

**Characterizing Trajectories of Work, Disability and Health in Work and Retirement:  
A Multistate Analysis**

Amal Harrati  
Stanford University

Peter Hepburn  
University of California, Berkeley

Mark Cullen  
Stanford University and NBER

Prepared for the 19<sup>th</sup> Annual Joint Meeting of the Retirement Research Consortium  
August 3-4, 2017  
Washington, DC

The research reported herein was pursuant to a grant from the U.S. Social Security Administration (SSA), funded as part of the Retirement Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA; any agency of the federal government; Stanford University; the University of California, Berkeley; or the NBER Retirement Research Center.

## **Abstract**

Administrative data offer unique and all-too-rare opportunities to track individuals through their working lives. Using multiple, linked sources of employment data at a large manufacturing firm, this paper characterizes trajectories of work and disability across the tenure of a cohort of over 40,000 employees. We employ sequence and cluster analysis to group workers across common trajectories of work and disability. We then describe demographic and job-specific characteristics associated with these trajectories. These trajectories and their correlates provide insight into disability processes and their relationship to demographic characteristics, health, and working conditions of employees.

## **Introduction**

A growing body of research explores the complex relationships between disability, employment, and health. Transitions from work to disability have a range of direct, negative effects on labor force participation (CITE), unemployment (CITE), lifetime earnings (Breslin et al. 2007), and permanent exclusion from the labor market (CITE). Transitions into short- and long-term disability are associated with increased medical costs (Sears et al. 2012) and psychological distress (Bultmann 2002). A number of health conditions are associated with increased risk of work-place disability, including rheumatoid arthritis (CITE), diabetes (Virtanen 2015), musculoskeletal problems (CITE), depression (Kessler et al., 1999), and neuroticism (CITE). Aspects of working conditions, including psychosocial factors (Sullivan, 2013, Iles 2008), task monotony (CITE), and experienced stress of daily activities, are predictors of disability (Appellberg et al., 1996).

Previous research on these topics is limited by at least two problems. First, little research attempts to account for the endogeneity between employment, health, and disability. That is, studies explore effects of health on disability or vice versa, but rarely in a manner that acknowledges the bidirectional relationships at play. Second, most studies rely on data collected over short time-frames, with either limited or delayed follow up (CITE). Such data limit the conclusions that can be drawn about long-term trajectories and may mask significant variations. Many studies that observe rates of transitions back to work after a health shock or disability episode, for example, find that rates of return to work (a measure of success) to be quite high (CITE). Longer-term data, however, may reveal a different pattern. In a study of Ontario workers, the rate of successful returns to employment measured by first return to work is 85%, but the rate of success evaluated over a longer period is only 50% (Butler et al. 1995).

In this paper, we bring to bear a large set of administrative data that allow us to track the employment, health, and disability of a large cohort (n=42,146) of workers at Alcoa, a major American manufacturing firm. 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. We use these data to characterize trajectories of work and disability across the employment tenure and to explore variations in trajectory by demographic characteristics, health, and working conditions. To do so, we employ two techniques. First, we use sequence and cluster analysis to derive a typology of working

tenures. Second, we use multinomial logistic regression to model varying likelihood of cluster membership by worker and job characteristics.

In the following section we briefly describe our data and the methods used; additional information can be found in the supplemental materials. In the subsequent section we describe the trajectories of work and disability we find in these data and present regression results. We close with a brief discussion of these results and a summary of future research directions.

## **Data and Methods**

Our analysis is based on multiple sources of administrative data from Alcoa. The primary data are 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. We combine these records with health claims and disease diagnoses, prescription drug records, and data on job characteristics. Our final analytic sample includes 42,146 individuals employed by Alcoa at their 26 largest plants (across 15 states) for at least 20 weeks between January, 1996 and December, 2013. Employment can be both left- and right-censored; individuals could have started working for Alcoa before or at any time during the time period and may still be working when the period ends. The richness of these data allow for detailed exploration of the interactions between work, disability, health for a large number of employees across much of the life course. While the sample is not nationally representative, sample characteristics are close to national averages across a number of demographic characteristics (Cullen, 2006).

### *Measures*

#### *Disability and Leave of Absence Events*

The HR data records all changes in work status. Using these reports, we develop a recoding of employment status wherein each worker's tenure is represented as a vector of months. Each month is assigned to one of five states: 1) working, 2) absent due to short-term disability (STD), 3) absent due to long-term disability (LTD), 4) on-leave, and 5) having left the company (referred to as "terminated"). STD and LTD are company-specific administrative codes.

STD insurance is an employer-provided benefit for all Alcoa workers. This coverage provides wage replacement during spells of medical work absence of up to six months. For

hourly employees, work absence due to injury, hospitalization, or outpatient surgery is compensable beginning on the first day of the absence spell; there is a seven-day waiting period for illnesses. Benefits are payable only during periods of total disability. Benefits are not payable during periods of partial disability (i.e., modified duty) and employees who decline offers of modified return to work lose their eligibility for continued benefits. Claims are processed by a third-party administrator. Employees must submit medical documentation when requested in order for the claim to be initiated or continued, and the plan administrator can conduct medical examinations at its discretion. There is no employee premium for this lowest level of wage replacement. However, employees may select a higher level of wage replacement in exchange for a premium.

Alcoa offers LTD benefits to all active, full-time employees. LTD coverage is available after STD benefits expire for employees who are deemed “totally disabled” and unable to work. LTD benefits for workers under age 60 are available for the total period that they are disabled up until the day of their 65<sup>th</sup> birthday. For workers over age 60, benefits are available for a maximum of five years. If a second disability occurs within three months of the end of the first, in which the workers goes back to work between the two occurrences, the LTD is considered part of the first. If more than three months has expired, it is considered a new LTD event and workers must first exhaust their STD benefits for 26 months. Importantly, workers will not receive LTD benefits unless they show proof that they have applied for SSDI benefits. All other forms of leave were coded as Leave of Absence (LOA). This includes family leave, maternity leave, union-based leave, military leave, unpaid leave, paid leave and leaves due to disciplinary action or suspension.

If an employee leaves the company, they are coded as “terminated” after their last day of work. This is a general term for having left the company and does not differentiate between having been laid-off/fired and voluntary decisions to leave the company. Workers can, and sometimes do, come back to work at Alcoa; they remain in the data if they do come back to regular work. The HR data does have some records of retirement, but the data are incomplete and quality varies by plant. We do not use these data to assess retirement and instead we plan to use linkages to Social Security Administration data, which are currently in process.

Demographic, health, and occupational exposure characteristics are defined in the Supplemental Materials. Given that the sample is a manufacturing, the sample is largely white and male. Based on their risk score, this group of manufacturing workers are, by and large, relatively healthy and rates of chronic disease mirror those in the general population. The median income of the sample is higher than that of the United States as a whole.

### *Methods*

After editing and reshaping the HR data (see Supplemental Materials for a detailed description of the process), we have a dataset with 49,592 rows (individuals) and 216 columns (months). Because of left censoring, employees hired after January, 1996 are reported to be in state “unhired” until the start of their tenure with Alcoa. We reduce the complexity of the data by transforming them into distinct-state sequence format. This format removes the duration aspect of state-sequence data: rather than have one state for each of 216 months, these sequences only list states as they change. Every individual only has as many states in their vector as they experience transitions between states.

These distinct-state sequence data are the basis for clustering. We set aside 699 employees who ever enter the LTD state. For the remaining employees we develop a dissimilarity matrix using optimal distance matching with insertion-deletion costs set using the observed transition rates between the remaining states. We attempted clustering using a number of algorithms and find the most robust results (judging by average silhouette width) with the Partitioning Around Medoids method.

Results from the logistic regression are restricted to only the hourly workers, since the job-exposure characteristics are only available for hourly workers.

### **Results**

The clustering process described above yields eight groups (ASW of 0.723). Table 2 provides a schematic of the clusters; Table 3 offers descriptive statistics of the workers in each cluster. We collapse these clusters into three typologies based on their similarity: “regular work,” “short STD,” and “disruptive work.”

The “regular work” group is made up of two large clusters. The largest cluster is of workers who work regularly and then leave Alcoa permanently. Workers in this cluster are on average younger and have shorter tenure at the company. The second group is made up of those in continuous work who are right-censored. Similarly, this group is also younger and works longer. These two groups represent a large majority (64%) of all workers in our sample. A closer look at this group shows that this is a combination of longer-term workers who are relatively free of injury and disability, as well as a group of younger workers who start at Alcoa and leave soon afterwards. There may be a selection process at play in which some employees realize quickly that they are not be “cut out” for this work and leave the company after a relatively brief tenure.

Three of the clusters (Cluster numbers 3, 4, 5) involve regular work followed by one or two periods of short-term disability; these are combined in the “Short STD” group. Together, these workers represent 26 percent of the sample. The mean age of workers in these clusters is approximately two years older than those in the “Regular Work” clusters; years of tenure are fairly similar. Further work will detail the duration of these disruptions.

Finally, there are a number of clusters that are characterized by a startling number of work disruptions through both short-term and long-term disability. Clusters 6, 7, 8 all involve a trajectory of work that have at least two or three episodes of short-term disability or long-term disability. These are grouped together as “Disruptive Work.” These workers are, on average, older and have shorter tenure with Alcoa, suggesting that these workers may have started work at older ages. They have worse health than in any other cluster, as measured by both the number of outpatient and inpatient hospitalizations for a variety of chronic diseases and by the composite risk score measure of health.

### *The Sequence Distribution Plots*

Multinomial logistic regressions examine the associations between a number of demographic, job, and health-related characteristics and membership in the three composite groups. All regressions include year and plant fixed effects and are relative to the probability of being in the “Regular Work” typology. All time-varying covariates are included only for the first year when the worker is observed in the data, with the exception of total particulate matter,

which is measured in the last observed year, and hospitalizations, which are summed over a worker's entire tenure.

Neither age at first employment nor race have significant effects on tenure classification. Sex is significantly associated with membership: being female increases the odds of inclusion in the "Disruptive Work" type.

The number of outpatient hospitalization visits for heart disease, hypertension, and musculoskeletal conditions are all associated with significantly lower odds of membership in the "Regular Work" clusters. Health appears to be strongly associated with work trajectories in the "Disruptive Work" typology in a number of ways. First, with regards to risk score, a one standard deviation in risk score increases the odds of inclusion in this typology more than 1.4 times. Secondly, hospitalizations of arthritis, heart disease, hypertension, asthma and depression all increase the risk of inclusion in this category. Of course, these results imply statistical association but not directional causality; they could be the result of a bi-directional relationship between health shocks, health expenditures (as measured by risk score), and disability episodes.

Finally, we explore job characteristics and job exposures in these models. Exposure to cumulative total particulate matter increases the odds of inclusion in both "Short STD" and "Disruptive Work" (with highest odds for the latter). Exposure to total particulate matter increases the likelihood of being in the disruptive work category, has nearly no effect on the "Disruptive Work" group and slightly decreases to the likelihood of being in the "Short STD" group.

## **Discussion**

In this paper we sought to characterize working trajectories over a long period of time for a sample of American workers. To do so, we used administrative data to track work "states" for active workers, including regular work, disability episodes, leaves of absences, and terminations on a monthly basis for 18 years. Using these data, we carry out a number of analyses. First, we used sequence and cluster analysis to characterize long-term trajectories of work and to group workers according to their patterns of work experience. Second, we use multinomial logistic regression models to examine demographic, health and job-related characteristics that were associated with membership into particular typologies of work.



Our analysis reveals a number of interesting conclusions. First, there are a large number of distinct and diverse work patterns. While the majority of workers in this sample have very stable working patterns, there are a number divergent patterns, some that can be viewed as quite disruptive to job performance and work productivity. Moreover, there do appear to be age-related patterns into and out of certain transitions in states. For example, there appears to be selection out of work when a disability episode occurs in younger ages, but a reversal at older ages, which may be a function of pension incentives or limited job prospects.

Some demographic and health characteristics are particularly salient in this analysis. Of particular interest, for example, is the high likelihood of women to be included in the “Disruptive Work” typology. Certainly, given that this sample refers to manufacturing, the females working in this sample may be selected in a number of observable and unobservable ways. Little is known about women working in manual labor and these results point to the importance of further exploration into this special population.

There also appear to be important gradients related to health characteristics and chronic disease. A particular highlight is the finding that depression increases the likelihood of being in a “Disruptive Work” typology but not any other category. Depression is often overlooked as a potential driver of job disruption relative to other chronic disease, though evidence does point to its importance in labor market participation and worker productivity. (Lerner et al., 2008)

## References

- Appelberg, Kirsi, Kalle Romanov, Kauko Heikkilä, Marja-Liisa Honkasalo, and Markku Koskenvuo. "Interpersonal conflict as a predictor of work disability: a follow-up study of 15,348 Finnish employees." *Journal of psychosomatic research* 40, no. 2 (1996): 157-167.
- Breslin, F. Curtis, Emile Tompa, Ryan Zhao, Benjamin C. Amick III, Jason D. Pole, Peter Smith, and Sheilah Hogg-Johnson. "Work disability absence among young workers with respect to earnings losses in the following year." *Scandinavian journal of work, environment & health* (2007): 192-197.
- Bültmann, Ute, Ijmert J. Kant, Piet A. Van Den Brandt, and Stanislav V. Kasl. "Psychosocial work characteristics as risk factors for the onset of fatigue and psychological distress: prospective results from the Maastricht Cohort Study." *Psychological medicine* 32, no. 02 (2002): 333-345.
- Butler, Richard J., William G. Johnson, and Marjorie L. Baldwin. "Managing work disability: why first return to work is not a measure of success." *Industrial & Labor Relations Review* 48, no. 3 (1995): 452-469.
- Cullen, Mark R., Sally Vegso, Linda Cantley, Deron Galusha, Peter Rabinowitz, Oyebode Taiwo, Martha Fiellin et al. "Use of medical insurance claims data for occupational health research." *Journal of occupational and environmental medicine* 48, no. 10 (2006): 1054-1061.
- Elzinga, Cees H. 2003. "Sequence Similarity: A Nonaligning Technique." *Sociological Methods & Research* 32(August 2003):3–29.
- Elzinga, Cees H. 2005. "Combinatorial Representations of Token Sequences." *Journal of Classification* 22(1):87–118.
- Elzinga, Cees H. 2007. *Sequence Analysis: Metric Representations of Categorical Time Series*.
- Eriksen, Willy, Bård Natvig, and Dag Bruusgaard. "Marital disruption and long-term work disability A four-year prospective study." *Scandinavian journal of public health* 27, no. 3 (1999): 196-202.
- Horner, Elizabeth Mokyr, and Mark R. Cullen. "Linking individual medicare health claims data with work-life claims and other administrative data." *BMC public health* 15, no. 1 (2015): 1.
- Iles, Ross Anthony, Megan Davidson, and Nicholas F. Taylor. "Psychosocial predictors of failure to return to work in non-chronic non-specific low back pain: a systematic review." *Occupational and environmental medicine* 65, no. 8 (2008): 507-517.

- Kessler, Ronald C., Catherine Barber, Howard G. Birnbaum, Richard G. Frank, Paul E. Greenberg, Robert M. Rose, Gregory E. Simon, and Philip Wang. "Depression in the workplace: effects on short-term disability." *Health affairs* 18, no. 5 (1999): 163-171.
- Lerner, Debra, and Rachel Mosher Henke. "What does research tell us about depression, job performance, and work productivity?." *Journal of Occupational and Environmental Medicine* 50.4 (2008): 401-410.
- Noth, Elizabeth M., et al. "Development of a job-exposure matrix for exposure to total and fine particulate matter in the aluminum industry." *Journal of Exposure Science and Environmental Epidemiology* 24.1 (2014): 89-99.
- Pedersen, Jacob, et al. "Transitions between sickness absence, work, unemployment, and disability in Denmark 2004—2008." *Scandinavian journal of work, environment & health* (2012): 516-526.
- Sears, Jeanne M., Laura Blonar, Stephen M. Bowman, Darrin Adams, and Barbara A. Silverstein. "Predicting work-related disability and medical cost outcomes: estimating injury severity scores from workers' compensation data." *Journal of occupational rehabilitation* 23, no. 1 (2013): 19-31.
- Studer, Matthias. 2013. *WeightedCluster Library Manual: A Practical Guide to Creating Typologies of Trajectories in the Social Sciences with R*.
- Sullivan, Michael JL, Heather Adams, and Tamra Ellis. "A psychosocial risk-targeted intervention to reduce work disability: Development, evolution, and implementation challenges." *Psychological Injury and Law* 6, no. 3 (2013): 250-257.
- Virtanen, M., M. Kivimäki, M. Zins, R. Dray-Spira, T. Oksanen, J. E. Ferrie, A. Okuloff et al. "Lifestyle-related risk factors and trajectories of work disability over 5 years in employees with diabetes: findings from two prospective cohort studies." *Diabetic Medicine* 32, no. 10 (2015): 1335-1341.