CENTER for RETIREMENT RESEARCH at BOSTON COLLEGE

A STUDY OF LONGITUDINAL TRAJECTORIES OF HEALTH AND JOB DEMAND ON RETIREMENT AGE

Amal Harrati and David Rehkopf

CRR WP 2020-1
January 2020

Center for Retirement Research at Boston College
Hovey House
140 Commonwealth Avenue
Chestnut Hill, MA 02467
Tel: 617-552-1762 Fax: 617-552-0191
https://crr.bc.edu

Amal Harrati is an instructor of medicine at the Stanford University School of Medicine. David Rehkopf is an associate professor of medicine at the Stanford University School of Medicine.
The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA, any agency of the federal government, the Stanford University School of Medicine, or Boston College. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.

© 2020, Amal Harrati and David Rehkopf. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

About the Steven H. Sandell Grant Program

This paper received funding from the Steven H. Sandell Grant Program for Junior Scholars in Retirement Research. Established in 1999, the Sandell program's purpose is to promote research on retirement issues by scholars in a wide variety of disciplines, including actuarial science, demography, economics, finance, gerontology, political science, psychology, public administration, public policy, sociology, social work, and statistics. The program is funded through a grant from the Social Security Administration (SSA). For more information on the Sandell program, please visit our website at: http://crr.bc.edu/?p=9570, send e-mail to crr@bc.edu, or call (617) 552-1762.

About the Center for Retirement Research

The Center for Retirement Research at Boston College, part of a consortium that includes parallel centers at the National Bureau of Economic Research, the University of Michigan, and the University of Wisconsin-Madison, was established in 1998 through a grant from the U.S. Social Security Administration. The Center's mission is to produce first-class research and forge a strong link between the academic community and decision-makers in the public and private sectors around an issue of critical importance to the nation's future. To achieve this mission, the Center sponsors a wide variety of research projects, transmits new findings to a broad audience, trains new scholars, and broadens access to valuable data sources.

Center for Retirement Research at Boston College
Hovey House
140 Commonwealth Ave
Chestnut Hill, MA 02467
Tel: 617-552-1762 Fax: 617-552-0191
https://crr.bc.edu

Affiliated Institutions:
The Brookings Institution
Mathematica – Center for Studying Disability Policy
Syracuse University
Urban Institute

Abstract

In this paper, we characterized health trajectories over an 18-year period for a cohort of American workers. We used administrative data to track monthly, health-related events for six chronic diseases (asthma, arthritis, diabetes, depression, ischemic heart disease, and hypertension) including the diagnoses of new disease, hospitalizations, and outpatient visits. Using these data, we first used sequence and cluster analysis to characterize long-term trajectories of health and to group workers according to their patterns of work experience. We then modeled the relationship of these health trajectories to retirement age, accounting for baseline underlying health, as well as a number of demographic and job-related characteristics. Finally, we consider the role of physical and psychosocial job demand in retirement age.

Our analysis produces a number of findings that should be of interest to those studying retirement policy. In our data, workers can be categorized into a small number of distinct work trajectories. While the majority of workers in this sample remain relatively healthy for much of the observation window, others exhibit patterns of health-related events that are often marked by hypertension and/or arthritis. We find that clusters characterized with health events related to hypertension and arthritis are likely to retire later, not earlier. We offer a possible interpretation that these numerous health-related events are signaling the proper management of these chronic diseases, allowing workers to extend their working life. Moreover, we find an association with job demands and retirement, even after controlling for health. Specifically, we find that increases in exposure to heat are associated with lower retirement age, as is less decision-making autonomy. The limitations of this paper include the lack of a representative sample, a relatively small sample size, and the strong incentives of retirement pensions in this cohort that may overwhelm other factors related to retirement decisions.

The paper found that:

- There are six distinct clusters of health trajectories, which are largely characterized by diagnosis and management of hypertension and arthritis.
- Workers with health trajectories that involve more health-related events, and in particular hypertension, were actually more likely to retire later relative to their healthier peers. This may reflect better management of hypertension through more frequent outpatient physician visits, or a greater need for health insurance coverage.

- Increases in one physical job demand—heat —and decreases in one psychosocial job demand —decision-making autonomy—were associated with lower retirement age.

The policy implications of the findings are:

- Interventions to decrease the prevalence of, or help in the management of, hypertension and arthritis may help keep workers healthier while working.
- Retirement pensions and health insurance provide strong incentives for employees to work longer.
- Lessening demanding job conditions, and in particular exposure to heat, could allow workers to retire later. Likewise, creating opportunities in the workplace for workers to have greater decision-making autonomy may also incentivize later retirement.

Background

The transition from work to retirement is a complex process that can follow various pathways and evolve from multiple influences. Health is often considered one of the most important factors in the transition to retirement (Munnell, Sanzenbacher, & Webb, 2015). As such, there a is long literature exploring the role of health in retirement (Duggan, Singleton, & Song, 2007; Dwyer & Mitchell, 1999; McGarry, 2004; Munnell, Soto, & Golub-Sass, 2008). Indeed, research suggests that the growing gap in retirement ages may be a function of growing inequality in health and longevity (Rutledge, 2018). However, most work studying the role of health on retirement is based on short time frames, with little follow up and/or long lags of time between follow ups, or centered on the time period immediately preceding or following the retirement decision.

However, the state of an individual's health state at retirement is the result of a cumulative process of deteriorating health across the entire span of adulthood. While it has been assumed in both the scientific and policy-oriented literatures that life course experiences are important factors for understanding retirement, the relevance of mid-life experiences have often remained implicit or been neglected in empirical studies. The limited evidence focused on the relationship of mid-life health and retirement often focuses on specific health shocks and not longer-term health trajectories. Little is known about the cumulative patterns of health during active employment on retirement patterns.

Scholarly efforts in this domain have been stymied, because opportunities to track longitudinal data through long periods of working histories are extremely rare. We address this gap in knowledge through use of a unique database of nearly 5,000 retirees who were previous employees at a large national firm, with detailed, longitudinal health claims data on their tenures at the firm.

We seek to answer two questions: First, can early retirement of workers be explained by midlife health trajectories? Second, given that previous work with these data suggests that workers whose jobs have high physical job demand retire earlier (Modrek & Cullen, 2012) we also ask: How is that relationship influenced by job demand? We explore the relationship between these health trajectories and job demand, and ask whether job demand mediates the health-retirement association. We explore patterns in job demand over employment and job

characteristics, how those patterns are related to past or subsequent changes in health, and how these effects cumulate jointly to inform retirement decisions.

This project seeks to answer the following questions:

1. How are trajectories of health across mid-life related to retirement age?
2. How is the relationship between long-term health trajectories and retirement decisions mediated by job characteristics and job demand?

Data

This study employs a set of administrative data, the *American Manufacturing Cohort* (AMC), that track the employment, health, and job characteristics of a large cohort of workers at a major American manufacturing firm between 1996-2013. These data are both dynamic—capturing changes to employees' job characteristics, health, and employment status as they occur—and long term, following individuals so long as they are employed at the firm within the observation period. We draw on multiple sources of administrative data from AMC. The two central components of the data are: i) the human resources (HR) records that detail all changes in work status (i.e., hiring, firing, retirement, entering or returning from leave status, promotion, etc.) for all employees, and ii) health claims data, which record all inpatient and outpatient visits, as well as disease diagnoses for all workers eligible for the employer-sponsored medical insurance. We combine these records with data on job characteristics and job demand, as well as other worker-level covariates including sex, age, race, plant location, state of birth, and tenure at the firm.

Our primary analytical sample contains 4,672 retirees from 26 plant locations who were employed at the firm for at least 12 consecutive months between 1996 and 2013 and retired during that period, and who are eligible for employee-sponsored health care benefits. For analysis involving questions related to job demand, we must restrict our analytical sample to the plants at which the job demand survey (see description in measures section below) was administered. As such, our analytical sample for analysis related to job demand is 1,707 from seven principal plants.

Measures

Health Measures

Health status is assessed through linkages to medical claims records from employee-sponsored health insurance and includes all inpatient and outpatient visits and diagnoses of health conditions. Nearly all workers are eligible and covered under the employee-sponsored health insurance so data are available for any eligible workers during the 18-year time period between 1996-2013.

Six chronic diseases will be assessed using ICD-9 codes from the health claims data: asthma/COPD, arthritis, diabetes, depression, hypertension, ischemic heart disease, and musculoskeletal problems. We assess three different health-related events: a hospitalization, a new diagnosis, and a change to treatment or care of a chronic disease through outpatient visits.

We also measure baseline underlying health through the use of a health risk score. The health risk score is computed using an individual's Current Procedural Terminology (CPT) and ICD codes and use of healthcare services from the previous year and is a measure of individual-level health expenditures in the last year. These inputs are used to predict an individual's health expenditures in the coming year. The scores are standardized such that a score of 1 indicates that the individual's health expenditures are likely to fall at the median in the following year. Each unit increase predicts a one-fold increase in expenditures above the median. This score is produced by Verisk Health. Past work has shown that these (Hamad, Modrek, Kubo, Goldstein, & Cullen, 2015) scores are a close measure of general health. These scores are available for all workers on a yearly basis; we use baseline data for every worker, which is used to represent baseline general health.

Retirement

The HR records contain a dated record of all employee records. When an employee leaves the firm, retirement is coded as the reason for leaving. We calculate age at retirement by subtracting the date (day, month and year) of a retiree's listed birthdate from the HR records from their date of retirement.

Job Demand

Job demand data come from a job demand survey that includes externally rated measures of physical and psychosocial work for the majority of hourly workers. We will include two measures of physical demand: exposure to heat and physical demand. Both are measured on a Likert scale with the lowest value representing the lowest demand. We also include four measures of psychosocial demand based on the following questions: frequency of the need to work fast, importance of doing work without mistakes, frequency of conflicting demands, and level of decision-making and autonomy. A single expert rater at each plant rated the job demand by department and determined the job demand score. The raters were safety and health mangers who received training on the criteria to rate each job prior to the data collection. While the survey was done in 2003, the standardized titles allow us to assign workers job demand scores going both forward and back in time because the nature of work for hourly workers (the majority of the sample) has not changed drastically. We match job titles from 1996 through 2013. In order to avoid time-varying confounding, we use job demand at baseline, which represents either 1996 for workers hired before that year, or the year of hire for workers hired after that year. The vast majority of workers do not transition to lower or higher demand jobs across the observation window so baseline job demand is highly representative of a worker's demand for the entire observation window. In all cases, job demand variables are coded such that higher values equal more difficult or challenging demands.

Disability

The pathway from poor health to retirement decisions sometimes passes through short- or long-term disability. Previous work (Harrati, Hepburn, Meausoone & Cullen, under review) explored detailed trajectories of transitions into and out of employer-sponsored short- and long-term disability for this cohort. We will use these data in our current analysis as well. Gathered from HR records, the data include daily dates and duration of every episode of employee sponsored short and long-term disability. In previous work, we use cluster analysis to group workers by their trajectories of disability. We use worker cluster-membership in our analysis. We also flag any worker who ends up on company long-term disability insurance (N=65).

Other Covariates

The HR records also contain a number of other metrics that we will use as covariates or additional measures of job characteristics. These include: age, sex, tenure at company (based on original hire date, which is often before our observation window of 1996), whether the worker is unionized, and whether the employee works in a smelter or a fabrication plant.

Methods

Our empirical strategy consists of two parts. First, we estimate individual trajectories of health and then group, or cluster, workers with similar trajectories. Secondly, we model retirement age as a function of these health trajectories, as well as job demands and other personal and job-related characteristics. To do so, we employ three analytic techniques.

First, using the detailed health claims data for our sample of workers, we record any health-related incident (new diagnosis, hospitalization, new prescription, or office visit) for six chronic diseases for every month between 1996-2013. Importantly, these trajectories reflect their medical claims data and remain free of the bias inherent in self-reported data.

We use these data to create a matrix spanning 4,672 workers over 216 months (running from January 1, 1996 to December 31, 2013). Each month is coded with a measure of health if, in that month, there is: i) a new diagnosis, or ii) an outpatient visit for any one of the six health conditions detailed above, or iii) an inpatient hospitalization for any one of the health conditions. If more than one health event takes place in a month, both are listed (the cluster analysis, described below, can appropriately handle multiple events in the same time period). This matrix, thus, tracks trajectories of any new health events for any and all months for which a worker is employed during this 18-year observation period.

Second, we use cluster analysis to group workers into common trajectories of health. We use the technique of discrepancy analysis (Studer et al., 2011) as our methodological framework. The result of this analysis gives us groups of individuals clustered under the same typology of health patterns. We attempted clustering using a number of algorithms and find the most robust results (judging by average silhouette width) with both the Ward's minimum variance method (Blashfield, 1976) and the Partitioning Around Medoids (PAM) method (Kaufman &

Rousseeuw, 2005), yielding similar clusters. We use the Ward method for our reporting and our modeling.[1]

Third, we model the relationship between membership in a specific health cluster and retirement age, controlling for a number of demographic, job, and plant-level characteristics at baseline (1996) listed above. We carefully consider physical and psychosocial job demand, as well as transitions into employer-sponsored disability to understand pathways between health and retirement decisions. Because the distribution of retirement age is normally distributed, we use ordinary least squares (OLS) models with plant and year fixed effects. In order to ensure that we are not controlling for factors that are on the causal pathway between health trajectories and retirement age, we first model membership into a cluster by the list of covariates using a multinomial logit model. We exclude the covariate of unionized because it is collinear with smelter and the plant fixed effects.

Similarly, we model the distribution of retirement age as a function of job demand for the subsample for whom data are available (N=1707). We include health trajectories as well as the same covariates as in the model specification focused on health. Given the small sample size, we are underpowered to run a formal mediation analysis between health trajectory clusters and the various metrics of job demand but instead describe relationships between these health trajectories and job demand in the results section. We exclude smelter as a covariate in these specifications because of collinearity with a number of job demands.

Results

Descriptive Statistics

Table 1 displays the summary statistics for the entire analytical sample of retirees. The mean age at retirement is 58 (s.d. 4.59 years). This is much lower than the national average and can be explained by the strong pension system available to workers in the company. Workers become pension eligible under a variety of circumstances, but typically after 30 years of service. As such, many workers who start their employment in their early or mid-20s are eligible for full retirement by their mid-50s. This is further underscored by the mean tenure, which is 30.7 years of service. Given that this is a manufacturing company, the sample is majority male and also majority White. Seventy percent of the sample are hourly workers (as opposed to salaried) at

[1] Results for analysis using clusters created with PAM are available upon request.

baseline and about 60 percent work in smelters (versus fabrication plants). Workers in this firm have high rates of use of disability insurance benefits. More than 40 percent of workers had at least one spell of short-term disability during the observation window, 21.5 percent had two or more spells, and 1.4 percent of workers (N=65) went onto long-term disability during the observation period.

Table 3 displays the prevalence of disease for the six health conditions for our retirees while they were employed between 1996-2003. This prevalence is calculated using the date of new diagnoses and includes incident of disease before and during the observation window between 1996-2013 but does not include any new diagnosis after workers have retired. Hypertension and arthritis have the highest prevalence in this cohort, followed by diabetes and ischemic heart disease. The mean total number of inpatient and outpatient visits for workers is 5.8 visits for hypertension and 3.5 visits for diabetes (Table 1). One quarter of the sample has been diagnosed with one health condition, 12 percent have two health conditions, and 5 percent have been diagnosed with three or more of the health conditions in this study (Table 4).

Health Trajectories and Clusters

The cluster analysis using Ward's minimum variance method yields six different groups of workers that share similar health trajectories during the observation time period prior to retirement. The average silhouette width, a metric to validate the clusters, is 0.66. A silhouette width is a measure of how similar an object, in this case an individual health trajectory, is to its own cluster relative to other clusters. The silhouette ranges from -1 to 1, where a high value indicates that a trajectory is well matched to its cluster and poorly matched to other clusters (Rousseeuw, 1987). As the name implies, the average silhouette width is the mean of each of the 4,672 individual silhouette widths.

The largest of the six clusters, which we have dubbed Mostly Healthy, comprises 63 percent of the sample and is characterized by workers who remain disease free for most or all of the observation time. The second cluster, Hypertension (HTN), is comprised of workers whose health trajectories are dominated by hypertension, and represent 4.5 percent of the sample. The third cluster, Diabetes (DM), is made up of workers who have been diagnosed with diabetes as their only or primary health condition; they consist of 4 percent of the sample. The fourth cluster is called Few HTN-Arthritis. These are individuals whose trajectories are characterized by the

diagnosis and management of both hypertension and arthritis, but who experience relatively few health-related events over the observation window. This cluster is the second largest cluster and represents 27.5 percent of the sample. The fifth cluster, Many HTN-Arthritis, is similar to the cluster previously described, but with many health-related events associated with hypertension and arthritis. Finally, the sixth cluster, which represents less than 1 percent of workers, is characterized by workers who have a diagnosis of diabetes, but also myriad other health conditions, and whose trajectories involve a large number of health-related events.

Figure 1 plots the frequency of each of the six clusters. These plots describe the 10 most frequent sequences, or trajectories, for each of the six clusters. The x-axis represents a transition in the trajectory: for example, a new diagnosis, an inpatient visit, or a health-related event of a different health condition. Each value of s1-s*n*, therefore, represents a "transition" in the sequence. The y-axis represents the percentage of the total workers in that cluster that this particular trajectory represents; thicker bars, thus, represent a greater share of workers sharing this trajectory. Each health event is color-coded, following the legend.

Figure 2 plots the sequence density plots for each of the six clusters. These plots represent all of the individual-level worker trajectories in that cluster. The x-axis can be interpreted similarly to Figure 1, as markers of a health-related transition in the trajectory. The y-axis represent the frequency of each sequence. In essence, a sequence density plot can be interpreted as all of the individual-level, color-coded, health trajectories stacked vertically on top of each other to represent the entire cluster.

Table 5 displays summary statistics for each of the six clusters. We observe small differences in mean retirement age; an anova test (Table 6) confirms that these mean ages are statistically different from each other. We also observe higher shares of women and Whites in the Mostly Healthy cluster relative to the other groups and the total sample. As expected, large health gradients exist between clusters, though baseline underlying health, as measured by the risk score, is lowest in the hypertension cluster and highest in the DM mix group, a cluster where workers have higher numbers of health-related transitions and more co-morbidities. There also appear to be health-related gradients by job and plant characteristics. Individuals who work in smelters, which are considered to be environments with more physically-intensive jobs and higher levels of occupational exposures, are not as prevalent in the "Mostly Healthy" cluster relative to any other group.

Modeling of Retirement Age

Prior to our formal modeling, we run anova tests to ensure that mean retirement age does indeed differ between groups. Table 6 displays the results for anova tests with all of the job demand categories as well as the health clusters. There are statistically significant differences among mean retirement age for the health clusters and for all job demand metrics except the need to work without making mistakes.

Table 7 shows the results from the modeling of retirement age on the health clusters. The first column displays the results of the association for each of the five health clusters, relative to the mostly healthy group, on retirement age. The model is an OLS and coefficients can be interpreted in the usual linear fashion. Two clusters are positively associated with retirement age: the cluster characterized by workers with hypertension and the cluster characterized by a relatively limited number of health events related to hypertension and arthritis. In both cases, membership in these clusters *increases* the average age of retirement by 0.8 of a year and 0.33 of a year, respectively. The second column in Table 7 adds demographic and health controls including sex, race/ethnicity, and baseline underlying health, as measured by the risk score. The coefficients on the health clusters remain relatively unchanged and statistically significant. The third column adds job-related characteristics including whether an individual works in a smelter, their total tenure, and whether they are hourly (rather than salaried). Again, the coefficients on health clusters HTN and Few HTN-Arthritis remain stable and statistically significant. In other specifications not shown here, we include an indicator for workers who experience long-term disability. Perhaps unsurprisingly, spending time on long-term disability decreases average retirement age by 1.07 years ($p=0.079$) in the full model.

Table 8 displays the associations between retirement age, health trajectories, and job demand for the subset of 1,707 workers with information on job demand. The relationship between retirement age and health clusters remains fairly consistent to the full model, with both the HTN cluster and the Few HTN-Arthritis cluster positively associated with retirement age and with slightly higher coefficients than in Table 7. After the inclusion of demographic and job characteristics (column 3), only two facets of job demand are associated with retirement age. Increased exposure to heat at work is also positively associated with a later retirement age. With regards to psychosocial demand, lower autonomy of decision-making is associated with lower retirement age.

Discussion

In this paper, we characterized health trajectories over an 18-year period for a cohort of American workers. To do so, we used administrative data to track monthly, health-related events for six chronic diseases (asthma, arthritis, diabetes, depression, ischemic heart disease, and hypertension), including the diagnoses of new disease, hospitalizations, and outpatient visits. Using these data, we first used sequence and cluster analysis to characterize long-term trajectories of health and to group workers according to their patterns of work experience. We then modeled the relationship of these health trajectories to retirement age, accounting for baseline underlying health, as well as a number of demographic and job-related characteristics. Finally, we consider the role of physical and psychosocial job demand in retirement age.

Our analysis produces a number of findings that should be of interest to those studying retirement policy. In our data, workers can be categorized into a small number of distinct work trajectories. While the majority of workers in this sample remain relatively healthy for much of the observation window, others exhibit patterns of health-related events that are often marked by hypertension and/or arthritis. We find that clusters characterized with health events related to hypertension and arthritis are likely to retire later, not earlier. Moreover, we find an association with job demand and retirement, even after controlling for health. Specifically, we find that increases in exposure to heat are associated with higher retirement age, and less decision-making autonomy with lower retirement age.

A number of the results that are not particularly intuitive. In particular, membership in health clusters characterized by hypertension and arthritis, relative to the mostly healthy cluster, are *positively* associated with retirement age. We offer a possible interpretation that these numerous health-related events are signaling the proper management of these chronic diseases, thus allowing workers to extend their working life. Trajectories in which there are a greater number of health-related events could largely reflect workers' outpatient visits used to properly manage their health conditions through regular medical follow-up. This is even more likely given that the diseases in question, hypertension and arthritis, are chronic rather than acute in nature, and patients have a number of treatment options to consider. In this way, trajectories may reflect an intensity of health care and management rather than work-limiting health problems. Moreover, workers with chronic conditions that need to be actively managed may be at higher need for health insurance, which is provided through employment at the firm. The two

coupled together could create a powerful incentive to stay at work longer. Further work could explore measures of the severity of disease in the data as a way to distinguish between health events that may be considered management of chronic diseases versus work-limiting illnesses. Moreover, the relationship between exposure to heat and retirement appears counterintuitive. This relationship may be explained by healthy worker selection (Li & Sung, 1999), a form of bias whereby the healthiest workers are more able to withstand harsher work environments and remain more consistently employed. Healthy worker selection bias has been demonstrated in this cohort in past work (Brown et al., 2017). On the other hand, the results demonstrating lower decision-making abilities associated with earlier retirement is consistent with extant literature demonstrating that lower perceived control at work is often associated with lower participation at work, lower work satisfaction, and earlier retirement (Spector, 1986).

We acknowledge a number of limitations. The primary limitation has to do with the nature of our sample: workers at Alcoa are not a nationally representative group. While this does limit the generalizability of our findings, the workforce in the AMC cohort is also like the American public on a number of key dimensions (Cullen et al., 2006). Second, the methods employed here allow us to describe overall trajectories of health, but do not do they account for the timing or duration of the spells. Subsequent analyses will pursue such possibilities, and the strong incentives of retirement pensions in this cohort that may overwhelm other factors related to retirement decisions.

References

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, *83*(3), 377–388. https://doi.org/10.1037/0033-2909.83.3.377

Brown, D. M., Picciotto, S., Costello, S., Neophytou, A. M., Izano, M. A., Ferguson, J. M., & Eisen, E. A. (2017). The Healthy Worker Survivor Effect: Target Parameters and Target Populations. *Current Environmental Health Reports*, *4*(3), 364–372. https://doi.org/10.1007/s40572-017-0156-x

Cullen, M. R., Vegso, S., Cantley, L., Galusha, D., Rabinowitz, P., Taiwo, O., Sircar, K. (2006). Use of Medical Insurance Claims Data for Occupational Health Research. *Journal of Occupational and Environmental Medicine*, *48*(10), 1054–1061. https://doi.org/10.1097/01.jom.0000241049.23093.a4

Duggan, M., Singleton, P., & Song, J. (2007). Aching to retire? The rise in the full retirement age and its impact on the social security disability rolls. *Journal of Public Economics*, *91*(7–8), 1327–1350. https://doi.org/10.1016/j.jpubeco.2006.12.007

Dwyer, D. S., & Mitchell, O. S. (1999). Health problems as determinants of retirement: Are self-rated measures endogenous? *Journal of Health Economics*, *18*(2), 173–193. https://doi.org/10.1016/S0167-6296(98)00034-4

Hamad, R., Modrek, S., Kubo, J., Goldstein, B. A., & Cullen, M. R. (2015). Using "Big Data" to Capture Overall Health Status: Properties and Predictive Value of a Claims-Based Health Risk Score. *PLOS ONE*, *10*(5), e0126054. https://doi.org/10.1371/journal.pone.0126054

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Hoboken, N.J: Wiley.

Li, C.-Y., & Sung, F.-C. (1999). A review of the healthy worker effect in occupational epidemiology. *Occupational Medicine*, *49*(4), 225–229. https://doi.org/10.1093/occmed/49.4.225

McGarry, K. (2004). Health and Retirement Do Changes in Health Affect Retirement Expectations? *Journal of Human Resources*, *XXXIX*(3), 624–648. https://doi.org/10.3368/jhr.XXXIX.3.624

Modrek, S., & Cullen, M. R. (2012). Job Demand and Early Retirement. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2127722

Munnell, A. H., Sanzenbacher, G., & Webb, A. (2015). *What Causes Workers to Retire Before They Plan?* (SSRN Scholarly Paper No. ID 2665786). Retrieved from Social Science Research Network website: http://papers.ssrn.com/abstract=2665786

Munnell, A. H., Soto, M., & Golub-Sass, A. (2008). *Will People Be Healthy Enough to Work Longer?* (SSRN Scholarly Paper No. ID 1315487). Retrieved from Social Science Research Network website: http://papers.ssrn.com/abstract=1315487

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rutledge, M. S. (2018). *What Explains the Widening Gap in Retirement Ages by Education? Issue in Brief* 18-10. Chestnut Hill, MA: Center for Retirement Research at Boston College.

Spector, P. E. (1986). Perceived Control by Employees: A Meta-Analysis of Studies Concerning Autonomy and Participation at Work. *Human Relations*, *39*(11), 1005–1016. https://doi.org/10.1177/001872678603901104

Table 1. *Summary Statistics for Entire Sample of Retirees (N=4,672)*

	Mean	SD
Retirement age	58.6	4.59
Demographics		
Female (%)	11.2	
Nonwhite (%)	10.1	
Health		
Baseline Risk Score	1.4	1.51
Total Inpatient and Outpatient Visits in Observation Period		
Asthma	0.7	3.19
Arthritis	2.5	4.98
Depression	0.3	1.45
Diabetes	3.5	10.06
Ischemic heart disease	1.9	5.64
Hypertension	5.8	8.64
Job Characteristics		
Smelter (%)	62.0	
Hourly (%)	70.2	
Unionized (%)	57.7	
Tenure at company	30.7	8.45
Disability		
Ever on long-term disability (%)	1.4	
At least one short-term disability event (%)	44.3	
More than one short-term disability event (%)	21.5	
N	4,672	

Table 2. *Frequency of Physical and Psychosocial Job Demands (N=1,707)*

Job demand	Frequency (%)
Physical demand	
Sedentary	2.5
A little heavy	14.5
Somewhat heavy	54.6
Mostly heavy	27.1
Very heavy	1.4
Exposure to heat	
Never or almost never	35.3
Only occasionally	39.5
Sometimes	14.5
Almost all the time	10.7
Frequency of consecutive tasks difficult to combine	
Never or almost never	29.4
Only occasionally	19.1
Sometimes	39.1
Often	12.4
Extremely important to do work with NO mistakes	
Never or almost never	33.6
Sometimes	27.5
Often	38.9
Importance of Working Fast	
Never or almost never	36.4
Only occasionally	21.2
Sometimes	23.8
Often	18.6
Complete discretion of how work gets done	
Often	25.3
Sometimes	35.3
Only occasionally	18.6
Never or almost never	20.8
N	1,707

Table 3. *Prevalence of Health Conditions While Employed for Retiree Sample (N=4,672)*

Health condition	Frequency	Percent
Asthma	203	4.4
Depression	200	4.3
Diabetes	426	9.1
Hypertension	1,082	23.2
IHD	295	6.3
Arthritis	866	18.5
Sample size	4,672	

Table 4. *Prevalence of Multiple Health Conditions while Employed for Retiree Sample (N=4,672)*

Number of health conditions	Frequency	Percent
0	2,715	58.1
1	1,155	24.7
2	553	11.8
3	196	4.2
4	48	1.0
5	7	0.1
Sample size	4,672	

Table 5. *Summary Statistics of the Sample by Health Trajectory Cluster*

	Mostly healthy		Hypertension		Diabetes		Few HYP-ART		Many HYP-ART		Many DM-Mix	
Cluster type	Mean/ Percent	S.D.	Mean/ Percent	S.D.	Mean/ Percent	S.D.	Mean/ Percent	S.D.	Mean/ Percent	S.D.	Mean/ Percent	S.D.
Retirement age	58.46	4.68	59.26	3.93	58.98	4.09	58.78	4.57	59.79	3.68	59.14	3.09
Demographics												
Female (%)	12.74		5.71		5.38		9.70		6.06		7.69	
Nonwhite (%)	9.35		10.00		14.52		10.95		9.09		23.08	
Health												
Baseline Risk Score	1.35	1.68	1.20	0.74	1.51	1.25	1.37	1.23	1.36	1.20	1.59	0.81
Total inpatient and outpatient visits												
Asthma	0.27	1.23	2.45	7.36	0.52	1.80	1.33	4.55	0.29	1.13	0.33	1.00
Arthritis	1.46	2.87	3.87	5.76	3.31	4.64	4.21	6.68	8.19	16.02	5.22	8.81
Depression	0.16	0.93	0.23	0.70	0.22	0.98	0.45	2.27	0.19	0.60	0.00	0.00
Diabetes	1.13	3.98	4.09	7.20	32.21	21.18	3.40	7.13	3.55	8.50	61.33	46.56
Heart disease	0.86	2.95	6.48	11.66	4.74	8.83	2.84	6.60	4.39	10.29	5.67	11.67
Hypertension	2.12	3.94	21.72	12.55	10.54	10.99	8.93	7.65	37.48	12.61	11.67	13.56
Job characteristics												
Smelter (%)	56.34		75.27		73.81		69.68		77.42		77.78	
Hourly (%)	63.25		83.33		87.10		80.75		90.91		92.31	
Unionized (%)	52.13		69.89		69.64		63.96		87.10		100.00	
Disability												
Ever on long-term disability (%)	1.30		1.61		1.19		1.66		0.00		0.00	
At least one short-term disability event (%)	35.92		63.98		58.33		55.48		61.29		55.56	
More than one short-term disability event (%)	14.60		37.10		40.48		28.85		51.61		55.56	
Percent of total sample (N=4,672)	63		4.49		3.98		27.35		0.71		0.28	

Table 6. *Anova Test of Retirement Age with Job Demand and Health Trajectories Clusters*

	Prob > F	F-test value
Job demand metric (N=1,707)		
Physical demand	0.013	3.18
Heat	0.000	8.47
Combining tasks	0.075	2.31
No mistakes	0.698	0.36
Working fast	0.002	4.83
Discretion how work gets done	0.001	5.52
Health trajectory clusters (N=4,672)	0.024	2.59

Table 7. *Results from OLS Regression of Retirement Age on Health Clusters*

Variables	(1)	(2)	(3)	(4)
Clusters (ref= healthy)				
Hypertension (HTN)	0.807**	0.806**	0.890***	0.928***
	(0.328)	(0.328)	(0.325)	(0.334)
Diabetes (DM)	0.520	0.494	0.466	0.318
	(0.347)	(0.347)	(0.344)	(0.351)
Few HTN-arthritis	0.328**	0.320**	0.348**	0.397**
	(0.154)	(0.154)	(0.153)	(0.164)
Many HTN-arthritis	1.330*	1.333*	1.356*	1.087
	(0.803)	(0.803)	(0.796)	(0.788)
Many diabetes-mix	0.679	0.607	0.557	-0.204
	(1.275)	(1.275)	(1.264)	(1.455)
Female		0.0290	-0.0894	0.558**
		(0.214)	(0.214)	(0.241)
Nonwhite		0.536**	0.471**	0.103
		(0.225)	(0.224)	(0.246)
Baseline Risk Score			0.370***	0.297***
			(0.0442)	(0.0473)
Tenure (years)				0.0710***
				(0.00951)
Smelter				-1.188***
				(0.153)
Hourly				1.717***
Constant	58.46***	58.40***	57.89***	55.03***
	(0.0844)	(0.0904)	(0.107)	(0.337)
Observations	4,672	4,672	4,672	4,672
R-squared	0.003	0.004	0.019	0.076

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8. *Results from OLS Regression of Retirement Age on Health Clusters and Job Demand*

	(1)	(2)	(3)
Cluster (ref=healthy)			
Hypertension (HTN)	1.301***	1.380***	1.214***
	(0.429)	(0.434)	(0.421)
Diabetes (DM)	0.730	0.564	0.382
	(0.447)	(0.456)	(0.443)
Few HTN-arthritis	0.839***	0.845***	0.619***
	(0.236)	(0.239)	(0.234)
Many HTN-arthritis	2.373**	2.253*	1.817
	(1.158)	(1.233)	(1.196)
Many diabetes-mix	1.601	1.583	1.090
	(1.414)	(1.393)	(1.354)
Physical job demand (ref=low/sedentary)			
A little heavy		-1.273	-1.077
		(0.777)	(0.755)
Somewhat heavy		-0.943	-0.692
		(0.739)	(0.719)
Mostly heavy		-0.283	-0.158
		(0.773)	(0.750)
Very heavy		-3.111**	-2.258*
		(1.250)	(1.216)
Exposure to heat (ref= none)			
Only occasionally		0.714**	0.838***
		(0.334)	(0.324)
Sometimes		1.013**	1.106***
		(0.423)	(0.410)
Often		0.772*	0.964**
		(0.458)	(0.444)
Tasks are difficult to combine (ref=never)			
Only occasionally		-0.278	-0.0960
		(0.365)	(0.354)
Sometimes		-0.294	-0.343
		(0.317)	(0.308)
Often		0.154	0.184
		(0.463)	(0.449)
Cannot make mistakes (ref=never)			
Sometimes		0.256	0.325
		(0.321)	(0.312)
Often		0.322	0.300
		(0.326)	(0.316)

-continued-

Table 8. *Results from OLS Regression of Retirement Age on Health Clusters and Job Demand (cont'd)*

	(1)	(2)	(3)
Must work fast (ref=never)			
Only occasionally		0.461	0.166
		(0.347)	(0.340)
Sometimes		-0.210	-0.529
		(0.358)	(0.349)
Often		0.486	0.552
		(0.378)	(0.368)
Ability to make decisions (ref=always)			
Often		-0.221	-0.160
		(0.333)	(0.324)
Only occasionally		-1.107***	-1.103***
		(0.382)	(0.371)
Never		-0.907**	-0.911**
		(0.397)	(0.385)
Female			1.189**
			(0.530)
Non white			0.521
			(0.384)
Tenure (years)			0.150***
			(0.0179)
Hourly			3.425***
			(0.563)
Constant	58.16***	58.79***	50.36***
	(0.154)	(0.854)	(1.164)
Observations	1,707	1,707	1,707
R-squared	0.013	0.045	0.105

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1. *Frequency Plots of the 10 Most Common Sequences (Trajectories) for Each Health Cluster*

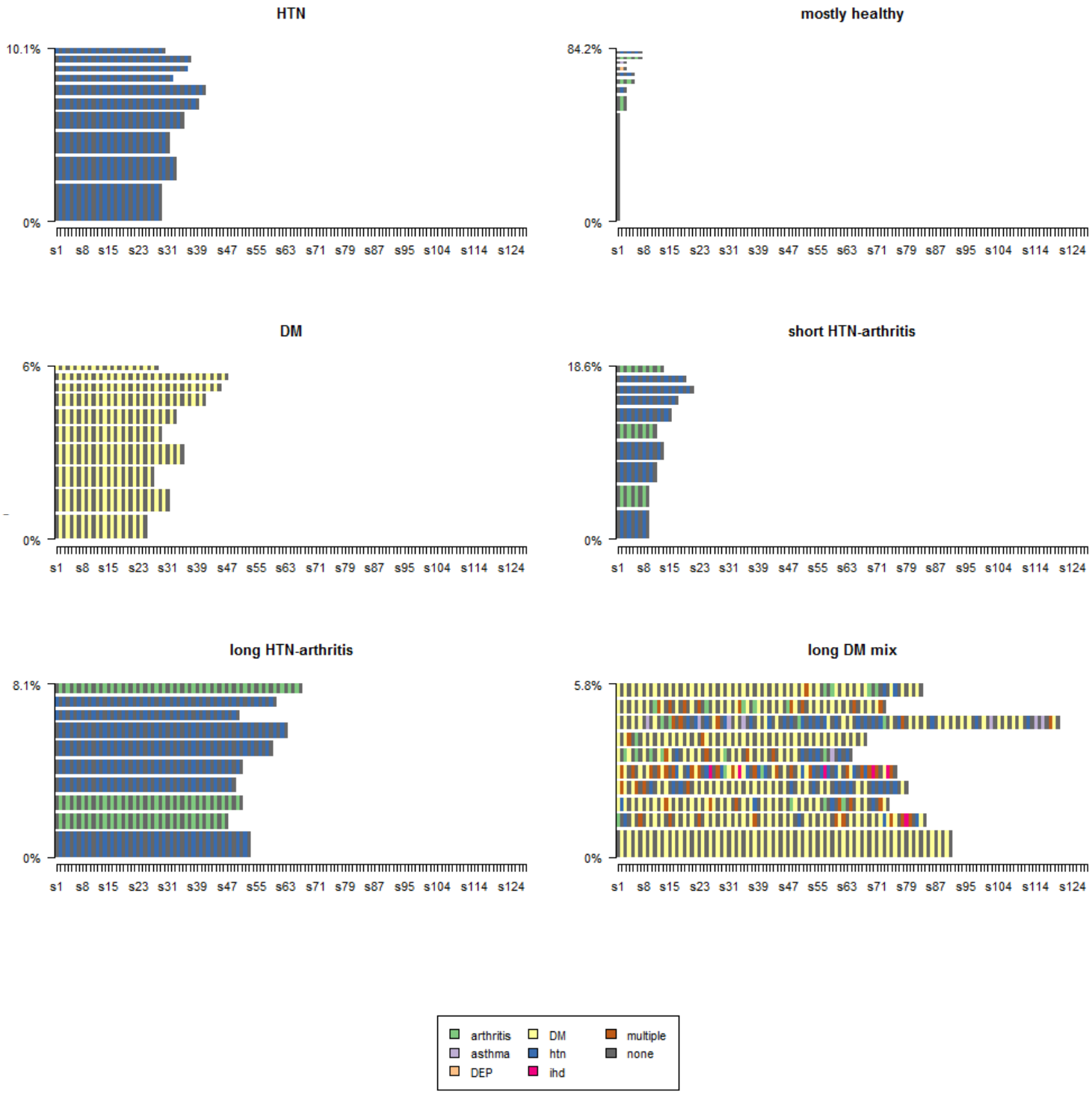
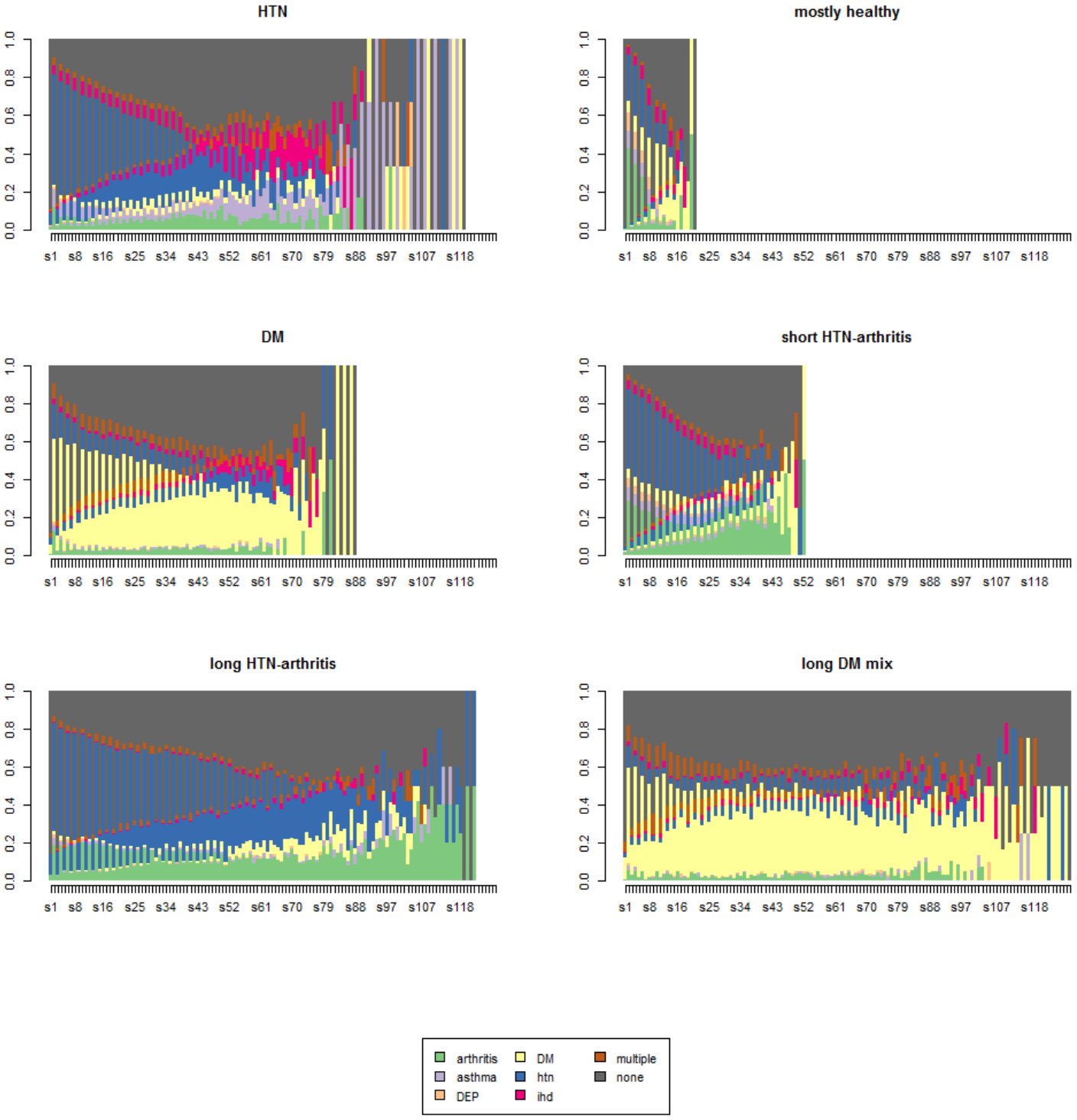


Figure 2. *Density Plots for All Six Health Trajectory Clusters*



RECENT WORKING PAPERS FROM THE CENTER FOR RETIREMENT RESEARCH AT BOSTON COLLEGE

Participation and Leakages in Oregon's Auto-IRA
Laura D. Quinby, Wenliang Hou, Anek Belbase, and Geoffrey T. Sanzenbacher, November 2019

Why Are 401(k)/IRA Balances Substantially Below Potential?
Andrew G. Biggs, Alicia H. Munnell, and Anqi Chen, November 2019

How Best to Annuitize Defined Contribution Assets?
Alicia H. Munnell, Gal Wettstein, and Wenliang Hou, October 2019

How Do Older Workers Use Nontraditional Jobs?
Alicia H. Munnell, Geoffrey T. Sanzenbacher, and Abigail N. Walters, October 2019

Will More Workers Have Nontraditional Jobs as Globalization and Automation Spread?
Matthew S. Rutledge, Gal Wettstein, and Sara Ellen King, July 2019

Do States Adjust Medicaid Enrollment in Response to Capitation Rates? Evidence from the Medicare Part D Clawback
Laura D. Quinby and Gal Wettstein, June 2019

The Effect of Medicare Part D on Evergreening, Generic Entry, and Drug Prices
Geoffrey T. Sanzenbacher and Gal Wettstein, May 2019

Is the Drop in Fertility Due to The Great Recession or a Permanent Change?
Alicia H. Munnell, Anqi Chen, and Geoffrey T. Sanzenbacher, March 2019

Will Fewer Children Boost Demand for Formal Caregiving?
Gal Wettstein and Alice Zulkarnain, March 2019

The Relationship Between Occupational Requirements and SSDI Activity
Matthew S. Rutledge, Alice Zulkarnain, and Sara Ellen King, February 2019

How Does Contingent Work Affect SSDI Benefits?
Matthew S. Rutledge, Alice Zulkarnain, and Sara Ellen King, February 2019

Do Pension Cuts for Current Employees Increase Separation?
Laura D. Quinby and Gal Wettstein, January 2019

Competition, Asymmetric Information, and the Annuity Puzzle: Evidence from a Government-Run Exchange in Chile
Gastón Illanes and Manisha Padi, January 2019

All working papers are available on the Center for Retirement Research website (https://crr.bc.edu) and can be requested by e-mail (crr@bc.edu) or phone (617-552-1762).