UCSF
UC San Francisco Previously Published Works

Title

Identifying Latent Subgroups of High-Risk Patients Using Risk Score Trajectories

Permalink

https://escholarship.org/uc/item/6xs0z6f2

Journal

Journal of General Internal Medicine, 33(12)

ISSN

0884-8734

Authors

Wong, Edwin S
Yoon, Jean
Piegari, Rebecca I
et al.

Publication Date

2018-12-01

DOI

10.1007/s11606-018-4653-x

Peer reviewed

Identifying Latent Subgroups of High-Risk Patients Using Risk Score Trajectories

Edwin S. Wong, PhD[1,2], Jean Yoon, PhD[3,4], Rebecca I. Piegari, MS[5],
Ann-Marie M. Rosland, MD, MS[6,7], Stephan D. Fihn, MD, MPH[5,8], and Evelyn T. Chang, MD, MSHS[9,10]

[1]Center of Innovation for Veteran-Centered and Value-Driven Care, VA Puget Sound Health Care System, Seattle, WA, USA; [2]Department of Health Services, University of Washington, Seattle, WA, USA; [3]Health Economics Resource Center, VA Palo Alto Healthcare System, Livermore, CA, USA; [4]Department of General Internal Medicine, UCSF School of Medicine, San Francisco, CA, USA; [5]Office of Clinical Systems Development and Evaluation, Veterans Health Administration, Seattle, WA, USA; [6]Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Pittsburgh, PA, USA; [7]Department of Internal Medicine, University of Pittsburgh, Pittsburgh, PA, USA; [8]Department of Medicine, University of Washington, Seattle, WA, USA; [9]Center for the Study of Healthcare Innovation, Implementation and Policy, VA Greater Los Angeles Health Care System, Los Angeles, CA, USA; [10]David Geffen School of Medicine, University of California, Los Angeles, CA, USA.

OBJECTIVE: Many healthcare systems employ population-based risk scores to prospectively identify patients at high risk of poor outcomes, but it is unclear whether single point-in-time scores adequately represent future risk. We sought to identify and characterize latent subgroups of high-risk patients based on risk score trajectories.

STUDY DESIGN: Observational study of 7289 patients discharged from Veterans Health Administration (VA) hospitals during a 1-week period in November 2012 and categorized in the top 5th percentile of risk for hospitalization.

METHODS: Using VA administrative data, we calculated weekly risk scores using the validated Care Assessment Needs model, reflecting the predicted probability of hospitalization. We applied the non-parametric k-means algorithm to identify latent subgroups of patients based on the trajectory of patients' hospitalization probability over a 2-year period. We then compared baseline sociodemographic characteristics, comorbidities, health service use, and social instability markers between identified latent subgroups.

RESULTS: The best-fitting model identified two subgroups: moderately high and persistently high risk. The moderately high subgroup included 65% of patients and was characterized by moderate subgroup-level hospitalization probability decreasing from 0.22 to 0.10 between weeks 1 and 66, then remaining constant through the study end. The persistently high subgroup, comprising the remaining 35% of patients, had a subgroup-level probability increasing from 0.38 to 0.41 between weeks 1 and 52, and declining to 0.30 at study end. Persistently high-risk patients were older, had higher prevalence of social instability and comorbidities, and used more health services.

CONCLUSIONS: On average, one third of patients initially identified as high risk stayed at very high risk over a 2-year follow-up period, while risk for the other two thirds decreased to a moderately high level. This suggests that multiple approaches may be needed to address high-risk patient needs longitudinally or intermittently.

KEY WORDS: high risk; risk stratification; trajectory; latent subgroups; machine learning; patient-centered medical home.

INTRODUCTION

To improve patient outcomes and reduce costs, healthcare systems have implemented care management programs, often with mixed results.[1, 2] Typically, healthcare systems use risk stratification, or the process of categorizing patients by health risk status, to identify patients at high risk of adverse health outcomes. Then, they often apply the same intervention strategy to all patients identified as high risk. Previous research has found heterogeneity in the characteristics and clinical needs of high-risk patients,[3] in relation not only to clinical characteristics and use of health services, but also to the duration of being at high risk.[4] Little is known about matching clinical services to patients' patterns of risk to optimize their outcomes. Thus, it may be possible to identify longitudinal subgroups of high-risk patients who benefit from brief (e.g., transition in care or short-term services embedded within primary care) or long-term case (e.g., services embedded within primary care or dedicated primary care clinic for high-risk patients) management interventions.[1, 5, 6]

Previous research has not examined heterogeneity among high-risk patients using trajectories of risk or comorbidity score, or the potential for group-based trajectories that could inform clinical decision making. Current approaches to identifying and stratifying high-risk patients primarily rely on cross-sectional measurement of healthcare costs or comorbidity risk scores.[4] Leveraging trajectories of healthcare use, clinical

outcomes, and health markers may more accurately inform patients' clinical care needs. For example, patients at high risk of hospitalization over extended periods of time may have different care needs and outcomes than those who are at high risk for a short period. Group-based longitudinal modeling has already been applied to classify patients based on trajectories of healthcare spending, obesity, glycemic control, kidney function, disability levels, and medication adherence.[7–11]

Our objective was to identify and describe latent subgroups of high-risk patients with a recent hospitalization and receiving primary care from the Veterans Health Administration (VA). By latent subgroup, we refer to a subset of the high-risk population that is defined by patients sharing underlying unobserved characteristics. Subgroups are referred to as latent because they would not be easily observed in data with other analysis methods.[12–14] Our approach is designed to be a data-driven analysis using established methods in machine learning rather than hypothesis testing. We describe discrete patterns of risk score trajectories based on weekly measures that VA generates to help primary care providers identify and improve care for patients at high risk for hospitalizations. Based on prior research findings demonstrating variation in patterns of utilization and cost, we expected multiple distinct group-level risk score trajectories that differed by direction and magnitude of change over time.[4, 15]

METHODS

Data Sources

The primary source of data was the VA Corporate Data Warehouse (CDW), which contains comprehensive administrative data tracking patient risk scores, health service utilization, provider panel assignments, demographics, and clinical measures including ICD-9 diagnostic codes. VA data were linked to Medicare claims data, which were used to measure utilization of outpatient care through Fee-For-Service (FFS) Medicare.

Study Population

We identified 12,930 patients who had a medical/surgical hospitalization at one of 125 VA hospitals and discharged during a 1-week period in November 2012. We excluded 2243 patients who did not have a risk score in the week following discharge, which included those that were not assigned to a VA primary care provider. We also excluded 321 patients who had a hospitalization lasting more than 30 days during the follow-up period. Finally, we excluded 3077 patients who had a risk score following the hospitalization that was not in the top 5th percentile. In total, the study population consisted of 7289 patients.

Trajectories of Patient Risk Scores

Weekly risk scores were derived from the previously validated Care Assessment Needs (CAN) model.[16] The CAN model assesses patient risk by estimating the probability of

hospitalization and death, respectively, as a function of six sets of patient characteristics: demographics (including limited data elements from the Department of Defense), clinical diagnoses, vital signs, utilization, pharmacy, and laboratory values. We conducted separate analyses of predicted probabilities of hospitalization over the subsequent 90 days and 1 year. Using weekly hospitalization probabilities from the CAN model, we constructed patient-level trajectories over a 104-week (2 year) period. The predictive accuracies (c-statistic) of 90-day and 1-year hospitalization CAN scores were 0.83 and 0.87, respectively.[16]

Statistical Analysis

We used the non-parametric k-means clustering algorithm to identify latent subgroups of patients based on trajectories of hospitalization probability.[17, 18] This machine learning approach determined the following: (1) the optimal number of subgroups and (2) a mean trajectory for each subgroup, given the optimal number of subgroups. The k-means algorithm identified the best fitting model as the one which maximizes the distance between subgroup trajectories while minimizing variation in patient-level trajectories within a subgroup. The k-means algorithm iteratively partitions patient trajectories into a minimum of two and a maximum of six clusters. The number of subgroups was determined based on the Calinski-Harabatz criterion that maximizes the ratio of the between-subgroup variance and the within-subgroup variance. Hospitalization probabilities from the CAN model were generated at the start of each week for all patients who were alive, listed on a primary care panel, and not hospitalized at the end of the prior week. Estimation of the k-means algorithm applied the copy mean assumption, which imputes missing values by filling the subgroup-level mean for a given week.

Characterizing Latent Subgroups

We compared latent subgroups in relation to patient characteristics measured in fiscal year 2012, 12 months prior to the baseline date: demographics, comorbidities, markers of social instability, health service utilization, and socioeconomic status of patients' residence area. Demographics included age, gender, marital status, race/ethnicity, eligibility for copayment-exempt VA care, distance from patients' residence to the nearest VA clinic, and residence in a rural geographical area. Comorbidities included indicator variables for the presence of conditions in the validated Gagne (2012) index.[19] Also, because of the high prevalence of mental illness in VA and the limited number of mental health comorbidities available in the Gagne index, we added several mental health condition indicators from a prior VA study.[20] Markers of social instability included number of residence ZIP codes and history of hospital discharge against medical advice (AMA), which were strongly associated with 30-day readmissions in prior research.[21] We also examined homelessness, which was defined using a previously validated algorithm of VA clinic identifier

codes and ICD-9 diagnosis codes.[22] Utilization included counts of the number of VA hospitalizations, emergency department (ED) visits, office-based visits to primary care and specialty care, and palliative care encounters. Outpatient visit counts included care received from VA and FFS Medicare, and were derived using a previously developed algorithm that classified visits based on provider specialty and Current Procedure Terminology codes.[23] Socioeconomic status in patients' residence areas was captured using a composite Census tract level measure.[24] We calculated standardized mean differences (SMDs) for each variable, and considered SMDs with a magnitude of greater than 0.1 as meaningful, consistent with prior studies.[25, 26]

We conducted two additional analyses to characterize latent subgroups identified by the k-means algorithm. First, we estimated a logistic regression model including characteristics described above as covariates and calculated the pseudo R-squared to assess the proportion of variation in subgroup assignment explained by observable characteristics. Second, we compared variation in patient-level trajectories around mean subgroup-level trajectories by comparing intraclass correlations (ICCs) between latent subgroups. We calculated ICCs as the proportion of variation in hospitalization probabilities explained at the patient level.

Sensitivity Analysis

We performed several sensitivity analyses. First, because mortality occurred in 17% of the study population, subgroup-level trajectories could have been influenced by missing values related to inherent censoring of relatively sicker patients who died during the study. To examine this possibility, we estimated k-means clustering models for the subpopulation of patients that were alive at study end. Second, hospitalization probabilities from the CAN model were generated exclusively from VA data and may have underestimated risk for patients who also obtained care from other non-VA providers.[23, 27, 28] We explored potential differences among these patients by estimating clustering models among the subgroup of patients who had > 50% reliance on VA outpatient care. To measure reliance, we identified all face-to-face office visits with primary care and specialty care providers in VA and FFS Medicare over the 104-week follow-up period. We defined reliance as the proportion of all visits obtained from VA. Third, we estimated trajectory models excluding patients assigned to the VA home-based primary care program during the follow-up period to assess whether group-level trajectories were influenced by this intervention for patients with complex health needs.[29] Fourth, the distribution of hospitalization probabilities generated by the CAN model exhibited right skewness and subgroup-level trajectories may be influenced by outliers (Online Appendix, Fig. 1). To address this possibility, we identified latent subgroups using k-medians clustering, which calculated the median trajectory for patients assigned to a given subgroup. All statistical analyses were performed

using RStudio version 0.99.902 (RStudio, Inc., Boston, MA) and STATA version 13.0 (STATA Corp., College Station, TX).

RESULTS

Latent Subgroups

For 90-day hospitalization probability, the best fitting model consisted of two latent subgroups (Fig. 1). The first subgroup, characterized by moderately high risk, included 65% of patients. Subgroup-level hospitalization probability decreased from 0.224 in week 1 to 0.097 at week 66 and remained around 0.096 between weeks 67 through week 104. The second subgroup, characterized by persistently high risk, comprised the remaining 35% of patients. The subgroup-level trajectory for persistently high-risk patients exhibited a modest increase in hospitalization probability between weeks 1 and 12. Between weeks 13 and 52, the subgroup-level probability was steady around 0.38. Then, the subgroup-level probability decreased from 0.344 in week 53 to 0.296 by week 104. For both subgroups, marked probability decreases occurred after weeks 12 and 52. These stepwise declines reflect the expiration of the increase in risk score stemming from the baseline hospitalization at weeks 12 and 52 of follow-up, respectively,
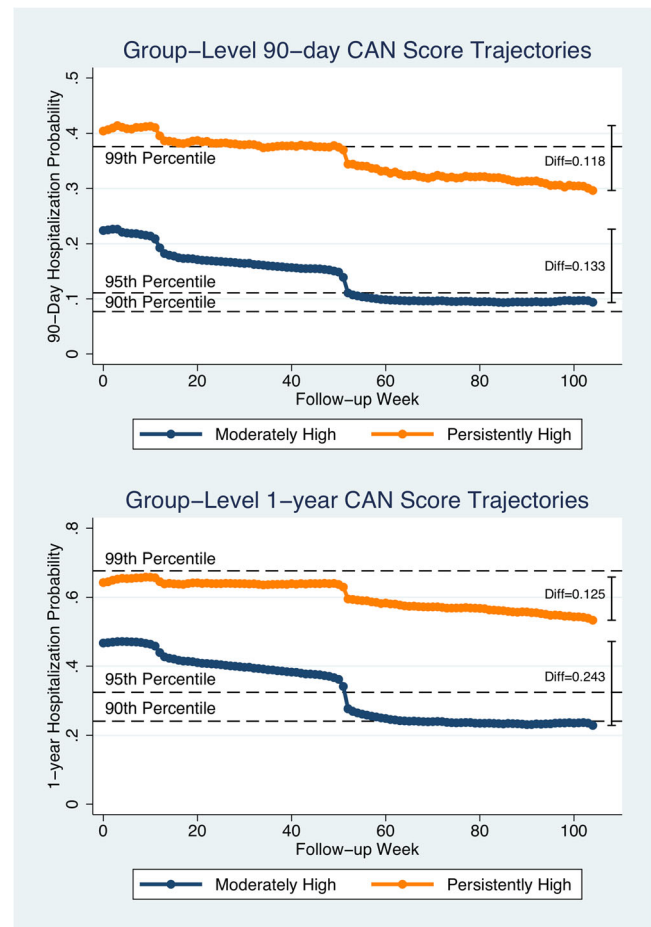

Fig. 1 Subgroup-level trajectories based on 90-day and 1-year hospitalization probability.

in this sample of recently hospitalized patients. Analysis of 1-year hospitalization probabilities also produced two latent subgroups exhibiting similar group-level changes over time.

Sensitivity analyses examining the respective subpopulations of patients alive at end of follow-up, mostly reliant on VA and not enrolled in home-based primary care produced similar results (Online Appendix, Figs. 2–3). Analysis using k-medians to address skewness in the distribution of hospitalization probabilities also identified two latent subgroups exhibiting similar subgroup-level trajectories; however, a greater proportion of patients were assigned to the persistently high-risk subgroup (49%) (Online Appendix, Fig. 4).

Characteristics of Latent Subgroups

Table 1 presents baseline characteristics of the two latent subgroups of patients based on trajectories of 90-day hospitalization probabilities. Mean hospitalization probability at baseline was 0.22 (range = 0.10–0.69) for the moderately high-risk subgroup and 0.40 (range = 0.10–0.80) for the persistently high-risk subgroup (Fig. 2). For demographics, persistently high-risk patients were older (66.9 vs. 65.0 years, SMD = − 0.15) and less likely to be married (34% vs. 42%, SMD = 0.15). Patients in the persistently high-risk subgroup had higher rates of nearly all comorbidities. The largest differences were for heart failure (36.4% vs. 13.9%, SMD = 0.54), fluid and electrolyte disorders (36.8% vs. 15.7%, SMD = 0.49), controlled hypertension (25.0% vs. 8.2%, SMD = 0.46), chronic obstructive pulmonary disease (47.4% vs. 26.9%, SMD = 0.43), and renal failure (31.3% vs. 14.2%, SMD = 0.42). Persistently high-risk patients also had markedly higher rates of mental comorbidities with the largest differences for alcohol abuse (29.9% vs. 18.4%, SMD = 0.27) and substance use disorder (34.0% vs. 23.3%, SMD = 0.24). Two markers of social instability [unique residence ZIP codes (1.3 vs. 1.2, SMD = 0.17) and history of discharge AMA (4.8% vs. 2.8%, SMD = 0.11)] were greater among persistently high-risk patients. Finally, rates of all types of health service utilization, including inpatient and outpatient care, were markedly higher among persistently high-risk patients. The most substantial differences were for number of hospitalizations (2.2 vs. 0.8, SMD = 0.69) and ED visits (3.9 vs. 1.8, SMD = 0.58).

The pseudo R-squared from the logistic regression estimating subgroup assignment as a function of baseline patient characteristics was 0.23. ICC estimates measuring proportion of variation in trajectories explained at the patient level was 0.42 and 0.36 for the trajectories of persistently high-risk and moderately high-risk patients, respectively.

DISCUSSION

We sought to identify patterns in the longitudinal profile of risk scores among patients who were receiving care from an integrated healthcare delivery system and considered high risk. Using patient-level trajectories of risk scores over a 2-year period, we identified two latent subgroups of patients. In both subgroups, mean trajectories declined over the follow-up period, likely due in part to mean reversion. One subgroup of patients exhibited lower mean hospitalization probabilities and a more pronounced decline in risk over time, and another subgroup of patients exhibited higher mean hospitalization probabilities that remained persistently high over time. Nevertheless, patients in both subgroups, on average, remained in the top 10th percentile of risk throughout the 2-year follow-up period. However, there was substantial heterogeneity in patient-level trajectories within both latent subgroups.

We identified marked differences in characteristics at baseline between latent subgroups. Although the mean hospitalization probability at baseline was higher for the persistently high-risk subgroup, the range of initial hospitalization probabilities exhibited substantial overlap, indicating subgroup assignment was not completely determined by initial probability for hospitalization alone. The largest differences in patient characteristics between subgroups were in measures of utilization in the prior year, particularly for hospitalizations and ED visits. We identified marked differences in rates of heart failure and chronic obstructive pulmonary disease, which are two of the leading causes of hospitalization in VA.[30, 31] Notably, we found persistently high-risk patients had a greater prevalence of markers of social instability including history of discharge AMA and number of residence ZIP codes. Although several differences between subgroups were identified, we found observable baseline characteristics explained 23% of the variation in subgroup assignment. This corresponds with prior studies estimating 20–30% of variation in health service use and mortality explained by similar patient characteristics.[32, 33]

To our knowledge, this is the first study using latent class modeling to examine heterogeneity in the subpopulation of high-risk patients by identifying patterns in the longitudinal profile of risk scores. Latent class models are increasingly being applied to characterize heterogeneous groups of patients by costs and utilization to better inform more efficient delivery of clinical care.[4, 10] Others are already exploring the use of latent models to describe various clusters by disease groupings. It may be challenging, however, for primary care to use these groupings[34] because many high-risk patients have multiple chronic medical conditions; creating disease management programs for each disease group may result in greater care fragmentation.[35] Our study differs from prior work by examining trajectories of a composite measure of risk, and focusing on the subpopulation of patients at highest risk, excluding variation introduced by lower risk individuals. A future direction is the examination of risk score trajectory patterns among disease-specific populations, which may augment findings in this study.

Findings from this study contribute to efforts by healthcare systems to develop population-level tools to identify and manage high-risk patients, as espoused by the National Committee for Quality Assurance[36–38] for patient-centered medical home model certification. We adopted a population-level

Table 1 Selected Descriptive Statistics by Latent Subgroup

	Moderately high risk N = 4647	Persistently high risk N = 2547	Standardized difference
Age (mean/sd)	65.0 (13.5)	66.9 (12.2)	− 0.15
Male (%)	94.3	96.2	− 0.09
Race			
White (%)	62.6	63.7	− 0.02
Black (%)	17.8	21.2	− 0.09
Marital status			
Married (%)	41.6	34.1	0.15
Divorced (%)	31.1	34.2	− 0.07
Widowed (%)	10.7	13.5	− 0.09
Not married (%)	12.1	12.5	− 0.01
Urban (%)	67.2	71.4	− 0.09
Drive distance to nearest VA			
0–4.9 miles (%)	22.9	25.8	− 0.07
5–9.9 miles (%)	26.0	28.0	− 0.04
10–19.9 miles (%)	24.1	24.8	− 0.02
20–39.9 miles (%)	20.4	15.7	0.12
> 40 miles (%)	6.6	5.8	0.04
Branch of service			
Air Force (%)	12.7	12.2	0.01
Army (%)	55.4	55.1	0.01
Marine Corps (%)	11.1	11.7	− 0.02
Navy (%)	19.7	20.2	− 0.01
Other (%)	1.1	0.8	0.03
Medical comorbidities			
Heart failure (%)	13.9	36.4	− 0.54
Fluid and electrolyte disorders (%)	15.7	36.8	− 0.49
Hypertension—controlled (%)	8.2	25.0	− 0.46
COPD (%)	26.9	47.4	− 0.43
Renal failure (%)	14.2	31.3	− 0.42
Arrhythmia (%)	20.3	35.7	− 0.35
Liver disease (%)	7.3	18.0	− 0.32
Diabetes—complicated (%)	13.8	25.9	− 0.31
Diabetes—uncomplicated (%)	36.5	50.2	− 0.28
Hypertension—uncontrolled (%)	71.1	81.5	− 0.25
Deficiency anemia (%)	6.0	13.0	− 0.24
Coagulopathy (%)	4.3	10.1	− 0.23
Pulmonary circulatory disorder (%)	3.6	8.8	− 0.22
Peripheral vascular disorder (%)	12.4	19.8	− 0.20
Myocardial infarction (%)	4.8	9.9	− 0.20
Mental comorbidities			
Alcohol abuse (%)	18.4	29.9	− 0.27
Substance use disorder (%)	23.2	34.0	− 0.24
Depression (%)	40.2	49.8	− 0.19
Psychoses (%)	10.1	14.8	− 0.14
Serious mental illness (%)	12.1	17.1	− 0.14
Bipolar disorder (%)	5.6	8.1	− 0.10
Social instability			
Homelessness (%)	7.1	9.7	− 0.10
Discharge AMA (%)	2.8	4.8	− 0.11
Unique ZIP codes (mean/sd)	1.2 (0.6)	1.3 (0.8)	− 0.17
Utilization			
Any hospitalization (%)	43.1	74.6	− 0.68
No. of hospitalizations (mean/sd)	0.8 (1.4)	2.2 (2.5)	− 0.69
ED visits (mean/sd)	1.8 (2.8)	3.9 (4.4)	− 0.58
Total outpatient visits (mean/sd)	7.7 (11)	14.7 (63)	− 0.16
Palliative use (%)	1.2	5.6	− 0.24

COPD chronic obstructive pulmonary disease, AMA against medical advice, ED emergency department

approach in the identification of latent subgroups to help inform clinic- and system-level planning. This may help stakeholders and practitioners who are responsible for the management of populations by identifying patients' expected risk score trajectory, which may be used to inform assignment to specialized programs for managing high-risk patients. For example, study findings can help practitioners determine assignment level within the context of the stepped care model that adjusts the level of care to patients' needs.[39]

The ability to categorize patients' trajectories of risk has potential implications for further research examining the effectiveness of intensive management interventions for patients at high risk of adverse outcomes, where most randomized trials have not shown significant benefits. For example, VA investigators conducted two randomized evaluations of an intensive, multidisciplinary, team-based approach for veterans with high hospitalization probability, as predicted by the CAN model, and showed no reduction in use of acute care or total healthcare
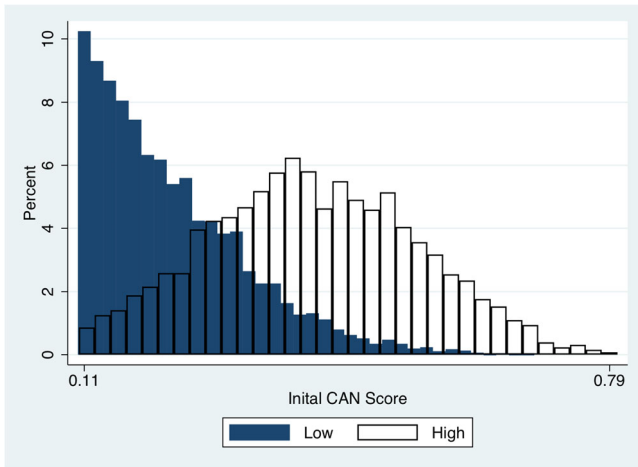
Fig. 2 Distribution of initial hospitalization probabilities by latent subgroup.

costs among intervention patients relative to patients receiving standard VA care.[40, 41] Future research should ascertain whether these programs exhibit variable effects on patients with different expected risk trajectories, characterized using baseline characteristics identified in this study.

This study has several limitations. First, we examined the population of VA enrollees using a measure of patient risk currently used by VA to inform clinical decision making. The VA population differs from populations served by other health systems in characteristics such as gender and race/ethnicity. Further, risk scores from the CAN model were computed using VA and veteran-specific variables including branch of service and level of service-connected disability. Thus, the number of latent subgroups of high-risk patients and characteristics associated with subgroup assignment may differ in other non-VA settings or for different measures of risk. Nevertheless, in prior population-level studies, a high degree of comparability was identified between VA enrollees and Medicare beneficiaries[42] and with patients in other super-utilizer programs.[43–45] Second, hospitalization probabilities from the CAN model were estimated using data capturing utilization of VA healthcare, so risk may be underestimated for patients who also used health services from non-VA providers. However, results from subgroup analysis among patients who were mostly reliant upon VA outpatient care were similar to the primary results. Finally, the k-means clustering approach does not account for potential informative missingness arising from death during follow-up in the determination of subgroup-level trajectories. However, the analysis of patients who survived through study end validated our primary results, mitigating these concerns.

CONCLUSION

In this study, we applied a machine learning approach to identify two latent subgroups of high-risk patients based on longitudinal trajectories of risk scores over a 2-year period.

Subgroup-level trajectories for both subgroups declined somewhat over time; however, risk for one subgroup remained persistently high, while the other subgroup stabilized more quickly and at a lower level. Study findings help address current limitations in understanding of patterns of high risk over time and offer suggestions of how this knowledge can be adopted in clinical practice. Future research should examine whether the effects of clinical interventions targeting high-risk patients differ across specific subgroups of patients defined by risk score trajectories, such as those identified in this study.

Corresponding Author: Edwin S. Wong, PhD; Center of Innovation for Veteran-Centered and Value-Driven CareVA Puget Sound Health Care System, 1660 S. Columbian Way, HSR&D MS S-152, Seattle, WA 98108, USA (e-mail: eswong@uw.edu).

Compliance with Ethical Standards:

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs, the University of Washington, the University of California, and the University of Pittsburgh.

Conflict of Interest: Dr. Wong reports ownership of common stock in UnitedHealth Group Inc. All other authors have no conflicts of interest to report.

REFERENCES

1. Edwards ST, Peterson K, Chan B, Anderson J, Helfand M. Effectiveness of Intensive Primary Care Interventions: A Systematic Review. J Gen Intern Med. 2017;32(12):1377–1386.
2. Peterson K, Helfand M, Humphrey L, Christensen V, Carson S. Evidence Brief: Effectiveness of Intensive Primary Care Programs. https://www.hsrd.research.va.gov/publications/esp/intensive-primary-care.pdf. Accessed Nov 1, 2017.
3. Powers BW, Chaguturu SK. ACOs and High-Cost Patients. N Engl J Med. 2016;374(3):203–205.
4. Lee NS, Whitman N, Vakharia N, Ph DG, Rothberg MB. High-Cost Patients: Hot-Spotters Don't Explain the Half of It. J Gen Intern Med. 2017;32(1):28–34.
5. Hochman M, Asch SM. Disruptive Models in Primary Care: Caring for High-Needs, High-Cost Populations. J Gen Intern Med. 2017;32(4):392–397.
6. Blumenthal D, Chernof B, Fulmer T, Lumpkin J, Selberg J. Caring for High-Need, High-Cost Patients - An Urgent Priority. N Engl J Med. 2016;375(10):909–911.
7. Gill TM, Gahbauer EA, Han L, Allore HG. Trajectories of disability in the last year of life. N Engl J Med. 2010;362(13):1173–1180.
8. Franklin JM, Shrank WH, Pakes J, et al. Group-based trajectory models: a new approach to classifying and predicting long-term medication adherence. Med Care. 2013;51(9):789–796.

9.   Ravona-Springer R, Heymann A, Schmeidler J, et al. Trajectories in glycemic control over time are associated with cognitive performance in elderly subjects with type 2 diabetes. PLoS One. 2014;9(6):e97384.

10. Lauffenburger JC, Franklin JM, Krumme AA, et al. Longitudinal Patterns of Spending Enhance the Ability to Predict Costly Patients: A Novel Approach to Identify Patients for Cost Containment. Med Care. 2017;55(1):64–73.

11. O'Hare AM, Batten A, Burrows NR, et al. Trajectories of kidney function decline in the 2 years before initiation of long-term dialysis. Am J Kidney Dis. 2012;59(4):513–522.

12. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol. 2010;6:109–138.

13. Lanza ST, Rhoades BL. Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. Prev Sci. 2013;14(2):157–168.

14. Institute of Medicine. *Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1*. Washington, DC 2014.

15. Yoon J, Chee CP, Su P, et al. Persistence of High Health Care Costs among VA Patients. Health Serv Res. 2018.

16. Wang L, Porter B, Maynard C, et al. Predicting risk of hospitalization or death among patients receiving primary care in the Veterans Health Administration. Med Care. 2013;51(4):368–373.

17. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics). 1979;28(1):100–108.

18. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R Packages to Cluster Longitudinal Data. *Journal of Statistical Software.* 2016;65(4).

19. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. J Clin Epidemiol. 2011;64(7):749–759.

20. Sloan KL, Montez-Rath ME, Spiro A, 3rd, et al. Development and validation of a psychiatric case-mix system. Med Care. 2006;44(6):568–580.

21. Amarasingham R, Moore BJ, Tabak YP, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Med Care. 2010;48(11):981–988.

22. Peterson R, Gundlapalli AV, Metraux S, et al. Identifying Homelessness among Veterans Using VA Administrative Data: Opportunities to Expand Detection Criteria. PLoS One. 2015;10(7):e0132664.

23. Liu CF, Manning WG, Burgess JF, Jr., et al. Reliance on Veterans Affairs outpatient care by Medicare-eligible veterans. Med Care. 2011;49(10):911–917.

24. Nelson K, Schwartz G, Hernandez S, et al. The Association Between Neighborhood Environment and Mortality: Results from a National Study of Veterans. J Gen Intern Med. 2017;32(4):416–422.

25. Normand ST, Landrum MB, Guadagnoli E, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol. 2001;54(4):387–398.

26. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399–424.

27. Hynes DM, Koelling K, Stroupe K, et al. Veterans' access to and use of Medicare and Veterans Affairs health care. Med Care. 2007;45(3):214–223.

28. Petersen LA, Byrne MM, Daw CN, et al. Relationship between clinical conditions and use of Veterans Affairs health care among Medicare-enrolled veterans. Health Serv Res. 2010;45(3):762–791.

29. Department of Veterans Affairs. Home Based Primary Care. https://www.va.gov/geriatrics/docs/Home_Based_Primary_Care.pdf. Accessed Nov 1, 2017.

30. U.S. Department of Veterans Affairs. Improving the Health and Care for Veterans with Cardiovascular Disease. https://www.hsrd.research.va.gov/publications/emerging_evidence/cardiovascular-disease.cfm. Accessed Nov 1, 2017.

31. Rinne ST, Elwy AR, Bastian LA, et al. Impact of Multisystem Health Care on Readmission and Follow-up Among Veterans Hospitalized for Chronic Obstructive Pulmonary Disease. Med Care. 2017;55 Suppl 7 Suppl 1:S20-S25.

32. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998;36(1):8–27.

33. Fortin Y, Crispo JA, Cohen D, et al. External validation and comparison of two variants of the Elixhauser comorbidity measures for all-cause mortality. PLoS One. 2017;12(3):e0174379.

34. Prenovost K, Maciejewski ML, Vijan S, et al. Mixture Distribution Item Response Theory As a Method for Identifying Clinically Relevant Subgroups of Patients https://academyhealth.confex.com/academyhealth/2016arm/mediafile/Handout/Paper9891/Prenovost_AHposter_1methods_FINAL_061716.pdf. Accessed Jun 12, 2018.

35. Porter ME, Pabo EA, Lee TH. Redesigning primary care: a strategic vision to improve value by organizing around patients' needs. Health Aff (Millwood). 2013;32(3):516–525.

36. Scholle SH, Saunders RC, Tirodkar MA, Torda P, Pawlson LG. Patient-centered medical homes in the United States. J Ambul Care Manage. 2011;34(1):20–32.

37. National Committee for Quality Assurance. PCMH Recognition. https://www.ncqa.org/Portals/0/PCMH%20brochure-web.pdf. Accessed Aug 1, 2017.

38. Sinaiko AD, Landrum MB, Meyers DJ, et al. Synthesis Of Research On Patient-Centered Medical Homes Brings Systematic Differences Into Relief. Health Aff (Millwood). 2017;36(3):500–508.

39. Ewigman NL, Vella L, Eng JA. Implementing evidence-based interventions to prevent readmissions in the real world. *J Gen Intern Med.* 2018;33(5):578–580.

40. Yoon J, Chang E, Rubenstein LV, et al. Impact of Primary Care Intensive Management on High-Risk Veterans' Costs and Utilization: A Randomized Quality Improvement Trial. Ann Intern Med. 2018;168(12):846–854.

41. Zulman DM, Pal Chee C, Ezeji-Okoye SC, et al. Effect of an Intensive Outpatient Program to Augment Primary Care for High-Need Veterans Affairs Patients: A Randomized Clinical Trial. JAMA Intern Med. 2017;177(2):166–175.

42. Wong ES, Wang V, Liu CF, Hebert PL, Maciejewski ML. Do Veterans Health Administration Enrollees Generalize to Other Populations? Med Care Res Rev. 2016;73(4):493–507.

43. Johnson TL, Rinehart DJ, Durfee J, et al. For many patients who use large amounts of health care services, the need is intense yet temporary. Health Aff (Millwood). 2015;34(8):1312–1319.

44. Bodenheimer T. Strategies to Reduce Costs and Improve Care for High-Utilizing Medicaid Patients: Reflections on Pioneering Programs: Center for Health Care Strategies; 2013.

45. Hasselman D. *Super-Utilizer Summit: Common Themes from Innovative Complex Care Management Programs*: Center for Health Care Strategies;2013.