Absenteeism, Health, and Disability in a Large Working Cohort

Amal Harrati[1]

Sepideh Modrek[2]

Mark R. Cullen[3]

1. Stanford University School of Medicine, aharrati@stanford.edu
2. San Francisco State University, smodrek@sfsu.edu
3. Stanford University School of Medicine, mrcullen@stanford.edu

Corresponding author:

Amal Harrati
Department of Medicine
Stanford University
1070 Arastradero Road
Palo Alto, CA 94305
aharrati@stanford.edu

This research was supported by the U.S. Social Security Administration through grant #5 DRC12000002-06 to the National Bureau of Economic Research as part of the SSA Disability Research Consortium. The findings and conclusions expressed are solely those of the author(s) and do not represent the views of SSA, any agency of the Federal Government, or the NBER. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA or any agency of the Federal Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. We are grateful to Chen Chen for research assistantship and to Yonathan-Ben Shalom for excellent discussant comments at the Disability Research Consortium Annual Meeting.

Keywords: Short-term disability, absenteeism, long-term disability, employment, labor market

Abstract

Because transitions from employment to disability leave are often the process of gradual health declines, these effects could operate through labor market mechanisms before the onset of disability. Specifically, increased absenteeism may be a precursor to eventual disability leave. Efforts to elucidate answers on this topic have been stymied by limited data available on absenteeism in the American workforce. We utilize a unique database of daily absenteeism records of nearly 10,000 employees at a large firm to examine the patterns of absenteeism and health that are associated with the use of employer-sponsored disability benefits. We answer the following questions: What are the patterns of absenteeism in this working cohort and are they disease-specific? Do these patterns differ for workers experiencing disability leave? Are patterns of absenteeism predictive of subsequent disability leave, and if so, for what diseases? Finally, do workers use absenteeism as a short-term substitute for disability benefits when opportunities for disability benefits are unavailable or limited? We find strong evidence that absenteeism predicts subsequent short-term disability leave at work, and that these relationships are patterned across diseases. We show that absenteeism increases substantially when disability applications are denied, suggesting that absenteeism may be used as a substitute for disability insurance when benefits are unavailable.

1. Introduction and Background

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 (Jones et al. 2006, Virtanen et al., 2006) unemployment (Bratsberg et al. 2010, Stattin 2005) lifetime earnings (Breslin et al. 2007), and permanent exclusion from the labor market (Gallo et al 2009). Transitions into short- and long-term disability are also associated with increased medical costs (Sears et al. 2014) and psychological distress (Bültmann et al. 2005). A number of health conditions are associated with increased risk of work-place disability, including rheumatoid arthritis (Backman et al. 2004, Sokka et al. 1999, Wolfe et al. 1999), diabetes (Virtanen et al. 2015), depression (Druss et al. 2000, Kessler et al. 1999) and asthma (Hakola et al. 2011, Eisner et al. 2006).

Moreover, transitions to disability leave are often the cumulative process of gradual health declines. As such, the effects of these declines could operate through a number of mechanisms related to employment before the onset of use of disability benefits. One such marker may be absenteeism at work; increased absenteeism may be a precursor to eventual disability leave. Moreover, since the process of applying for and getting approval for disability benefits is a lengthy one, absences might be used as a way to fill in the gaps of time while waiting for disability approval. Finally, the relationships between absenteeism and subsequent disability leave may differ by underlying health.

Efforts to elucidate answers on this topic have been stymied by limited data available on absenteeism in the American workforce. While evidence in European contexts provides growing evidence of a relationship between disease-specific absenteeism (often referred to as "sickness absence") and negative employment outcomes (Wallman et al. 2009, Karlsson et al. 2008), the

United States generally lacks data similar to the registry data commonly available in Europe. Evidence from Scandinavian countries often finds that use of sick leave and worker absences are strong predictors of future disability pension take-up, and the relationship between absence and disability spells is found to be strongest for absences spells with long durations (Andren 2007, Kivimaki et al.2007, Gjesdal and Bratberg 2003).

We utilize a unique database of daily absenteeism records of nearly 10,000 employees at a large manufacturing firm with a diversity of jobs and geographic locations, to carefully examine the patterns of absenteeism and health that are associated with short-term and long-term disability leave in a working population.

In this paper, we aim to answer the following questions:

- What are the patterns of absenteeism in this working cohort? Do these patterns differ for workers experiencing a disability event?
- Are these patterns of absenteeism disease-specific?
- Are patterns of absenteeism predictive of subsequent disability events?
- If so, for what diseases?
- Do workers use absenteeism as a short-term substitute for disability benefits when opportunities for disability benefits are unavailable or limited?

2. Data

This study relies on a unique dataset which links payroll data to health claims data for hourly workers (largely in production) from a large geographically diverse multinational aluminum manufacturing company, the American Manufacturing Cohort (AMC). We examine absenteeism in workers at seven sites from the years 2003 to 2008. The seven sites include both smelting plants and fabricating plants. The physical and psychosocial demands of the work vary substantially, but in general these jobs require some physical labor and many require repetitive movements. We draw on multiple sources of administrative data from AMC. 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 and payroll data, which has detailed daily shift work for these plants. Our final analytic sample focuses on 9,215 individuals employed by the firm at seven plants (across seven states). Employment can be both left- and right-censored: individuals could have started working for the company before this time period and may still be working after it; this amounts to a sample of 35,902 person-years. These data allow for detailed exploration of the interactions between absenteeism, disability leave, and health for a large number of employees across a six-year time period. While the sample is not nationally representative, sample characteristics are close to national averages across a number of demographic characteristics (Modrek and Cullen, 2013).

Absenteeism

Absenteeism metrics are created using the hourly payroll, or "punch clock" data from seven production facilities. For each employee and each day between January 1, 2003 and December 31, 2008, we observe the date of the shift, its length, and whether the employee was absent for that shift. These data are aggregated to create 52 consecutive work weeks as defined by the United Steel Workers of America Union master agreement with the company and taking into account holidays and worker weekends (which do not always fall on Saturday and Sunday depending on an individuals' weekly work schedule). For each individual in the data set, daily payroll data consists of the number of hours recorded for each calendar day (zero to 24 hours) and the pay-type associated with these hours: work, vacation, and unpaid. Paid hours are paid work hours. Vacation hours are paid vacation hours and include both plant-wide holidays (ie. 4th of July) and personal vacation days as defined by the labor agreement. There are no separate paid "sick days" or "personal days" at the firm; workers take those days as "lost pay" or "unschedued vacation" days (Hill et al., 2008). Unpaid hours are the hours a worker was scheduled to work on a given calendar day, but was not at work. These unpaid hours are considered absent days. *Thus, absenteeism is defined as any shift in which a worker was scheduled to work and did not show up to work.*

From the daily payroll data, we create a number of metrics of absenteeism. First, we count the total number of absent days per year for each worker. Next, we calculate the total number of absent spells, measured as one or more consecutive days of absences on a scheduled shift and accounting for weekends and holidays. We also calculate the duration, in days, of each absence spell and note the maximum duration in number of days. Finally, we delineate workers who have at least one spell of at least 2 or more consecutive days, referred to as "Extended Absent." Many of the workers have infrequent spells of exactly one day of absence and we aim to differentiate absenteeism behavior that suggests more serious drivers, and therefore potentially more predictive of future labor market activity. These are described in more detail in the results section later in the text.

Health Conditions

We identify incident cases of six health conditions using linked health claims data. These health conditions are: hypertension, arthritis, diabetes, asthma/COPD, ischemic heart disease (IHD), and depression. These chronic conditions are selected because they are common and have previously been identified and validated in research using claims data (Modrek and Cullen, 2013, Horner and Cullen, 2016). To determine health events, we use CPT procedure codes and ICD9 diagnostic codes. We only examine diagnostic codes if the procedure code ensures that the claims include a face-to-face component. We obtain the first date of the claim and designate the month of the diagnosis based on that date. For hypertension, we use one or more inpatient or outpatient claims with ICD9 diagnosis codes of 401–404. For arthritis, we use one or more inpatient or outpatient claims with ICD9 diagnosis codes of 710–719. For diabetes, we use one or more inpatient or outpatient claims with ICD9 diagnosis codes of 250, 357, 362, or 366. We define depression cases with one or more inpatient or outpatient claims with ICD9 diagnosis codes of 296, 309, or 311. For COPD/asthma cases, we use one or more inpatient or outpatient claims with an ICD9 diagnosis code of 493 (Modrek and Cullen, 2013).

Disability

Disability data comes from the human resources data as well and is defined as employer-sponsored disability benefits. These data include dates for the start day and end day of a short- or long-term disability, the health reason linked to the disability episode, and, for the year 2004, whether the disability application was denied or approved.

Short-term disability (STD) insurance is an employer-provided benefit for all active, full-time workers at the firm. This coverage provides wage replacement during spells of medical work absence of up to 26 consecutive weeks. For hourly employees, work absence due to work 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. The company offers long-term disability (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 65th birthday. For workers over age 60, benefits are available for a maximum of five years. LTD recipients must demonstrate that they have applied for federal Social Security Disability Insurance (SSDI) and employer-sponsored insurance terminates if and when workers become eligible for SSDI.

These data also include information on the disability insurance coverage for workers. Workers may opt into plans that cover different amount of income replacement while they are on disability leave; this coverage ranges from 40-100 percent of their income and a small percentage of workers have no coverage at all (slightly less than 4% of workers). Prior to 2005, all unionized workers were on the same disability insurance with a 50% income coverage. Starting in 2005, workers were able to pick between several plans with a range of coverage (Einav et al., 2012).

Other Covariates

Age is defined using the date of birth and, depending on the model, is defined as age at baseline (2003) or age at health diagnosis. We measure the sample observation time as defined by the total time a worker is employed at the firm during the study sample period of January 1 2003 – December 31, 2008. In a few cases, workers are hired, terminated, and then nearly immediately rehired (this is a function of seasonal work and union contracts, among other reasons). Our sample includes only workers who are continuously employed. Racial category is self-reported based on HR forms and includes a classification in which Hispanic is not separately categorized as ethnicity. Risk score is a measure of health defined by predicted total health expenditures for a worker based on data from the past year, calculated using an algorithm produced by Verisk Health, and is described in detail elsewhere (Hamad et al. 2015). The scores are standardized such that a score of 1 indicates that the individual's health expenditures are likely to fall at the mean in the following year, in a nationally representative population defined by Verisk. Each unit increase predicts a one-fold increase in expenditures above the mean. Finally, we include the percent of income replacement coverage that workers choose for their STD and LTD insurance.

Table 1 describes the sample of 9,215 workers for the entire sample, as well as separately for those with and without an employer-sponsored disability event during the analytical time period (2003-2008).

3. Results

What are the patterns of absenteeism in this working cohort?

Table 2 describes the patterns of absenteeism for the full sample (Column 1), as well as by gender, race/ethnicity, and age. As described in brief above, we create a number of yearly metrics for absenteeism. We sum the total number of absent days per year for each worker (Total Absent Days). We also aggregate the total number of spells—defined as one or more consecutive days of absences—per year per worker (Absent Spells). We also measure the largest number of consecutive absent days in any spell, per year, denoted as maximum absent duration. Since workers in our sample can start work after the sample period and/or end work before the end of the sample period, we repeat these three metrics (total absent days, maximum duration, and number of spells) adjusting for the observation time of the worker during the sample study period. We do so by aggregating each of these metrics over the sample study period (Jan. 1, 2003 – Dec. 31, 2008), and dividing by the number of years they were working in the study period.

Finally, we create an indicator (Extended Absent) for any worker that has at least one spell of at least two consecutive absences. Since more than 50% of absences are exactly one day long, this is an indicator for workers who do not show up for work, without excuse, for more than one consecutive day. In the full sample, 60% of workers are designated as "Extended Absent." The Extended Absent group has higher absenteeism rates: the number of absent spells is 3.48 compared to 2.55 and the mean maximum duration of an absent spell is 2.42 days.

The distribution of absent days is highly skewed. As seen in Figure 1, the modal number of absent days is 1, but the mean (Table 2) is 3.97 days and the 99% percentile is 35 absent days per year. The mean number of annual absent spells is 2.5 spells--the average maximum duration of those spells is 1.71 days. Table 2 also demonstrates stark gender differences in absenteeism patterns, with female workers having higher levels of absenteeism across all metrics. Females have more total absent days and more frequent and longer absent spells, including when adjusting for their observed time in the study period. Racial/ethnic differences also exist, with African-American workers having higher levels and rates of absenteeism relative to their White and Hispanic counterparts for the yearly absenteeism metrics, though Hispanics have higher rates of absenteeism once adjusted for the observation time in the sample study period. Interesting age patterns also emerge. Absenteeism decreases consistently with age; workers 60-69 years old have a mean total number of absent days that is half that of the group under 30 years old.

Do these patterns differ for workers experiencing a disability event?

Figure 2 describes the distribution of STD leave events in the time period. Panel (a) plots the number of workers per year. Between 2003-2008, there are a total of 6,796 STD events across 3,682 unique workers; 40 percent of workers have at least one STD leave across the study period. Consistent with previous work exploring transitions to disability in this cohort (Harrati et al., under review), many workers experience multiple disability leaves during the study period at the firm. Panel (b) displays the distribution of one or more STD leave during the study period conditional on having at least one leave. Of the 3,682 workers who experience one STD leave, nearly half have at least a second (or more) STD leave in the six-year period.

Workers who use short-term disability benefits have higher absenteeism across a number of metrics, as seen in Table 3. Columns (2) and (3) represent the full sample of workers divided into two categories—those who have a STD leave at some point in the study period and those who do not. Workers who use STD benefits have higher absenteeism by all metrics relative to those who do not, including a higher number of total absent days and higher maximum days of absent durations for any absence spell. Column (4) subsets the full sample to only those with at least one absent spell of two or more consecutive absent days (Extended Absent), to demarcate workers with potentially more serious absenteeism behavior. As expected, among those workers that have at least one spell of two or more consecutive absent days, absenteeism is higher among those that have at least one episode of disability leave within the time period (column 5) relative to those that do not (column 6). Finally, the last column (7) represents the subset of workers (N=42) who utilize long-term disability benefits at work. Workers who end up on LTD have lower average yearly absenteeism than those who experience an STD, and higher than workers without a disability event. Unfortunately, given the small number of workers who use LTD benefits, our ability to fully elucidate patterns is somewhat limited.

Figure 3 shows the relationship between absenteeism and eventual short term disability using an unadjusted time-to-event framework (i.e., Kaplan-Meier curves). The panel on the left describes the time to the first absenteeism event (using the above-mentioned definition of at least two consecutive days) for those with and without STD leave (denoted as EverSTD=1 for those workers that take STD leave). Consistent with the data described in Table 3 above, workers that have an STD leave have a greater share of absent days and have those days sooner in the analytical time period. The panel on the right examines the hazard of use of STD benefits for workers with or without an absence of two or more consecutive days in the time period (denoted as Extended Absent=1 for those workers with at least one absent spell of 2+ consecutive days). Workers with at least one absence of two or more consecutive days are much more likely to experience at least one STD leave.

Taken together, these data show that there are distinct patterns of absenteeism between workers with at least one STD or LTD episode at work and those without. Workers who experience an STD have higher levels of unexcused absences across any metric, and workers with consistently higher absenteeism rates are much more likely to experience an STD. This provides evidence of a relationship between absenteeism and disability; next, we explore differences across diseases as well as issues of temporality.

Are these patterns of absenteeism disease-specific?

Figure 4 describes the new incidence of disease for the six chronic diseases of study during the study period. The highest incidences of new disease are hypertension and arthritis, respectively, consistent with previous work (Modrek and Cullen, 2013). Conversions to STD events for new incidence across these diseases are quite high, ranging from 40 percent for hypertension and diabetes to nearly 90 percent for ischemic heart disease, which typically involve hospitalization.

Table 4 describes absenteeism patterns for the 2,706 workers who experience a diagnosis of at least one chronic disease during the study period (we use the first diagnosis of disease). These data reveal that workers with depression exhibit distinctly higher levels of absenteeism across all metrics. Both observation time-adjusted and yearly total number of absent days are higher, as are the frequency and duration of absence spells. Asthma is also associated with higher levels of absenteeism; hypertension and arthritis have the highest incidence in this cohort but absenteeism patterns do not deviate substantially from the full sample or the non-diseased sample.

Absenteeism metrics that adjust for the observation time of the worker in the study period show similar patterns. Workers with depression and arthritis, diseases that are characterized by symptoms that are chronic in nature, rather than by health shocks, have longer observation times than those with other diseases. Absenteeism patterns for IHD are not markedly different from the full sample. Patterns of absenteeism for workers with different diseases appear to be influenced by the nature of disease. Diseases that are ongoing with cyclical acute manifestations, such as depression and arthritis, are characterized by higher levels of frequent and sustained absenteeism, despite not necessarily having higher rates of conversion to STD. Conversely, workers who experience a heart attack or other health shocks that involve hospitalization will often transition directly into STD without returning to work, thereby avoiding potential absent days.

Are patterns of absenteeism predictive of subsequent disability events? If so, for what diseases?

The central question of this inquiry is to examine whether absenteeism behavior, which appears to be patterned by disease, can be used to predict subsequent STD or LTD leave for workers. To do so, we use a Cox proportional hazards model to explore the role of absenteeism on the time to short-term disability, multiple short-term disability events, and long-term disability leave. We start by estimating a model for all workers for the entire sample period (Table 5, column 1) to examine the predictive role of absenteeism metrics on time to STD. We include three metrics of absenteeism—the indicator of Extended Absent, the number of yearly absence spells, and the maximum duration of a yearly absent spell. We do not include the total number of absent days because it is highly correlated with the other metrics (r=0.84 with number of spells, r=0.77 with maximum duration of spells). Sensitivity analysis reveals that the inclusion of total absent days does not attenuate the coefficients on the other absenteeism metrics. We also include measures of the percentage of income-replacement of worker disability insurance, which can be time-varying if workers elect different plans in different years, and worker yearly health risk score which captures underlying health. We control for age at baseline, ethnicity, gender, observed time in the sample period, and plant fixed effects.

We then model a second specification (column 2) that allows for multiple STD events, since approximately 45% of workers who experience one STD subsequently experience a second or more. Finally, we model time to LTD for the full sample using the same model specification; there are 42 workers in this sample that experience LTD.

Given the possibility that absenteeism patterns may persist for a number of years before a disability episode, we also ran similar Cox models using lagged absenteeism, from one to three years. Our results are robust to these specifications, and coefficients on our variables of interest remain virtually unchanged. Moreover, we tested the sensitivity of the "Extended Absent" indicator by restricting the indicator to workers with at least 3 or more consecutive days, 4 or more consecutive days, and 5 or more consecutive days. The main findings from our models remain substantively unchanged; the results of this sensitivity analysis can be found in Table S1 in the supplemental materials.

Table 5 displays the results of the three model specifications. In line with the Kaplan-Meier plots in Figure 3, the indicator for having at least one absence of two or more consecutive days is highly predictive of first disability, multiple disabilities, and long-term disability leave. The frequency and nature of spells are also predictive for short-term disability events; both the maximum duration of absence and the number of absenteeism spells are also statistically significant, albeit with smaller effect sizes. One additional absence spell is associated with a 1.02 (single failure, 95% CI 1.01-1.02) hazard ratio of a first-time short-term disability event and every additional day of absence for the longest spell is associated with a 0.032 increase hazard of first-time STD.

Associations with absenteeism behavior for long-term disability leave are difficult to ascertain with certainty, since the small sample size means the model lacks statistical power to identify associations. Still, we observe the indicator for "Extended Absent" is a much stronger predictor of LTD, with a hazard ratio of 2.59 (single failure, 95% CI 1.26-5.31). Conversely, the number of absent spells is negatively related to the hazard of an LTD event and the duration of these spells are not statistically significant.

The results of these models also show important differences in other demographic characteristics and work-related factors. Consistent with other work exploring disability in this sample (Harrati et al., under review), important gender differences emerge in the hazard of short-term disability, with females having both higher incidence and higher hazards of STD in this cohort (single failure, OR=1.62; 95% CI 1.45-1.81). Additionally, racial differences that were observed in the unadjusted descriptive analysis disappear, and race/ethnicity is not a statistically significant predictor. The extent of income-replacement coverage also has a significant association with hazard of STD leave; relative to full income coverage, partial income coverage is associated with lower hazard of STD leave. Likewise, observation time in the study sample period has a negative relationship with the hazard of STD leave. This likely points to a selection process out of the firm for those workers who are very sick, a phenomenon that has been demonstrated in this cohort (Clougherty et al., 2009) as well as the simple tautological fact that a worker with a longer time in the study sample has a longer time period to be at risk of disease and disability.

With the exception of the relationship of risk score and time in study sample, the above-described relationships do not hold for long-term disability, though the low number of LTD events results in insufficient power to reveal statistically significant results.

Next, we explore the role of disease in these relationships and whether absenteeism patterns are more or less predictive of subsequent disability across different diseases. To do so, we run similar Cox models, but limit our sample to those who have a new diagnosis for the six chronic diseases in study in the time period (so as not to capture absenteeism patterns linked to previous health problems.) We include a "wash out" period to exclude all workers with a short-term disability event one year prior to the start of our time period in 2003. First, we run a model of time to STD pooling all workers with any new diagnosis (N=2,706), using the same specification as the models above but with the additional inclusion of indicators for the six diseases and age at diagnosis. Next, we run separate models for the time to STD for the subset of workers diagnosed with each of the separate diseases. We exclude IHD and asthma because there were not enough new cases for the models to be sufficiently powered. All other covariates remain the same as the models on the full sample.

Table 6 present the results of the model from the onset of disease to first time STD leave. The first column of results pools all workers with any new diagnosis of the six chronic diseases; subsequent columns display results for workers with the onset of a specific disease for the diseases in which there were enough new diagnoses for the models to be sufficiently powered. We observe that in nearly all cases, the indicator of having at least one absence of two or more consecutive days prior to STD and after diagnosis is highly predictive of a subsequent employer-sponsored disability event. When all diseases are pooled (and thus statistical power is larger) we also observe that the maximum duration of an absent spell in the year of diagnosis is predictive of subsequent disability. We do not observe any effects for the number of absenteeism spells prior to disability leave or the total number of days absent. Finally, having insurance with income coverage in the case of disability is highly predictive of subsequent disability, likely because of added worker protections. We do not observe any relationship between insurance coverage and time to STD in the disease-specific models; however, we may not be sufficiently powered to do so.

Do workers increase absent days when opportunities for benefits are unavailable or limited?

Finally, we make an attempt to understand whether absenteeism may be used as a substitute for disability benefits when disability insurance is unavailable. We have data available for denials of disability applications for one year of the data, 2004. In Figure 4, we plot the mean number of absent days from 2003-2008 for the 137 workers that have a denied application for STD benefits in 2004 and compare to the rest of the sample, as well as those workers without a denied application but with at least one absence of two or more days. Indeed, denied workers miss substantially more days of work, and absenteeism increases significantly around the time of a denial of an application. Notably, the number of days absent peaks in the year of denial of claim and then gradually declines to be nearly on-par with the remaining workers by 2008. We interpret this evidence to suggest that workers may use absenteeism as a substitute for disability leave when leave is not an available option. Unfortunately, given the small sample size of workers with denied claims, we cannot explore these denials in a multivariate framework.

Conclusion and Discussion

In this paper, we explore the relationship between patterns of absenteeism and subsequent disability in a cohort of workers. Specifically, we ask whether patterns of absences can be used to predict subsequent episodes of disability leave, and whether these predictions are patterned by disease. We draw several conclusions from these findings. First, we see clear differences in absenteeism behavior between those workers who eventually use STD benefits and those who do not. Absenteeism, by any number of metrics, is higher and more frequent. Conversely, absenteeism behavior for workers who use LTD benefits follows different patterns; the total number of absent days is greater, but the length or frequency of those spells is only slightly higher than the full sample and lower than workers who experience an STD leave. Second, we also see distinct patterns in absenteeism by disease. Workers with diseases that are ongoing but cyclical in nature, such as depression and arthritis, exhibit higher levels of absenteeism relative to those who experience a sudden health shock, such as a heart attack.

Third, absenteeism behavior does appear to be predictive of subsequent disability events, in particular first and multiple STD leave. Absent spells and their duration, as well as an indicator of being absent at least two or more consecutive days ("Extended Absent"), are all highly significant predictors of an increased hazard of use of STD benefits. These associations hold when we explore relationships with chronic diseases. Though we are limited in our statistical power to explore the relationships across all individual diseases, we show some evidence of absenteeism being predictive for diseases such as hypertension and diabetes. Finally, using a small subset of workers with denied STD applications, we see that absenteeism increased substantially around the time of denial, but returned to levels near those of the rest of the sample within four years, suggesting a potential temporary use of absent days as a replacement to disability leave for workers for whom the severity of their health claim is arguably less definitive.

These findings shed new light on the relationship between the work patterns, worker health, and disability leave in a specific working sample. How generalizable are these results? National prevalence of absenteeism rates are difficult to find in the existing literature, as measurements of absenteeism and presenteeism vary by survey and study, but findings from this study appear to be consistent with previous studies. For example, according to Maestas et al., (2019), recent estimates from the National Health Interview Survey show that workers take an average of 3 to 3.7 absent days per year, depending on whether or not they have access to sick leave (Ahn and Yelowitz 2016), and estimates from the Commonwealth Fund Biennial Health Insurance Survey suggest that 64 percent of Americans take at least one absent day in a given year (Davis et al. 2005). Recent work by Maestas et al. (2019) using a nationally representative sample of workers shows that workers who do not have protected sick leave take an average of 2.2 absent days. Given that our measure of absence is restricted to “unexcused” absences, ie. those absences that are not otherwise covered by sick or vacation leave, prevalence rates for the cohort in this study are very much in line with national estimates, including the fact that workers with STD episodes have higher rates than the overall estimates.

One of the strongest and most consistent findings relates to our indicator of workers who have at least one spell of two or more consecutive unexcused absence, which we call the “Extended Absent” group, as a predictor of subsequent disability leave. Importantly, this indicator does not appear to simply be a comprehensive measure of absent behavior; when we exclude this variable from our models, the coefficients on our other metrics of absenteeism remain relatively unchanged. Moreover, the measurement of the metric using 2, 3, 4, or 5 days reveals similar patterns (see Table S1 in Supplemental Materials). We interpret this to suggest that this indicator is capturing something unique to absenteeism behavior that is not measured in the frequency or duration of absence spells. There are a number of possible explanations for what this metric might be capturing. One possibility is that this represents a group of workers that is simply in worse health. However, the prevalence of disease across the Extended Absent group is not markedly different from the rest of the sample. Table S2 (in Supplemental Materials) shows the rate of prevalence for the six chronic diseases. It is certainly possible that these workers are in worse health in ways that are not measured by the six chronic diseases in this study, but we do not have data to bear on this. The rate of STD leave is higher for the Extended Absent group relative to others; 51% of Extended Absent workers experience at least one STD leave compared to 32% of workers who are not. Organizational theorists and social psychologists studying work absenteeism have suggested that workers may use absence as a mechanism for control between other, non-work related activities (George, 1989), or a withdrawal from surveillance or an undesired activity (Chadwick-Jones), or perhaps simply a habit (Goodman, 1984). Further work is warranted on this question.

This study is not without limitations. First, the relationships we observe are associational and not causal. Despite our efforts to capture possible confounding, there may be unobserved factors that explain the estimated relationships between absenteeism and disability. Secondly, and perhaps most importantly, we have a limited ability to capture the full possibility of healthy worker selection. Sicker workers may either not be employed or may leave the firm sooner; our data can capture many health conditions, but there may be underlying differences in health that are unobserved and would bias our results. Moreover, the sample of manufacturing workers may not be representative of all working individuals, and the specific benefits and working conditions of the firm may not be generalizable. Finally, our samples of workers on long-term disability and for specific diseases are small and some of our results are underpowered.

While there is interest in the role of absenteeism in various labor market outcomes, data on absenteeism in the United States is extremely limited and relationships between absenteeism and labor market outcomes remains relatively understudied. This paper provides detailed evidence that patterns of absenteeism might be a means by which to assess eventual disability for a number of chronic diseases.

References

1. Backman CL. Employment and work disability in rheumatoid arthritis. Curr Opin Rheumatol. 2004 Mar;16(2):148–152.
2. Bratsberg B, Fevang E, Roed K. Disability in the Welfare State: An Unemployment Problem in Disguise? Rochester, NY: Social Science Research Network; 2010 Apr
3. Breslin FC, Tompa E, Zhao R, Amick BC, Pole JD, Smith P, et al. Work disability absence among young workers with respect to earnings losses in the following year. Scand J Work Environ Health. 2007;33(3):192–7.
4. Bültmann U, Kant I, Brandt P a. VD, Kasl SV. Psychosocial work characteristics as risk factors for the onset of fatigue and psychological distress: prospective results from the Maastricht Cohort Study. Psychol Med. 2002 Feb;32(2):333–45.
5. Chadwick-Jones J. Social Psychology of Absenteeism. New York: Praeger; 1982.
6. Clougherty JE, Eisen EA, Slade MD, Kawachi I, Cullen MR. Workplace status and risk of hypertension among hourly and salaried aluminum manufacturing employees. *Social Science and Medicine 2009* Jan;68(2):304-13.
7. Cullen MR, Vegso S, Cantley L, Galusha D, Rabinowitz P, Taiwo O, et al. Use of Medical Insurance Claims Data for Occupational Health Research. J Occup Environ Med. 2006 Oct;48(10):1054–1061.
8. Druss BG, Rosenheck RA, Sledge WH. Health and disability costs of depressive illness in a major U.S. corporation. Am J Psychiatry. 2000 Aug;157(8):1274–8.
9. Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen. "How general are risk preferences? Choices under uncertainty in different domains." *American Economic Review*102, no. 6 (2012): 2606-38.
10. Eisner MD, Yelin EH, Katz PP, Lactao G, Iribarren C, Blanc PD. Risk Factors for Work Disability in Severe Adult Asthma. Am J Med. 2006 Oct 1;119(10):884–91.
11. Gallo WT, Brand JE, Teng H-M, Leo-Summers L, Byers AL. Differential Impact of Involuntary Job Loss on Physical Disability Among Older Workers: Does Predisposition Matter? Res Aging. 2009 May 1;31(3):345–60.
12. Goodman RS. Absenteeism. San Francisco: Jossey-Bass Inc; 1984

13. Hakola R, Kauppi P, Leino T, Ojajärvi A, Pentti J, Oksanen T, et al. Persistent asthma, comorbid conditions and the risk of work disability: a prospective cohort study. Allergy. 2011 Dec 1;66(12):1598–603.
14. Hamad R, Modrek S, Kubo J, Goldstein BA, Cullen MR. Using “Big Data” to Capture Overall Health Status: Properties and Predictive Value of a Claims-Based Health Risk Score. Dalal K, editor. PLOS ONE. 2015 May 7;10(5):e0126054.
15. Hill III JJ, Slade MD, Cantley L, Vegso S, Fiellin M, Cullen MR. The relationships between lost work time and duration of absence spells: proposal for a payroll driven measure of absenteeism. Journal of occupational and environmental medicine/American College of Occupational and Environmental Medicine. 2008 Jul;50(7):840.
16. Horner, Elizabeth Mokyr, and Mark R. Cullen. "The impact of retirement on health: quasi-experimental methods using administrative data." *BMC health services research* 16, no. 1 (2016): 68.
17. Jones SRG, Riddell WC. Unemployment and Nonemployment: Heterogeneities in Labor Market States. Rev Econ Stat. 2006 May 1;88(2):314–23.
18. Karlsson NE, Carstensen JM, Gjesdal S, Alexanderson KA. Risk factors for disability pension in a population-based cohort of men and women on long-term sick leave in Sweden. European journal of public health. 2008 Feb 1;18(3):224-31.
19. Kessler RC, Barber C, Birnbaum HG, Frank RG, Greenberg PE, Rose RM, et al. Depression in the workplace: effects on short-term disability. Health Aff (Millwood). 1999 Sep 1;18(5):163–71.
20. Modrek, Sepideh, and Mark R. Cullen. "Health consequences of the ‘Great Recession’on the employed: evidence from an industrial cohort in aluminum manufacturing." *Social Science & Medicine* 92 (2013): 105-113.
21. Sears JM, Blanar L, Bowman SM. Predicting work-related disability and medical cost outcomes: A comparison of injury severity scoring methods. Injury. 2014 Jan 1;45(1):16–22.
22. Sokka T, Kautiainen H, Möttönen T, Hannonen P. Work disability in rheumatoid arthritis 10 years after the diagnosis. J Rheumatol. 1999 Aug;26(8):1681–5.
23. Stattin M. Retirement on grounds of ill health. Occup Environ Med. 2005 Feb 1;62(2):135–40.

24. Virtanen M, Kivimäki M, Vahtera J, Elovainio M, Sund R, Virtanen P, et al. Sickness absence as a risk factor for job termination, unemployment, and disability pension among temporary and permanent employees. Occup Environ Med. 2006 Mar 1;63(3):212–7.
25. Virtanen M, Ervasti J, Mittendorfer-Rutz E, Tinghög P, Lallukka T, Kjeldgård L, et al. Trends of diagnosis-specific work disability after newly diagnosed diabetes: a 4-year nationwide prospective cohort study. Diabetes Care. 2015 Oct;38(10):1883–90.
26. Wallman T, Wedel H, Palmer E, Rosengren A, Johansson S, Eriksson H, Svärdsudd K. Sick-leave track record and other potential predictors of a disability pension. A population based study of 8,218 men and women followed for 16 years. BMC Public Health. 2009 Dec;9(1):104.
27. Wolfe F, Hawley DJ. The longterm outcomes of rheumatoid arthritis: Work disability: a prospective 18 year study of 823 patients. J Rheumatol. 1998 Nov;25(11):2108

Table 1: Summary Statistics for Sample in AMC, 2003-2008 (STD=Short-term disability)

	Full Sample	With STD Leave	Without STD Leave
Extended Absent: 2+ consecutive days	58.6%	74.7%	47.8%
Female	6.9%	9.4%	5.2%
Race/Ethnicity			
White	79.7%	79.7%	79.7%
Black	11.4%	12.5%	10.6%
Hispanic	7.2%	6.5%	7.6%
Other	1.7%	1.2%	2.0%
Age (at Baseline)	42		40.5
Risk Score	1.2	1.5	0.9
Disability Insurance (Baseline)			
Coverage>=80%	13.9%	13.4%	14.2%
60%<=Coverage<80%	2.0%	1.0%	2.7%
40%<=Coverage<60%	81.5%	84.9%	79.2%
No Coverage	2.6%	0.6%	3.9%
Observations	9,215	3,682	5,533

Notes: Data from AMC from 2003-2008 for 7 plants.

Extended Absent refers to an indicator for a worker with at least one absence spells of 2+ consecutive days.

Table 2: Absenteeism statistics for AMC sample by gender, Race/Ethnicity, and Age

	Gender		Race/Ethnicity				Age				
	Female	Male	White	Black	Hispanic	Other	Under 30	30-39	40-49	50-59	60 and above
Extended absent: 2+ consecutive days	8%	92%	78%	13%	8%	2%	22%	23%	24%	28%	4%
Yearly Variables											
Total Absent Days (Mean)	6.3	3.8	3.8	5.0	3.9	4.4	5.9	5.3	3.4	3.1	2.7
Median Absent Days	3.0	2.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0	1.0	0.0
Maximum Absent Duration	2.4	1.7	1.6	2.1	1.9	2.1	2.5	2.1	1.5	1.4	1.4
Number of Absence Spells	4.1	2.4	2.5	3.1	2.2	2.8	3.5	3.4	2.4	2.0	1.5
Observation Time Adjusted Variables											
Total Absent Days (Mean)	7.1	5.4	5.2	7.5	6.6	6.8	8.2	7.0	4.5	3.9	3.5
Median Absent Days	4.4	2.7	2.6	3.8	3.2	3.7	5.4	4.2	2.3	1.8	0.9
Max Absent Duration (Mean)	4.9	3.5	3.4	4.5	3.9	4.3	3.8	4.2	3.5	3.4	2.2
Number of Absence spells	4.5	3.2	3.2	4.1	3.8	3.9	4.8	4.2	2.8	2.3	2.1
Observations (Person-Years)	2,408	33,494	28,966	4,117	2,267	552	4,801	7,111	10,139	12,325	1,526
Observations (Person)	631	8,584	7,343	1,050	662	160	1,897	1,905	2,158	2,639	616

Notes: Data from AMC from 2003-2008 for 7 plants.

Extended Absent refers to an indicator for a worker with at least one absence spells of 2+ consecutive days.

Table 3: Absenteeism statistics for AMC full sample and various disability leave sample subsets (STD=Short-term disability, LTD=Long-term disability)

	(1) Full Sample	(2) With STD Event	(3) Without STD Event	(4) With 2+ days consecutive absence	(5) Absent and With STD	(6) Absent and Without STD	(7) With LTD
Yearly Variables							
Extended Absent 2+ consecutive days	59.7%	74.7%	47.8%	100%	100%	100%	
Total Absent Days (Mean)	4.0	5.2	2.9	5.7	6.4	4.7	4.3
Median Absent Days	2.0	3.0	1.0	3.0	4.0	3.0	1.0
Number of absence spells	2.6	3.2	2.0	3.5	3.9	3.0	2.7
Maximum Absent Duration (Mean)	1.7	2.2	1.3	2.4	2.7	2.1	2.0
Observation Time Adjusted Variables							
Total Absent Days (Mean)	5.6	6.6	4.8	8.5	8.5	8.5	4.8
Median Absent Days	2.8	3.7	2.1	5.1	4.9	5.4	2.2
Maximun Absent Duration (Mean)	3.6	5.4	2.4	5.7	6.9	4.5	7.9
Number of Absence spells	3.3	3.8	3.0	4.7	4.7	4.6	2.2
Observations (Person-Years)	35,902	16,813	19,089	23,463	12,996	10,467	199
Observations (Person)	9,215	3,682	5,533	5398	2647	2751	45

Notes: Data from AMC from 2003-2008 for 7 plants.

Extended Absent refers to an indicator for a worker with at least one absence spells of 2+ consecutive days.

Table 4: Absenteeism statistics and age of onset by disease

	Full	All diseases	Hypertension	Diabetes	Depression	Asthma	IHD	Arthritis
Extended Absent 2+ consecutive days	58.5%	63.0%	31.1%	6.2%	6.0%	12.1%	37.0%	8.0%
Total Absent Days (Mean)	4.0	4.1	4.1	3.6	5.9	4.8	4.5	4.0
Median Absent Days	2.0	2.0	2.0	1.0	3.0	2.0	1.0	2.0
Maximum Absent Duration (Mean)	1.7	1.8	1.7	1.6	2.5	2.2	1.8	1.7
Number of absence spells	2.6	2.6	2.6	2.3	3.7	2.9	2.7	2.6
<u>Observation Time Adjusted Variables</u>								
Median Absent Days	2.8	2.4	2.4	2.2	4.8	2.7	2.0	2.5
Total Absent Days (Mean)	5.6	5.0	4.8	4.4	7.3	7.0	5.2	4.6
Maximum Absent Duration (Mean)	3.3	4.1	3.9	3.6	6.3	5.2	4.0	4.0
Number of Absence spells	3.3	3.0	2.8	2.7	4.3	3.6	2.9	2.9
Age at Baseline (full) or at onset	42.0	48.5	48.1	50.7	40.1	47.3	53.9	48.1
in study period	3.4	4.2	4.3	4.2	4.2	3.7	3.7	4.3
Observations (Person-Years)	35,902	10,456	4,527	1,594	595	696	990	3,925
Observations (Person)	9,215	2,706	973	346	131	169	245	842

Table 5: Results of Cox Model for Time to Event to First Short-Term Disability Leave and multiple Short-Term Disability leaves for all workers from 2003-2008

	Time to First STD	Time to Any STD (Multiple Failures)	Time to LTD
Extended Absent (2+ Days)	1.777***	1.930***	2.587***
Maximum Duration of Absence	1.022***	1.015***	1.123
Number of Spells	1.020***	1.028***	0.708***
Female	1.621***	1.637***	0.636
Age at Baseline	1.008***	1.004***	1.000
Sample Observation Time	0.865***	0.856***	0.400***
Health Risk Score	1.112***	1.107***	1.214***
Disability Insurance Coverage			
60%<=coverage<80%	0.744**	0.744**	0.0909
40%<=coverage<60%	0.691***	0.716***	0.000
No Coverage	0.636***	0.578***	0.000
Race/Ethnicity			
Black	1.051	0.966	1.47
Hispanic	1.089	1.004	1.269
Other	0.866	0.909	0.000
Person-Event Observations	1,290,298	1,825,688	1,822,540
Person Observations	9,215	9,426	9,426
Number of Events	3,676	7,187	42

*Note: Data from AMC 2003-2008 for all workers in payroll data for 7 plants. *** p<0.01, ** p<0.05, * p<0.1. Controls include sex, ethnicity/race, age at onset, observation time in the study period, and plant fixed effects. Person-events include all HR events (e.g. vacation days, non-disability leave, job changes). Extended Absent refers to an indicator for a worker with at least one absence spells of 2+ consecutive days.*

Table 6: Results of Cox Model for Absenteeism on Time to Short-Term Disability Leave

From Time of Health Diagnosis, 2003-2008

	All Diseases	Arthritis	Hypertension	Diabetes	Depression
Extended Absent (2+ days)	1.505***	1.196	1.997***	1.195	1.565
Maximum Duration of Absence	1.009**	1.022***	1.008	1.078***	1.018
Number of Absenteeism Spells	1.014	1.017	1.005	1.006	1.028
Disability Insurance					
60%<=Coverage<80%	0.609	0.837	0.588	0.427	2.342
40%<=Coverage<60%	0.677*	1.04	0.537	1.86	0.707
No Coverage	1.055	0.838	0.685	2.209	4.552*
Number of Person-Year Observations	184,802	58,603	90,697	28,921	9,450
Number of Unique Workers	1,887	703	823	280	106
Number of Events	889	390	318	106	54

Note: Data from AMC 2003-2008 for all workers in payroll data for 7 plants. Person-events include all HR events (e.g. vacation days, non-disability leave, job changes).

**** p<0.01, ** p<0.05, * p<0.1.*

Controls include sex, ethnicity/race, age at onset, observation time at study period, and plant fixed effects.

Extended Absent refers to an indicator for a worker with at least one absence spells of 2+ consecutive days.

Figure 1: Distribution of Total Absent Days, 2003-2008 for all person-year observations in AMC cohort

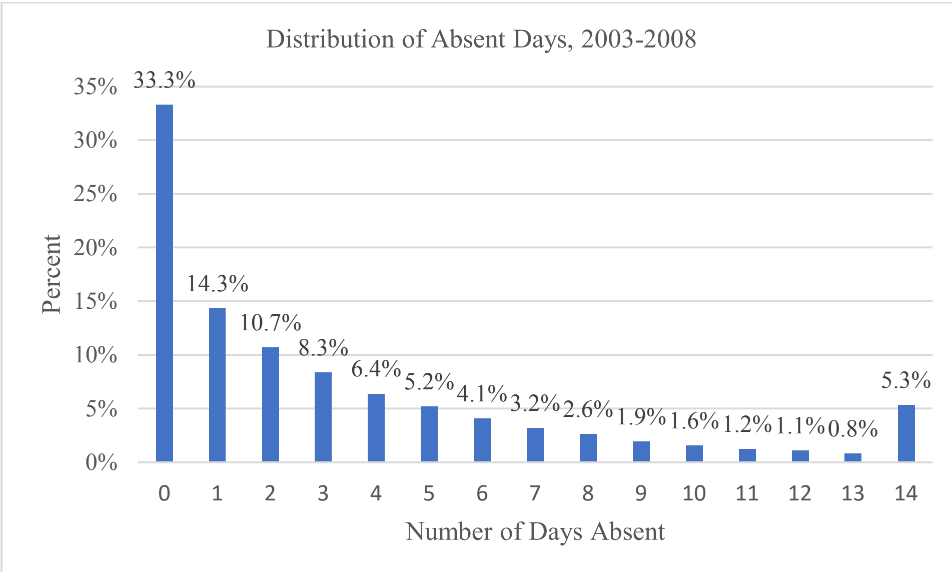


Figure 2: Percent workers with one or more Short-Term Disability leave, total and per year, AMC cohort, 2003-2008

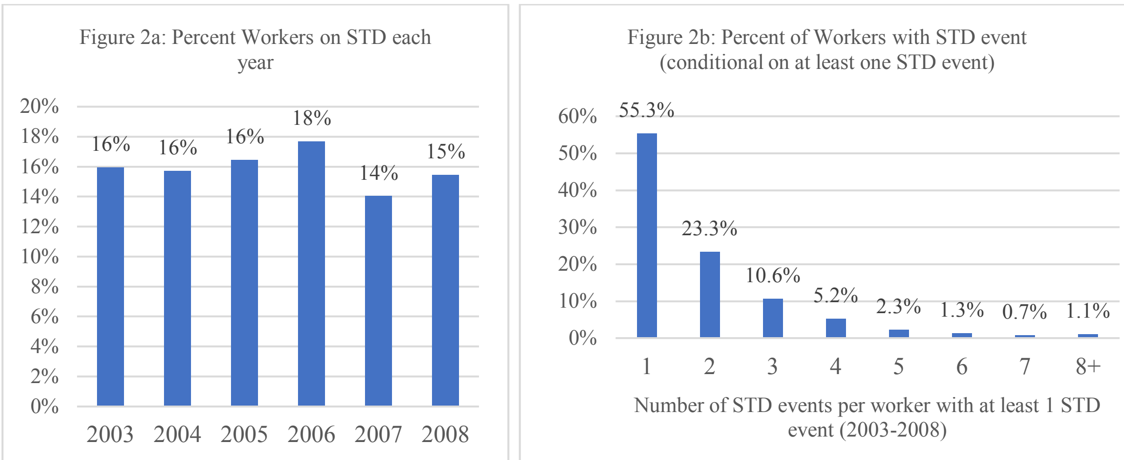
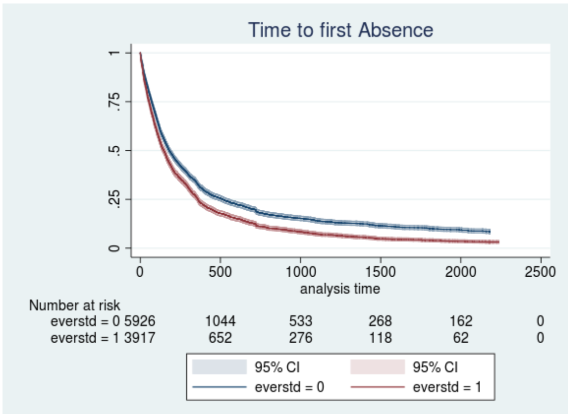


Figure 3a: Time to first absence for Workers with and without a short-term disability leave



Note: Everstd=0 indicates workers that never experience a short-term disability leave during the sample period, Everstd=1 indicate workers who do have at least one short-term disability leave.

Figure 3b: Time to short-term disability leave for group with and without at least absent spells of 2+ consecutive days

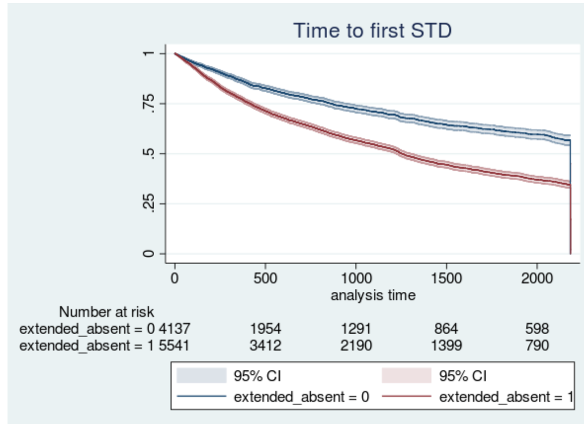


Figure 4: Percent of Workers with New Diagnosis and Conversion Rate to STD for Six Diseases, 2003-2008. The conversation rate is defined as the percent of individuals with diagnosis that result in STD.

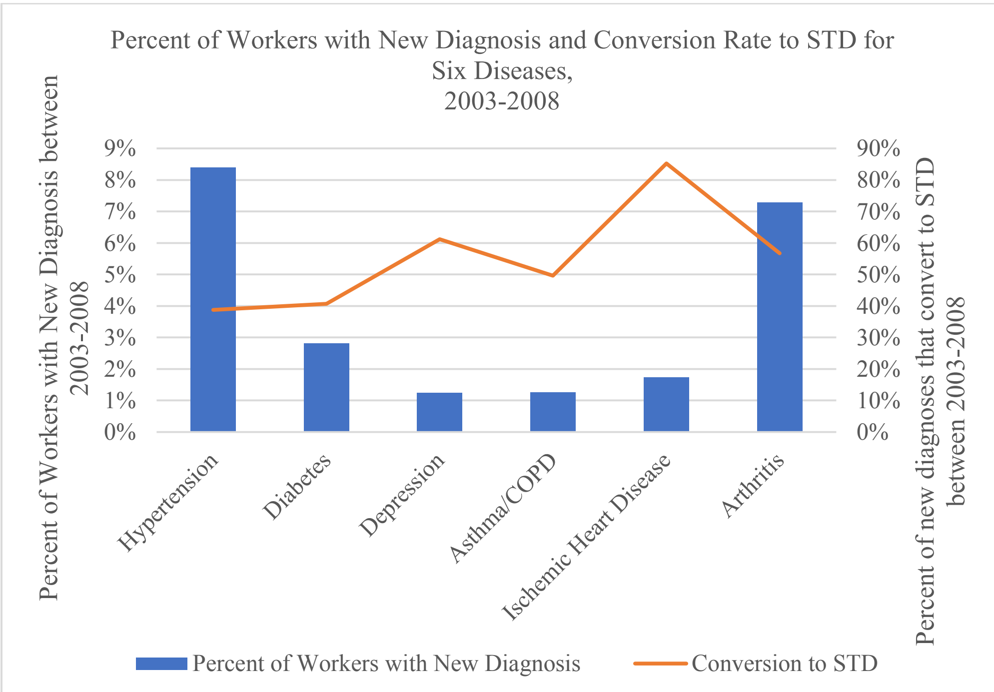
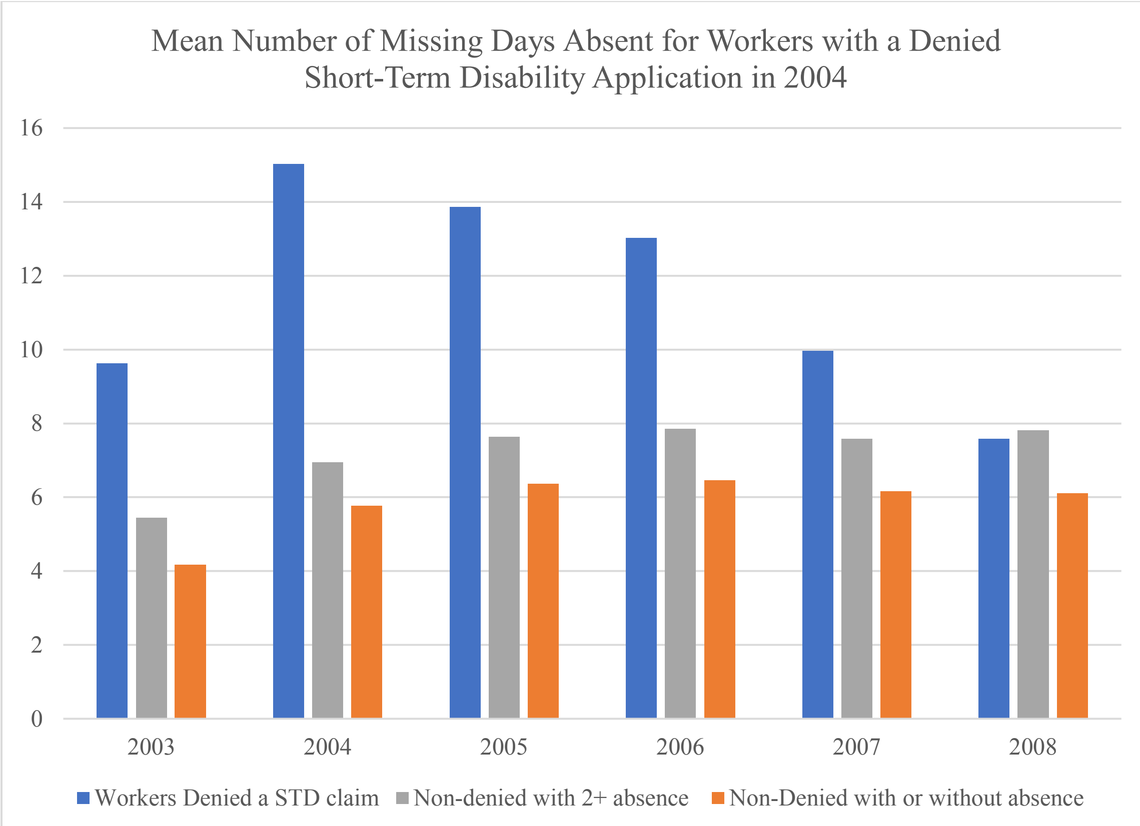


Figure 5: Mean Number of Days Missed for Workers Denied a Short-Term Disability Application in 2004, 2003-2008



Supplemental Materials

Supplement Table 1: Sensitivity analysis of extended absent indicator. Cox model of time to first short-term disability event. Columns represent the coefficients on covariates in the model, using definitions of the Extended Absent group for 2+ consecutive days (column 1), 3+ consecutive days (column 2), 4+ consecutive days (column 3), 5+ consecutive days (column 4).

	(1)	(2)	(3)	(4)
	Extended Absent(2+)	Extended Absent(3+)	Extended Absent(4+)	Extended Absent(5+)
Extended Absent (2+ Days)	1.755***	1.607***	1.565***	1.563***
Maximum Duration of Absence	1.032***	1.033***	1.033***	1.033***
Number of Spells	1.036***	1.040***	1.038***	1.036***
Female	1.592***	1.595***	1.595***	1.593***
Age at Baseline	1.008***	1.008***	1.008***	1.008***
Observation Time in the Study Period	0.864***	0.865***	0.862***	0.859***
Health Risk Score	1.112***	1.113***	1.113***	1.113***
<u>Disability Insurance Coverage</u>				
60%<=coverage<80%	0.745**	0.748**	0.746**	0.745**
40%<=coverage<60%	0.693***	0.689***	0.691***	0.688***
No Coverage	0.642***	0.638***	0.647***	0.647***
<u>Race/Ethnicity</u>				
Black	1.038	1.042	1.043	1.039
Hispanic	1.099	1.089	1.094	1.095
Other	0.862	0.854	0.853	0.862
Person-Event Observations	1,290,298	1,290,298	1,290,298	1,290,298
Person Observations	9,215	9,215	9,215	9,215

Supplement Table 2: Disease incidence, within study sample period 2003-2008, by extended absent group (i.e. workers at least one absent spell of 2+ consecutive days).

Disease	*Not Extended absent*		*Extended absent group*	
	Freq	Percent	Freq	Percent
Arthrities	307	31%	535	31%
Asthma/COPD	60	6%	109	6%
Depression	30	3%	101	6%
Diabetes	144	14%	202	12%
Hypertension	349	35%	624	37%
Ischemic Heart Disease	108	11%	137	8%
Total	998	100%	1708	100%