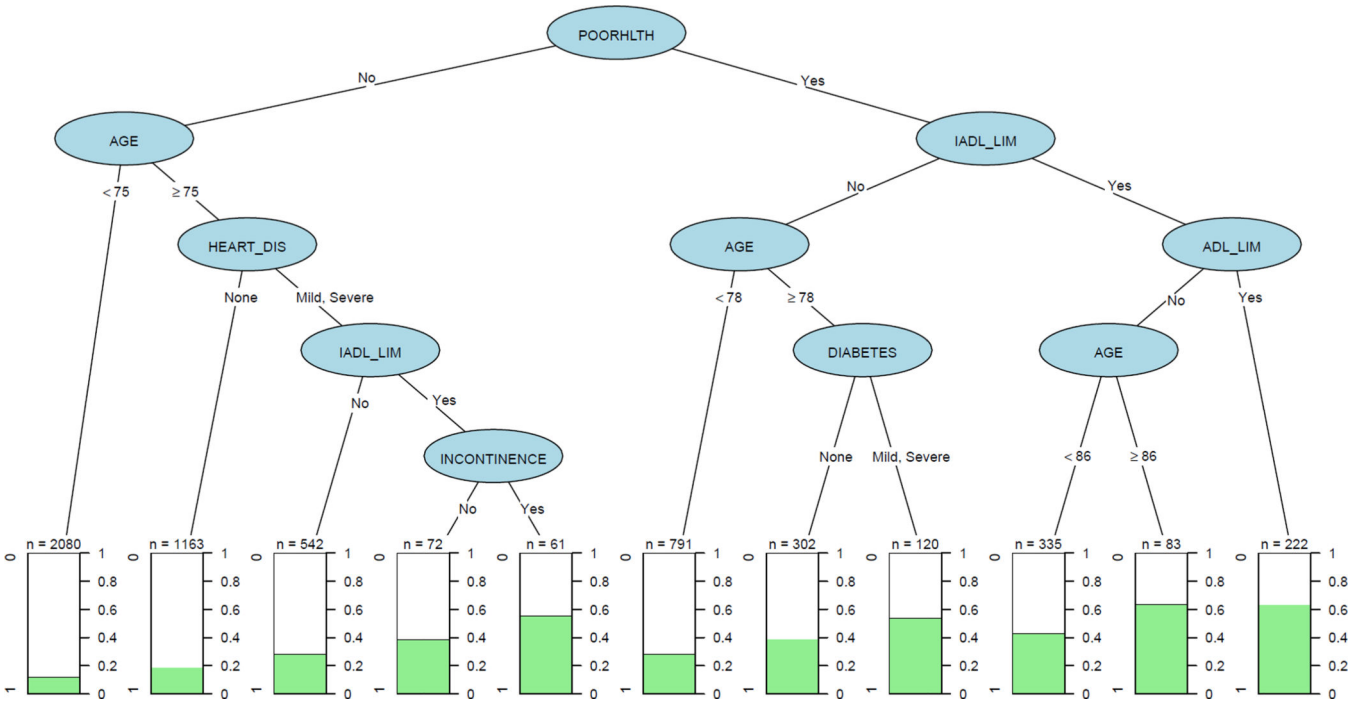

Figure 1. Classification and regression tree analysis of annual Medicare expenditures in top quartile

Each "path" in the tree concludes with a terminal node that shows the percentage of persons with that combination of characteristics that had high expenditures. For example, there were 222 people with self-rated poor health with IADL limitations and ADL limitations, and 64% of them had high Medicare expenditures. Abbreviations: POORHLTH = self-rated poor health, IADL_LIM and ADL_LIM = (Instrumental) Activities of Daily Living Limitations, HEART_DIS = heart disease.

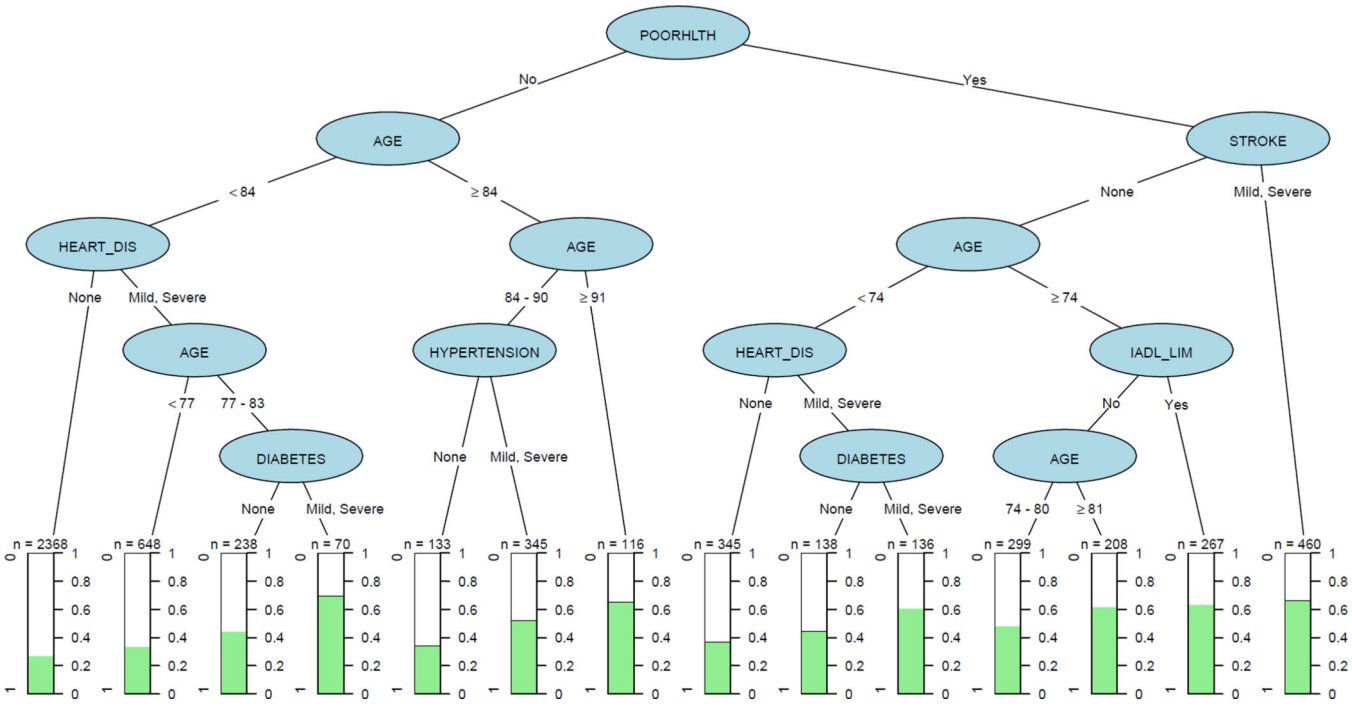

Figure 2. Classification and regression tree analysis of hospitalization

Each "path" in the tree concludes with a terminal node that shows the percentage of persons with that combination of characteristics that had an inpatient stay over a two-year period. For example, there were 460 people with self-rated poor health and history of mild to severe stroke, and 67% of them had an inpatient stay. Abbreviations: POORHLTH = self-rated poor health, IADL_LIM = Instrumental Activities of Daily Living Limitations, HEART_DIS = heart disease.

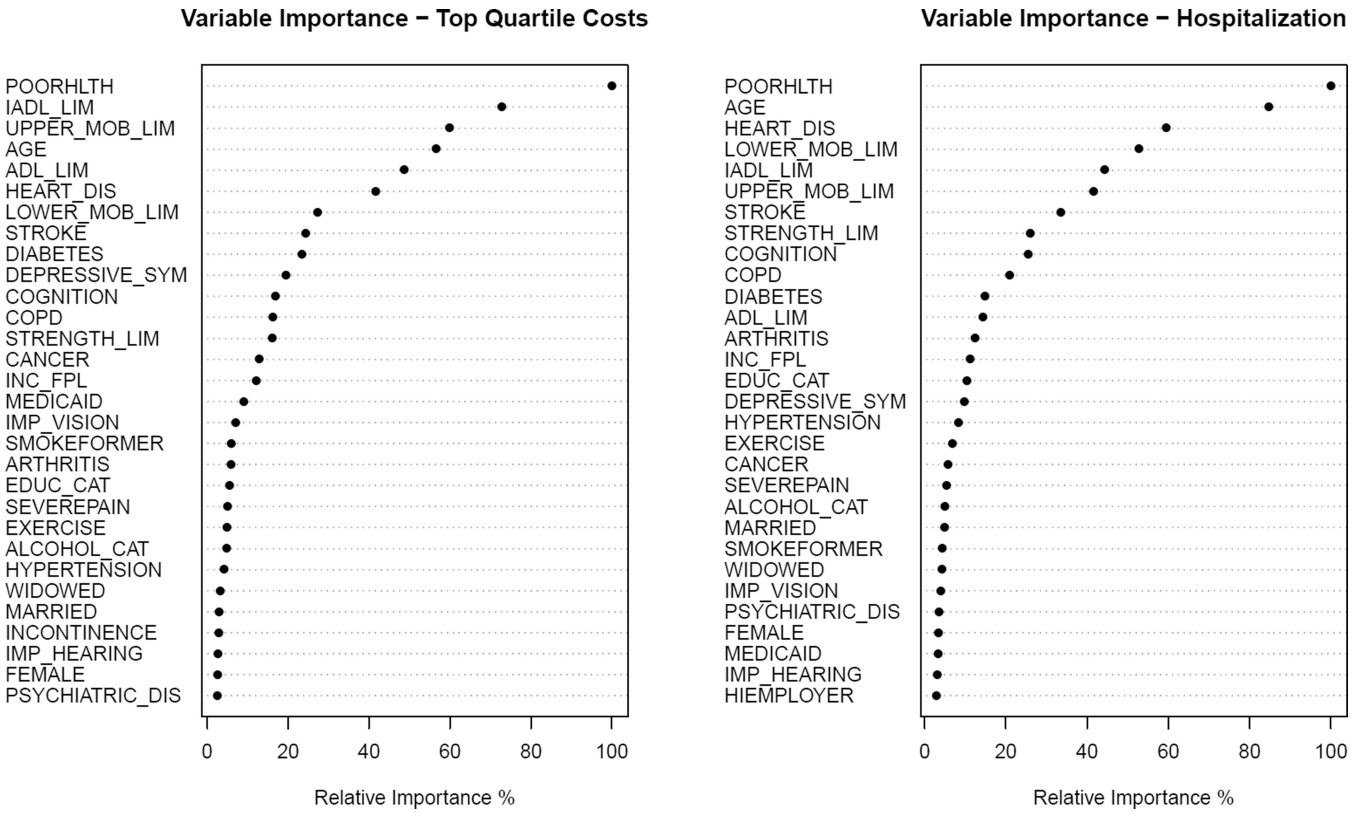

Figure 3. Random Forest Plots

The plots show the importance of each variable on explaining the outcomes for Medicare expenditures (left) and two-year hospitalization (right) using the Random Forest approach. Higher values indicate that a factor is an important factor associated with the outcome, while low values indicate the factor is not an important factor.

Table 1

Description of annual total Medicare expenditures in study population

Characteristic	N	Median	IQR	Percent in top quartile	Percent hospitalized
No. of subjects	5,771	$4,354	($1,438–$13,799)	25.0	40.7
Agecategories					
65–69	1,433	$2,593	($869–$8,998)	17.3	30.4
70–74	1,495	$3,563	($1,280–$10,404)	19.9	36.2
75–79	1,175	$4,772	($1,678–$14,547)	26.1	42.4
80–84	834	$5,650	($1,894–$16,319)	28.9	46.4
≥85	834	$10,565	($2,957–$26,182)	41.7	58.3
Sex					
Male	2,414	$4,245	($1,179–$14,101)	25.4	41.0
Female	3,357	$4,427	($1,594–$13,659)	24.7	40.5
Race/Ethnicity					
White, non-Hispanic	4,701	$4,192	($1,435–$13,094)	23.8	40.3
Black, non-Hispanic	635	$4,812	($1,392–$18,654)	29.5	42.7
Hispanic	345	$7,382	($1,758–$22,991)	35.9	42.0
Other	90	$3,633	($1,051–$9,931)	15.6	40.0
Marital status					
Married	3,317	$3,660	($1,294–$11,510)	21.8	37.1
Divorced	581	$5,865	($1,561–$17,901)	28.6	44.2
Widowed	1,731	$5,851	($1,735–$18,193)	30.3	46.6
Never married	142	$3,725	($1,205–$11,409)	21.1	38.0
Education, y					
<9	662	$7,379	($2,027–$22,991)	35.4	48.6
9-11	753	$5,497	($1,615–$17,823)	29.0	48.3
12	2,021	$4,150	($1,427–$13,621)	24.3	40.8
13–15	1,122	$3,578	($1,289–$11,275)	21.0	35.8
16	559	$4,001	($1,399–$12,446)	23.4	37.2
≥17	654	$3,427	($1,371–$10,849)	20.2	34.6
Income as % of federal poverty level					
<100%	486	$8,150	($1,973–$25,101)	38.9	49.0
100%–199%	1,147	$5,655	($1,634–$19,083)	30.5	48.2
200%–299%	1,069	$4,552	($1,548–$13,319)	24.1	40.5
≥300%	3,069	$3,563	($1,270–$11,437)	21.0	36.6

Characteristic	N	Median	IQR	Percent in top quartile	Percent hospitalized
Self-rated Poor Health					
No	3,918	$3,005	($1,117–$9,472)	17.7	33.7
Yes	1,853	$9,331	($3,240–$25,870)	40.4	55.3
Smoking status					
Never smoked	2,498	$3,867	($1,380–$11,598)	21.2	37.1
Former smoker	2,728	$4,918	($1,592–$15,935)	27.8	43.2
Current smoker	545	$4,164	($1,083–$15,954)	28.1	44.6
Alcohol use					
None	4,088	$4,792	($1,530–$15,523)	27.5	42.7
Moderate	1,416	$3,436	($1,312–$9,833)	18.4	35.2
Heavy	267	$3,860	($1,041–$11,457)	22.1	39.0
Body mass index					
Missing	51	$8,948	($3,026–$23,208)	39.2	43.1
Underweight	110	$11,725	($2,995–$27,274)	44.6	55.5
Normal/overweight	4,062	$4,146	($1,392–$13,311)	24.3	40.6
Obese	1,548	$4,519	($1,521–$13,796)	24.9	39.8
Vigorous exercise					
No	4,625	$4,849	($1,555–$15,537)	27.3	43.1
Yes	1,146	$3,002	($1,064–$8,478)	15.6	30.6
Proxy respondent					
Yes	259	$12,878	($3,395–$36,538)	24.0	39.8
No	5,512	$4,174	($1,414–$13,163)	46.7	59.9
Dual-Medicaid enrolled					
No	5,225	$4,064	($1,385–$12,757)	23.3	39.5
Yes	546	$9,570	($2,619–$26,112)	41.4	51.8

Table 2

Description of annual total Medicare expenditures by morbidity

Condition	N	Median	IQR	Percent in top quartile	Percent hospitalized
No. of subjects	5,771	$4,354	($1,438–$13,799)	25.0	40.7
Chronic conditions					
Hypertension, Mild	490	$5,041	($1,562–$18,300)	29.0	44.3
Hypertension, Severe	3,589	$5,031	($1,692–$15,185)	26.9	43.4
Arthritis, Mild	2,275	$4,070	($1,544–$11,857)	22.2	38.6
Arthritis, Severe	1,743	$6,507	($2,132–$20,061)	32.8	49.5
Heart Disease, Mild	1,729	$6,766	($2,322–$20,253)	32.9	49.5
Heart Disease, Severe	477	$10,213	($3,839–$28,337)	41.1	57.2
Lung Disease, Mild	544	$7,276	($2,347–$19,512)	34.0	49.5
Lung Disease, Severe	178	$16,528	($7,313–$37,347)	54.5	71.4
Stroke, Mild	653	$8,705	($3,118–$26,645)	40.7	57.1
Stroke, Severe	232	$11,283	($3,397–$27,364)	42.7	60.3
Cancer, Mild	1,199	$6,052	($1,975–$17,882)	30.4	44.0
Cancer, Severe	72	$19,792	($4,476–$46,179)	56.9	58.3
Diabetes, Mild	1,116	$5,647	($1,896–$17,892)	30.4	45.5
Diabetes, Severe	322	$13,541	($4,531–$41,302)	49.4	58.7
Psychiatric, Mild	322	$5,182	($1,412–$16,654)	28.9	41.9
Psychiatric, Severe	549	$6,686	($2,443–$17,823)	31.3	48.8
Functional Limitations					
Strength Limitations	3,748	$5,821	($1,966–$17,896)	30.1	46.2
Upper Body Mobility Limitations	2,340	$7,570	($2,457–$22,505)	36.3	50.8
Lower Body Mobility Limitations	4,067	$5,829	($1,939–$17,723)	29.9	46.3
Limitations in ADLs	352	$20,539	($7,224–$48,894)	59.1	66.2
Limitations in IADLs	1,026	$11,735	($3,565–$31,638)	46.1	58.5
Geriatric Syndromes					
Cognitive impairment, mild	150	$10,058	($3,840–$30,904)	41.3	64.0
Cognitive impairment, severe	212	$17,309	($6,520–$36,718)	54.7	67.5
Depressive Symptoms	720	$7,456	($2,267–$24,412)	38.1	49.7
Incontinence	1,711	$5,604	($1,928–$16,505)	29.4	44.8
Severe Pain	306	$10,690	($4,023–$28,133)	43.8	58.5
Visual Impairment	1,355	$7,189	($2,156–$21,605)	35.1	48.9
Hearing Impairment	1,547	$5,822	($1,758–$18,880)	30.8	46.2