Labor Supply Dynamics Among Employed Workers with Mental Illness

2024 - 2025 Job Market Paper

Kristin Vrona

February 2025

Abstract

This study explores the complex relationship between mental health, physical health, and absenteeism in employed adults, with a focus on how a prior mental illness diagnosis, combined with current distress levels and physical health conditions, predicts absence behavior. By examining interactions between mental health diagnoses, a physical health index, and chronic physical conditions, the study reveals distinct behavioral patterns across gender. Notably, men and women first exposed to a mental health diagnosis show fewer absences under higher distress and physical health challenges, suggesting that diagnosis may prompt preemptive health management or influence cautious behavior aimed at minimizing health risks. Gender differences emerge as well: while women's absence behavior remains stable regardless of physical health fluctuations, men exhibit marked sensitivity to declines in physical health, indicating gender-specific responses in the intersection of mental health, physical well-being, and workplace attendance. These findings contribute to understanding how mental health diagnosis timing and gender interplay shape absenteeism and underscore the need for tailored mental health support in workforce policies.

1 Introduction

Mental illness is one of the most prevalent and costly health conditions in the United States, affecting 1 in 5 adults annually. While mental health issues entail direct treatment costs, the substantial economic burden primarily stems from indirect costs, such as reduced productivity and heightened rates of disability. Depression and anxiety disorders alone contribute to an estimated $1 trillion in lost productivity worldwide each year. Despite the importance of early and comprehensive treatment, U.S. adults report an average delay of 11 years between symptom onset and treatment. This gap has severe implications, as untreated symptoms lead to higher rates of substance use, depression, anxiety, and suicide.

Legislative efforts, including the Affordable Care Act's (ACA) mental health parity mandates, have attempted to address some barriers, yet access to mental healthcare remains limited. As of 2020, only 42.6% of adults with mental illness received treatment, and 155 million Americans live in areas with shortages of mental health professionals (NAMI, 2023). While parity laws have reduced discriminatory coverage practices, significant barriers persist[1]. These include high rates of denial for necessary treatments, delays in reimbursement, and a professional workforce increasingly strained by retirements outpacing new entrants (Bishop et al., 2014).

The economic impact of untreated mental illness underscores the need for a shift in focus. Rather than solely viewing mental health care as a rising expense, employers could benefit from recognizing it as a critical investment in workforce productivity. Evidence suggests that access to mental health services positively impacts absenteeism and productivity outcomes (Cseh, 2008; Fletcher, 2013; Ashwood et al., 2016). Employers who implement robust wellness programs and promote timely access to mental healthcare may see improved productivity and reduced long-term costs associated with employee turnover and absenteeism. By fostering a supportive workplace environment and offering

[1] National Institute of Mental Health. (2024, September). Mental Illness. https://www.nimh.nih.gov/health/statistics/mental-illness.
World Health Organization. (2024). Mental Health. https://www.who.int/health-topics/mental-healthtab=tab$_1$.

comprehensive mental health benefits, employers can not only mitigate the economic toll of mental illness but also contribute to a healthier, more resilient workforce.

In this paper, I estimate the impact of diagnosed mental illness, a subset of chronic physical illnesses, point-in-time distress, and self reported physical health status on annual absence rates among employed adults in the United States. Further, I examine how diagnosed mental illness impacts the role of general psychological symptoms of distress on absenteeism in recognition of the varying degrees of symptom severity among the diagnosed and undiagnosed and under the suspicion that the undiagnosed facing general distress may face more barriers to productive outcomes. I also examine the the nature of the relationship between physical health and mental illness in determining absenteeism.

The remainder of this paper is organized as follows: section 2 lays out the theoretical approach; section 3 describes the data source and variables; section 4 describes econometric implementation; section 5 provides results; section 6 concludes.

2 Theoretical Approach

The theoretical approach described in what follows applies the Becker (1965) theory of time allocation and household production and extends the Grossman (1972) health capital model. Time allocation models of labor supply consider the decision making process associated with market goods as well as time, with time modeled as a scarce resource. Individuals in a household are modeled as producers that take intermediate inputs such as time and market goods to produce useful commodities that are then consumed by the household. Without loss of generality, consider final products c_k for $k = 1,..,K$. An individual faces the following utility function:

$$U = u(c_1, c_2, ..., c_K), \text{ with } \frac{\partial U}{\partial c_k} > 0 \; \forall k. \tag{1}$$

Each final good c_k can be defined by its respective production process that turns market goods and time into final units of consumption. For final good c_k, this process is characterized by

$$c_k = f_k(T_k, x_k; e_k), \tag{2}$$

where T_k is a vector of time inputs allocated to the production of final good c_k, x_k is a vector of market goods used in production, and e_k represents the efficiency of the process, characterized by exogenous factors such as one's age or education. In each case, the efficiency factor, e_k, impacts how much time and how many market goods are required to achieve a certain level of a final good, and may or may not be equivalent across production processes. Given each production process, f_k, utility function (1) can be rewritten as

$$U = u(f_1, ..., f_K) = u(x_1, ..., x_K; T_1, ..., T_K). \tag{3}$$

The separability between time and market inputs exhibited by (3) demonstrates that household production is bounded from above due to the scarcity of income and time resources.

Of particular interest in this paper is the process of producing health which acts as the final good to be "consumed". I define the production process resulting in health stock, H, in the following manner:

$$H = f_h(T_h, x_h; MH, PH, e_h), \tag{4}$$

where T_h is a choice of time allotted to the development of health, such as time spent exercising and utilizing healthcare services, x_h is a vector of market goods utilized in health production, such as vitamins and supplements or health insurance coverage, MH and PH are innate endowments of mental and physical health, respectively, which characterize an individual's health productive capacity, and e_h is a vector of non-health-related factors driving the efficiency of the production process, such as education and age. Equation (4) is a concave function exhibiting diminishing returns to factor inputs. Function (4) innovates the health production function proposed by Grossman (1972). In this set up, the baseline endowment of health (MH and PH prior to the commencement of an arbitrary time period) characterizes the feasible set of attainable levels of health given available time and market inputs and preferences over how these inputs are allocated.

Define C as a conglomerate final good that nests each of the production functions of final goods besides health, $c_k, ..., c_{K-1}$, within it so that one can write

$$C = f_c(T_c, x_c, e_c), \tag{5}$$

where $T_c = \sum_{k=1}^{K-1} T_k$ and $x_c = \sum_{k=1}^{K-1} x_k$. Then, equation (1) can be rewritten in the following manner:

$$U = u(C,H) \equiv u(f_c, f_h) \equiv u(x_c, x_h; T_c, T_h), \tag{6}$$

so that the level of utility realized depends on choices of market inputs and time allocations. It is assumed that an individual maximizes their utility subject to the time constraint,

$$T = T_c + T_h + N, \tag{7}$$

where T is total time available and N is time allocated to occupational work, as well as a budget constraint,

$$I = wN + V = p_c x_c + p_h x_h, \tag{8}$$

where I is total income, V is non-earned income such as monetary gifts or inheritances, w is the market wage rate and p_i for $i = \{c, h\}$ are vectors of input prices corresponding to the market goods utilized in the production processes (5) and (4). The time constraint can be substituted into (8) to yield a single "full income" constraint,

$$\begin{aligned} w(T - T_c - T_h) + V &= p_c x_c + p_h x_h \\ \implies wT + V &= p_c x_c + p_h x_h + wT_c + wT_h. \end{aligned} \tag{9}$$

The left-hand side of the second line in (9) represents full income – the income received when an individual chooses to allot all available time to occupational labor.

The results of the current utility maximization problem are more intuitive if the production functions (5) and (4) are redefined, noting that $T_c \equiv t_c C$, $x_c \equiv b_c C$, $T_h \equiv t_h H$, $x_h \equiv b_h H$, where t_i and b_i for $i = \{c, h\}$ are vectors of the input time per unit and market goods per unit required to produce levels of final goods C and H, respectively[2]. With this, a single resource constraint can be expressed as

$$(p_c b_c + w t_c)C + (p_h b_h + w t_h)H = wT + V, \tag{10}$$

where the full price of each unit of the final good, C and H, is the sum of both the direct costs (prices of market goods) and indirect costs (time away from work) associated with each unit produced (Becker, 1965, 6). Now the individual maximizes utility by choosing optimal levels of b_i and t_i for $i = \{c, h\}$.

The optimizing individual will allot additional units of expenditure and time up to the point at which the marginal utility resulting from an additional unit of the respective input equals zero; this is equivalent to saying that available resources will continue to be allotted to production processes until the marginal product of the input is zero. It should be noted that choices of market good inputs and time inputs are not independent. A condition of utility maximization is that the marginal rate of substitution (MRS) between these types of inputs be equal to the ratio of per-unit input costs.

2.1 Labor Supply

At this point of the analysis, it is assumed that an individual has enough information to maximize their utility by choosing optimal bundle $\{C^*, H^*\}$. Given this decision, more information on an individual's preferences over home production (and thus, preferences over labor) can be revealed. Decisions on labor supply in the current set up can be intuitively illustrated by examining the demand for "forgone income". Define the right-hand side of (10) S, which is

[2] Note that t_h and/or b_h are decreasing in measures of the efficiency of the health production process, MH, PH, and e_h. t_c and/or b_c are decreasing in efficiency factor e_c.

thus full income that would arise if allotting all available time to work. The demand for forgone income $L(C^*,H^*)$, is then

$$L(C^*,H^*) = S - I(C^*,H^*), \tag{11}$$

where I is an individual's observed income. $L(C^*,H^*)$ can be thought of as the indirect cost of utility-seeking; it is the potential earnings lost when an individual allots positive units of time to the production of health and the conglomerate consumption good.

The current paper focuses on labor supply *among the employed*, exhibited by work absence, so that I henceforth focus on the special case of individuals with a positive optimal level of labor supply at baseline. Equation (11) can be further defined as

$$L^* = w(T_c^* + T_h^*), \tag{12}$$

for short run fixed wage, w, which is greater than or equal to the individual's reservation wage that optimizes their utility in any given period.

$I(C^*,H^*)$ from (11) can be expressed as

$$I = b_c p_c C + b_h p_h H. \tag{13}$$

Plugging $T_c + T_h = T - N$ into (12) and then plugging it and (13) into (11) yields

$$w(T-N) = S - b_c p_c C - b_h p_h H, \tag{14}$$

which can be rearranged and simplified as follows:

$$\begin{aligned} -wN &= V - b_c p_c C - b_h p_h H \\ \implies N &= \frac{b_c p_c C + b_h p_h H - V}{w}. \end{aligned} \tag{15}$$

We now have hours of work expressed as a function of health status. Taking the partial derivative with respect to health, we have

$$\frac{\partial N}{\partial H} = \frac{b_h p_h}{w} > 0. \tag{16}$$

Equation (16) tells us that better health induces more short-term labor supply. In (16), $b_h p_h$ represents the marginal cost of producing an additional unit of health using market goods and w is the per-unit time cost associated with the level of time allotted to producing health rather than working. The derivative $\frac{\partial N}{\partial H}$ can additionally be thought of as the marginal product of labor supply with respect to health. Rearranging (16) gives

$$\frac{\partial N}{\partial H} w = b_h p_h, \tag{17}$$

which states that an individual will continue to produce health up to the point where the marginal benefit of an additional unit of health (the additional labor income received for an additional unit of health) equals the marginal cost of an additional unit of health.

Consider an individual with a relatively low mental (physical) health endowment, MH_L. In the present framework, this an exogenous level that determines the efficiency of the health production process. This study relies on exogenous diagnosis information to identify individuals with mental illness, so assume MH_L is for an individual that has a diagnosed mental health condition. In this case, it already takes more of each of the time and goods resources to produce an additional unit of health, but the individual adapts to ensure optimality. However, when a shock to health occurs, that is observed health, H, is less than H^*, the individual will be more likely to need to need to alter their level of optimal labor supply given H in the period in order to account for not only the shock, but the relatively less efficient production process defining the number of inputs required to rebuild their health stock. In summation, this model leads to the general hypothesis that individuals with diagnosed mental (physical) illness will be more sensitive to observed health shocks and thus, will likely exhibit higher levels of absenteeism in general, and exacerbated absenteeism in response to health shocks.∎

The following section describes the data source as well as steps taken to generate data used for analysis and a brief overview of variables.

3 Data

This section describes aspects of the data sources used to obtain a representative sample of employed adults in the United States. I then provide an overview of the variables used in the analysis.

The data used in this study comes from the Medical Expenditure Panel Survey (MEPS) which provides nationally representative information on demographic and employment characteristics, healthcare utilization, and measures of health and well-being at the individual level. I utilize three public-use data files from the MEPS: the Full Year Consolidated Data File (FYCD), Medical Conditions File (MCF), and the Jobs File (JF). The population of interest is employed adults. I use the MEPS years 2010 to 2014 to obtain a random sample of adults employed at some point during the year of participation in the MEPS. The following describes the edits performed to the raw pooled data files.

The MEPS tracks individuals across two consecutive calendar years for a total of five survey rounds. The reference period for the third interview round of the survey spans across the two consecutive calendar years; fortunately, the data sources used in the analysis ascribe round three data to the appropriate calendar year for many pertinent variables, including the variable counting absences. Any other variables are refined using round start and end dates and weighting round three variables by the number of days spent in the respective year (i.e., round 3 year 1 variables are weighted using number of days in round three in year one, the remainder of the total round 3 days are used to weight round three data in year 2).

Upon limiting the sample to observations reporting employment at some point within the the years of interest, some persons are observed once for a single calendar year while other individuals are observed twice – once for each calendar year they participate in the survey. While late entry of households into the survey is not permitted, individual participants may enter the survey late if entering a participating household during the survey period. For example, late entry may be observed for a newly married individual who moves in with a partner that was part of the original random sampling scheme. A variable indicating whether an individual entered the survey late relative to other members in the sampled household is utilized as a control variable in the analysis. Additionally, the MEPS follows individuals that are part of an originally sampled household that eventually move to a different household; this may be the case for persons moving out of parent's residence or a spouse moving out of the home during a separation, for example.

I include multiple control variables that identify possible sources of attrition for individuals observed once, including variables indicating which year of participation (first year or second) in the MEPS these individual's are observed for and whether these individuals participated for both years of MEPS and thus are observed once as a result of my own data-generating process. The inclusion of these variables is inspired by strategies proposed by Verbeek and Nijman (1992) to roughly control for some of the attrition bias induced by the unbalanced sample design. Round start and end dates and job start and end dates are used to obtain to approximate the number of days that a person is employed during the year, therefore accommodating heterogeneity in exposure to the hazard of being absent from work across individuals.

The final pooled sample consists of 31,929 observations at least 18 years old at the start of their first round of participation in the MEPS. Students, military personnel, and institutionalized persons are removed from the sample.

3.1 Variables

A description for each of the variables included in analysis along with descriptions and summary statistics are presented in Table A.1. In this section, I describe variables in analysis.

Dependent Variable (A_i): The dependent variable of interest is a count variable (*sickdays*) representing absences from work due to an injury or physical or mental illness or ailment.

Mental Health (MH_i): The main explanatory variable of interest in this study is a binary variable indicating diagnosed mental illness (*keyMHdis*). This variable is based on responses reported in the Medical Conditions File (MCF) which provide condition-specific codes for various forms of mental illness. This variable is equal to one for individual's reporting diagnosis(es) of mood, anxiety, personality, or psychotic disorders[3]; I henceforth refer to these categories

[3]These categories of mental illness are chosen due to population prevalence, standard treatment protocols, and comorbidity hazards among them.

of mental illness as "key disorders". If an individual indicates a diagnosis of one or more of these key disorders, *keyMHdis* equals one and is zero otherwise[4].

It is important to further stress that the *keyMHdis* variable represents only *diagnosed* mental illness across the sample. The consideration of potential homogenous symptoms between individuals with and without a given diagnosis, as well as heterogenous symptoms across individuals within a particular diagnostic category are highlighted in the DSM-5, which separates itself from earlier DSM editions by its focus on a spectral approach to mental illness[5]. In acknowledgment of within-disorder heterogeneity and that mental health may be, to some degree, independent of a particular diagnosis, I utilize an index variable from the FYCD that measures one's general level of emotional distress over the last 30 days on a Kessler-6 scale with scores ranging from 0-24 and higher values indicating more psychological distress (Kessler et al., 2003; Ashwood et al., 2016). This information is collected as part of the MEPS Self-Administered Questionnaire (SAQ).

General (Physical) Health (PH_i): The FYCD files of the MEPS provide self-rated general health scores for each of the three interview rounds of the respective calendar year. These three variables are summed to create a poor-health index (*physhlth*) ranging from 1-13 that acts as the main measure of present physical health.

As the variables just described may be prone to attenuation bias because they are based on self-perceptions of health, various other measures of health are utilized in the analysis. The second analytical variable in this category, named *prtycnds*, represents the number of diagnosed priority conditions as defined by the MEPS. The priority conditions specified by the FYCD are cancer, heart conditions, asthma, stroke, chronic bronchitis, emphysema, high cholesterol, high blood pressure, diabetes, arthritis, joint pain and ADHD[6]. The MEPS identifies these priority condition categories "because of their relatively high prevalence, and because generally accepted standards for appropriate clinical care have been developed."

Job Traits (J_i): A variable measuring an individuals binary variables to represent job tenure in years ($tenure$) is created using information on job start and end dates provided by the MEPS. Additional job-related controls include indicators for labor union membership (*union*) and dummy variables identifying seasonal workers (*ssnl*), temporary contracts ($temp$), and part-time employees ($parttime$), with individuals that typically work less than 35 hours per week considered part-time. Other job controls include indicator variables for public sector positions (*pubsect*), industry and occupational categories, and firm size (represented by variables $1to19$, $20to99$, $100to499$, and $500plus$ based on sample quartiles).

Demographics (X_i): One's highest level of educational attainment is controlled for using a group of dummy variables (*belowhs*, *hsdeg*, $somecoll$, *bachdeg*, *bachplus*). Other variables control for race ($black$ and *asian*), Hispanic ethnicity (*hispanic*), native-status (*bornus*), age (age), family size ($famsz$), the number of young children in the household (*yngchldrn*), marital status (*married*), and socioeconomic status (*poor*, *lowinc*,*midinc*, *highinc*).

Other Control Variables (C_i): Regional indicator variables (NE, MW, W, S) are generated using FYCD variables indicating one's region of residence for the majority of the calendar year. Monthly data from the U.S. Bureau of Labor Statistics on regional unemployment rates is utilized to estimate the average unemployment rate faced by an individual ($unemp.rt$). Reference period start and end dates are used to generate these estimates for individuals that move to a different region during the year of interest; the unemployment rates are averaged across the months for which an individual reported a certain region of residence, then the average is taken across each of the regions that one resided in during the year. This measure of the average unemployment rate faced by an individual in a particular year provides

[4]Diagnostic information is housed in the MCF for additional categories of mental illness such as sexual disorders, conduct disorders, and developmental disorders. I choose not to indicate these diagnosis categories with *keyMHdis* due to the differing nature of the diagnostic criteria associated with these groups of disorders according to the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM–5; American Psychiatric Association, 2013).

[5]DSM-5 states: "Earlier editions of DSM focused on excluding false-positive results from diagnoses; thus its categories were overly narrow, as is apparent from the widespread need to use NOS [not otherwise specified] diagnoses. Indeed the once plausible goal of identifying homogenous populations for treatment and research resulted in narrow diagnostic categories and did not capture clinical reality, symptom heterogeneity within disorders, and significant sharing of symptoms across multiple disorders"

[6]ADHD is indicated using a binary variable for the diagnosis of the condition; it is *not* included in the count *prtycnds* variable.

the benefit of controlling for heterogeneity induced by macroeconomic features of the regional economy. For details on other control variables, see appendix.

4 Empirical Methods

I now describe the model specifications to be estimated. It should be noted that in recognition of differences in how men and women supply labor, and thus, how changes in health status impact decisions on when to be present at work, I conduct separate analyses for each of the sexes.

All models considered in this paper will estimate the count of days absent from work due to *one's own* (as specified my MEPS survey prompts) physical or mental illness or ailment per year. Because of the count nature of the dependent variable, I consider two distributions: Poisson and negative binomial. I conclude that a negative binomial model best fits the data due to the presence of overdispersion[7]. I account for year fixed effects and cluster observations at the individual and family level as some household units have multiple individuals in the sample grouping by sex[8].

4.1 Baseline Model Specification

Define the dependent variable, A_i for observation i, a count of the number of days absent from work over the span of one calendar year. I specify an exponential conditional mean function with baseline specification as follows:

$$E[A_i|MH_i, PH_i, J_i, X_i, C_i] = \exp(\beta^0 + MH_i\beta^{MH} + PH_i\beta^{PH} + J_i\beta^J + X_i\beta^X + C_i\beta^C), \tag{18}$$

where each β^j for $j = \{MH, PH, J, X, C\}$ is a $k_j \times 1$ vector of parameters; k_j is an integer equal to the number of explanatory variables held in corresponding matrix j[9].

Coefficients are typically not directly interpretable in nonlinear models, especially if the goal of research is to form policy implications (Ai and Norton, 2003; Long and Freese, 2006; Williams, 2009; Braumoeller, 2004; Buis, 2010). The estimate of interest is thus the conditional mean which will be utilized to compute average marginal effect (AME) estimates. AME estimates provide the average difference in the conditional mean expectation of absences resulting from a discrete change in an explanatory variable for each observation in the sample, or, for binary variables, the difference in values of the conditional mean expectation for each level of the binary variable. Standard errors for AME estimates are computed using the delta method (Wooldridge, 2010, *Econometric Analysis of Cross Section and Panel Data*, 737).

4.2 Mental Health, General Health, and Absenteeism

In the current paper, I focus on how diagnosed mental illness (characterized by binary variable *keyMHdis*) interacts with other measures of health to predict absence behavior. I add additional terms to the baseline model illustrated by (18). The first model includes an interaction term between the binary variable indicating diagnosed mental illness, and the point-in-time general psychological distress index variable. Henceforth referred to as $I1$, the conditional mean expectation of this extended model is defined as follows:

$$\begin{aligned} E[A_i|MH_i, PH_i, J_i, X_i, C_i] = \exp(&\beta^0 + MH_i\beta^{MH} + PH_i\beta^{PH} + J_i\beta^J \\ &+ X_i\beta^X + C_i\beta^C + MH_{1i} \times MH_{2i}\beta^{MH_1 \times MH_2}), \end{aligned} \tag{19}$$

where MH_1 is the first column of matrix MH and it holds values of *keyMHdis* and MH_2 holds values of *distress*, the second variable present in matrix MH.

[7] A zero-inflated negative binomial is also considered but performs no better than a basic negative binomial.

[8] I consider an individual random effects negative binomial to account for individual-specific unobserved heterogeneity to the same data in a working paper (Vrona, 2024, Mental Health and Absenteeism: the Role of Fringe Benefits as Moderators [Unpublished manuscript], Northern Illinois University. *Preliminary results available upon request.*

[9] For example, β^{MH} is a 2×1 vector holding parameter values for the two explanatory variables represented in matrix MH: *keyMHdis* and *distress*. For a full list of the variables included in each matrix and their definitions, see the appendix.

Next, define $I2$, a model that includes an interaction term between *keyMHdis* and the variable representing general health status, *physhlth*. The first variable in matrix PH, PH_1 is the *physhlth* variable. Then for the $I2$ specification, we have

$$E[A_i|MH_i, PH_i, J_i, X_i, C_i] = \exp(\beta^0 + MH_i\beta^{MH} + PH_i\beta^{PH} + J_i\beta^J + X_i\beta^X + C_i\beta^C + MH_{1i} \times PH_{1i}\beta^{MH_1 \times PH_1}), \quad (20)$$

which will be used to examine the the relationship between diagnosed mental illness and self-reported health on absenteeism.

Finally, a third specification utilizes an interaction between the mental illness indicator variable and the variable representing the number of chronic physical condition diagnoses, *prtycnds*. Define the $I3$ specification as follows:

$$E[A_i|MH_i, PH_i, J_i, X_i, C_i] = \exp(\beta^0 + MH_i\beta^{MH} + PH_i\beta^{PH} + J_i\beta^J + X_i\beta^X + C_i\beta^C + MH_{1i} \times PH_{2i}\beta^{MH_1 \times PH_2}), \quad (21)$$

where PH_2 from matrix PH holds values of the *prtycnds* variable.

I hypothesize that there are significant interactions between mental illness and variables *distress*, *physhlth*, and *prtycnds* so that the total impact of diagnosed mental illness on absenteeism varies with changes in these variables, and vice versa. Each of these three measures of health are anticipated to have positive estimate for the effect they have on absenteeism *in isolation* (recall that higher values of *physhlth* indicate reports of *poorer* levels of general health). I anticipate that having a diagnosed mental illness will exacerbate the effect of these variables on the number of absence days reported by an individual and that higher levels of *distress*, poorer physical health (indicated by higher levels of *physhlth*), and a greater number of priority conditions (higher levels of *prtycnds*) will increase the total effect of diagnosed mental illness on absenteeism. In summary, I hypothesize

$$\beta^{MH_1 \times MH_2} > 0,$$
$$\beta^{MH_1 \times PH_1} > 0, \text{ and}$$
$$\beta^{MH_1 \times PH_2} > 0.$$

As these three additional measures of health are measured as either a discrete index or continuous variable, I plot the interaction effects, as it is likely that the relationship is nonlinear and may change in magnitude at extreme values of index or continuous variables. After observing plots of interactions, I choose a few interesting values of each of these three variables and compute conditional average marginal effects (CAME) for variable *keyMHdis*, conditioning on these interesting values for further inference.

5 Results

This section provides an overview of results for men and women separately. The negative binomial models are calculated using maximum likelihood estimation and a unique dispersion parameter to account for sample dispersion is estimated in this process.

5.1 Baseline Model Results

The estimated dispersion parameter is approximately 0.21 for the baseline negative binomial model for men with an estimated standard error of 0.004; the equivalent values for the baseline estimator for women in the sample are 0.28 and 0.01, respectively. Table 1 reports AME estimates for a subset of variables. Pseudo R^2 are also reported. AME estimates represent the expected change in the count of absence days per year in response to a unit change in the respective explanatory variable, on average.

Men with a diagnosed mental illness are estimated to report 1.10 days of additional absence, on average. The average sample female is expected to report 0.92 additional annual absences when diagnosed with a mental illness. On the other hand, AME estimates for one's level of general psychological distress are slightly larger for women, as are estimates of the AME of higher levels of poor health (for variable *physhlth*). A marginal increase in general distress is estimated to increase absences among men in the sample by an average of 0.11 days. AME estimates for women suggest that an increase in general distress is associated with an increase in work absence by a factor of 0.15 days, on average. AME estimates regarding self-reported measures of general physical health suggest that poorer degrees of physical health are associated with higher absences from work, on average, as anticipated. A one-point increase in the rating of one's own degree of poor physical health is estimated to increase absences by under half a day for men (by a factor of about 0.37) on average, and over half a day (a factor of 0.55) for women in the sample, on average. AME estimates suggest that men and women respond similarly to an additional priority condition diagnosis, on average.

I next report the results of the extended models including interactions between the *keyMhdis* variable and the three other measures of health.

5.2 Mental Health, General Health, and Absenteeism Results

The following serves to provide insight into the relationship between diagnostic status and general well-being. I choose to graphically illustrate these relationships. Each of the figures discussed hereafter plots predicted annual absence values of the respective general health variable, grouped by variable *keyMHdis*, which represents each level at which the estimates are generated for the entire sample. Blue curves illustrate the relationship between a measure of health and predicted absences when *keyMHdis* = 1 and red curves illustrate this relationship when *keyMHdis* = 0. After observing each plot, I estimate the CAME of *diagnosed* mental illness (*keyMHdis*) at three distinct levels of the health variable of interest, which are reported in Table 2, Table 3, and Table 4, along with coefficient estimates of the respective interaction term in the extended model of interest (one of $I1$, $I2$, or $I3$). Pseudo R^2 for the extended model specification and results of a likelihood-ratio test of the null hypothesis that the more restrictive baseline specification produces a better fit than the specification including the interaction term of interest are also reported.

Figures 1 and 2 depict the relationship between general psychological distress (index variable *distress*) and predicted absences for men and women, respectively, by levels of *keyMHdis*. At lower levels of *distress*, men with diagnosed mental illness see higher levels of absence than men without diagnosed mental illness. An intersection between the groups of men with and without a mental disorder diagnosis is observed around a scale score of 11. A score of 11 is close to the upper-bound score of 12 that separates individuals from crossing over from category of "moderate evidence of psychological impairment" to the category of "evidence of severe impairment" (Kessler et al., 2003). At this point of intersection, there is anticipated to be no statistically significant difference in expected absence days between the two groups. From this intersection onward, it appears that at higher levels of distress, men without a diagnosed mental illness are anticipated to exhibit higher levels of absence than men with a diagnosis.

A phenomenon similar to that exhibited for men can be observed in Figure 2 for women. At lower levels of general distress, women display a positive and significant difference in predicted absences between the group reporting diagnosed mental illness and the baseline group of women with no such diagnosis, while higher levels of *distress* suggest no statistically relevant difference in the absence behavior between the group of women with a diagnosed mental illness and no diagnosed mental illness. The coefficient estimate on the interaction term for the woman $I1$ specification is statistically significant at the five-percent level and is reported in Table 2.

CAME estimates are consistent with the figures for both men and women. These estimates are reported in Table 2 along with coefficient estimates for the interaction term $keyMHdis \times distress$. CAME results are consistent with the relationship suggested by Figure 1 – at a score of 5 on the *distress* index, the CAME of *keyMHdis* on expected absences is estimated to be positive and statistically significant; on the other hand, at a *distress* level of 11 which is approximately the point of intersection exhibited in Figure 1, this estimate is not statistically significant, nor is it at the higher index score of 15. Though the sign and magnitude of the coefficient on the interaction term is not directly interpreted here, the estimated degree of statistical significance of this coefficient is of value and illustrates that there is a distinct association between diagnosed mental illness and general psychological distress in determining absenteeism.

Figures 3 and 4 plot predicted absences by a self-reported score of one's own physical health with higher scores indicating poorer health for men and women, respectively. Figure 3 illustrates a similar trend among the group of men with a mental illness in the sample (linear and decreasing) as that observed in Figure 1, though men without a diagnosed mental illness clearly exhibit predicted absences that compile at a faster rate with increasing levels of *physhlth* compared to the rate of change exhibited by this group Figure 1. CAME estimates suggest that the AME of diagnosed mental illness on absenteeism changes sign as self-rated scores of poor mental health increase from lower levels to very high levels. The coefficient estimate for term $keyMHdis \times physhlth$ is significant at the 0.1 percent level for men after clustering on individual and family units. On the other hand, there is no clear difference in the rate of change of predicted absences across levels of physical health for either group of women. Sample women are anticipated to see higher absences at higher degrees of poor health, and this positive relationship grows at a visibly more rapid rate after passing an index score of about six or seven. women without a diagnosed mental illness are predicted to have a lower level of predicted absences across levels of poor health compared to women with a diagnosed mental illness. Confidence bands suggest that there is a positive and significant relationship between mental illness and physical health in predicting expected absences. Interestingly, prior to and after imposing cluster-robust inference upon estimates of the $I2$ specification for women, the interaction term between *keyMHdis* and *physhlth* is not reported to be statistically significant.

Figure 5 plots predicted absences by the count of the number of priority condition diagnoses as defined by the MEPS for sample men. These conditions include chronic heart conditions, high blood pressure, emphysema, asthma, cancer, and other chronic issues. In contrast to Figures 1 and 3, Figure 5 does not exhibit a downward sloping curve for the group of men with a diagnosed mental illness; instead, it appears that this curve may exhibit a slightly positive or near-zero slope on average across all levels of the *prtycnds* variable exhibited by men in the sample. This finding is consistent with the logic surrounding the argument that adverse health effects may pose less harm to the labor market outcomes of individuals who have knowledge of their diagnoses and risks; this is especially true if an individual has a typical healthcare provider, which reduces the search costs associated with improving health through the route of utilizing healthcare. Taking this into account, as well as Figure 6 for women, is an area of interest for future research that addresses these relationship more clearly by modeling structural equations of health rather than the reduced form focused on in this paper.

Noting the steeply increasing rate of change in predicted absences for the group of men without a diagnosed mental illness depicted in Figure 5, one can deduce that predicted absences are less flexible to increases in the count of chronic non-mental conditions for men with a mental illness. Figure 6 plots the predicted absences by condition count for women and depicts a higher positive rate of change per additional diagnosis for women without a diagnosed mental illness, and an intersection in prediction only at a very high number of diagnoses. This may suggest that the higher absence rates of women with mental illness are less responsive to additional condition diagnoses.

6 Concluding Remarks

Taken together, the findings of this study highlight the significant, often hidden, economic impact of undiagnosed moderate-to-severe mental illness on labor productivity. Individuals without a formal diagnosis may lack the resources or awareness to proactively address worsening symptoms, resulting in higher absenteeism and reduced productivity. Conversely, diagnosed individuals, while potentially facing more frequent health shocks, may have greater access to interventions that help mitigate extreme levels of distress and promote workplace attendance. This dynamic underscores the importance of understanding mental health's influence on labor supply, job retention, and match efficiency within the workforce.

Findings suggest that workers with a mental health diagnosis in severe distress and fair- to-low physical health exhibit better productive outcomes compared to similar but undiagnosed peers. This supports the notion that employers can mitigate productivity losses by promoting working wellness and by adopting health policies that encourage early treatment and support for mental health. Notably, AME estimates suggest that absenteeism for women in the sample may be more sensitive to presently observed health shocks than men. This may be due to a multitude of reasons, including differences over preferences for the sexes. Also of note is that when including the interaction of diagnosed mental illness and self-reported physical health status, women do not exhibit clear evidence that a diagnosed mental illness eventually makes them better off at poor levels of self-reported physical health.

The number of chronic physical health conditions is found to significantly influence absenteeism schedules for both men and women when comparing across groups of diagnosed mental illness and no such diagnosis. This finding is consistent with the logic surrounding the argument that adverse health effects may pose less harm to the labor market outcomes of individuals who have knowledge of their diagnoses and risks; this is especially true if an individual has a typical healthcare provider, which reduces the search costs associated with improving health through the route of utilizing healthcare. This is an area of interest for future research that addresses these relationship more clearly by modeling structural equations of health rather than the reduced form focused on in this paper. Of final note is the consideration of whether individuals facing multiple chronic physical conditions on top of mental illness may supply less labor in general or may be on disability leave that technically may not count as short-run absence in the current context, which can be studied in the future.

This paper has found empirical evidence that diagnosed mental illness leads to deviations from optimal labor supply, particularly in the short run, impacting job performance and, consequently, the labor market. Furthermore, individuals with a known diagnosis may face challenges in job matching and retention due to structural biases and employers' concerns over attendance stability. These factors contribute to limited productive capacity within the labor force and inefficient employment outcomes. Future research into the relationship between mental health, productivity, and job stability is critical, especially in today's economic environment, where high rates of mental illness and inadequate access to mental health resources continue to affect millions. A deeper understanding of these dynamics will inform policies and workplace practices that support better health outcomes and foster a more resilient, productive labor force.

References

National Alliance on Mental Illness. (2023, April). Mental Health by the Numbers.
Retrieved from https://www.nami.org/about-mental-illness/mental-health-by-the-numbers.

National Institute of Mental Health. (2024, September). Mental Illness.
Retrieved from https://www.nimh.nih.gov/health/statistics/mental-illness.

World Health Organization. (2024). Mental Health.
Retrieved from https://www.who.int/health-topics/mental-healthtab=tab_1.

Goetzel, R. Z., Ozminkowski, R. J., Sederer, L. I., Mark, T. L. (2002). The Business Case for Quality Mental Health Services: Why Employers Should Care About the Mental Health and Well-Being of Their Employees. *Journal of Occupational and Environmental Medicine, 44*(4), 320 − 330. doi:10.1097/00043764-200204000-00012

Rosenheck, R. A., Druss, B., Stolar, M., Leslie, D., Sledge, W. (1999). Effect of declining mental health service use on employees of a large corporation. *Health Affairs, 18*(5), 193 − 203

Bubonya, M., Cobb-Clark, D. A., Wooden, M. (2017). Mental health and productivity at work: Does what you do matter? *Labour Economics, 46*, 150 − 165. doi:10.1016/j.labeco.2017.05.001

Ashwood, J. S., Briscombe, B., Collins, R. L., C., W. E., Eberhart, N. K., Cerully, J. L., Burnam, M. A. (2016). *Investment in Social Marketing Campaign to Reduce Stigma and Discrimination Associated with Mental Illness Yields Positive Economic Benefits to California*. RAND Corporation.

Cseh, A. (2008). The Effects of Depressive Symptoms on Earnings. *Southern Economic Journal, 75*(2), 383 − 409. www.jstor.org/stable/27751391.

Fletcher, J. (2013). Adolescent Depression and Adult Labor Market Outcomes. *Southern Economic Journal, 80*(1), 26 − 49. doi:10.4284/0038-4038-2011.193

Agency for Healthcare Research and Quality. (n. d.). *MEPS-HC Panel Design and Collection Process*, Rockville, Md.

Long, J. S., Freese, J. (2001). *Regression Models for Categorical Dependent Variables Using STATA*. Stata Corporation.

Braumoeller, Bear F. (2004), Hypothesis Testing and Multiplicative Interaction Terms. *International Organization, 58*, 807 – 820.

Von Korff M, Ormel J, Katon W, Lin EH (1992). Disability and Depression Among High Utilizers of Health Care. *Arch Gen Psychiatry, 52*, 11 – 19.

Goldman, W., McCulloch, J., Sturm, R. (1998). Costs And Use Of Mental Health Services Before And After Managed Care. *Health Affairs, 17*(2), 40 – 52.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal, 75*(299), 493 – 517.

Hamermesh, D. S. (1993). *Labor Demand*. Princeton Press.

Shang, S., Nesson, E., Fan, M. (2017). Interaction Terms in Poisson and Log Linear Regression Models. *Bulletin of Economic Research, 70*(1), 89 – 96.

van Eeuwijk, F. A. (1995). Multiplicative Interaction in Generalized Linear Models. *Biometrics, 51*(3), 1017–1032.

Kohn, R., Saxena, S., Levav, I., Saraceno, B. (2004). The Treatment Gap in Mental Health Care. *Bulletin of the World Health Organization, 82*(11), 858 – 866.

French, M. T., Zarkin, G. A. (1998). Mental Health, Absenteeism and Earnings at a Large Manufacturing Worksite. *The Journal of Mental Health Policy and Economics, 1*(4), 161 – 172.

Cameron, A. C., Trivedi, P. (2004). *Microeconometrics: Methods and Applications.* Cambridge University Press.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data (2nd ed.)*. The MIT Press.

Karaca-Mandic, P., Norton, E. C., Dowd, B. (2012). Interaction Terms in Nonlinear Models. *Health Services Research, 47*(1), 255 – 274.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy, 80*(2), 223 – 255.

Simon, G. E., Barber C., Birnbaum, H. G., Frank, R. G., Greenberg, P. E., Rose, R. M., Wang, P. S., Kessler, R. C. (2001). Depression and Work Productivity: The Comparative Costs of Treatment Versus Nontreatment. *Journal of Occupational and Environmental Medicine, 43*(1), 2 – 9. doi:10.1097/00043764-200101000-00002.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., Zaslavsky, A. M. (2013). Screening for Serious Mental Illness in the General Population, *Archives of General Psychiatry, 60*(2), 84 – 189.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

American Psychiatric Association. (2020). *Stigma, Prejudice, and Discrimination Against People with Mental Illness*. Retrieved from https://www.psychiatry.org/patients-families/stigma-and-discrimination.

Cameron, A. C., Trivedi, P. K. (1986). Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics, 1*(1), 29 – 53. http://www.jstor.org/stable/2096536

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*(1), 69 – 85. https://doi.org/10.2307/1913646

Figures

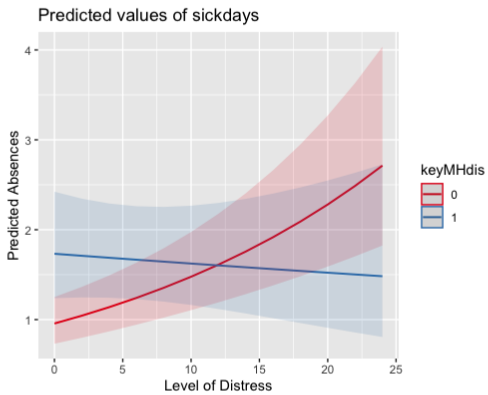

Figure 1: Predicted Absence Given Distress Index and Diagnostic Status, Men

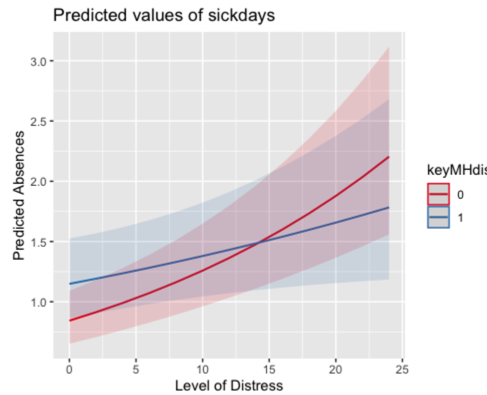

Figure 2: Predicted Absence Given Distress Index and Diagnostic Status, Women

Figure 3: Predicted Absence Given Poor-Health Index and Diagnostic Status, Men

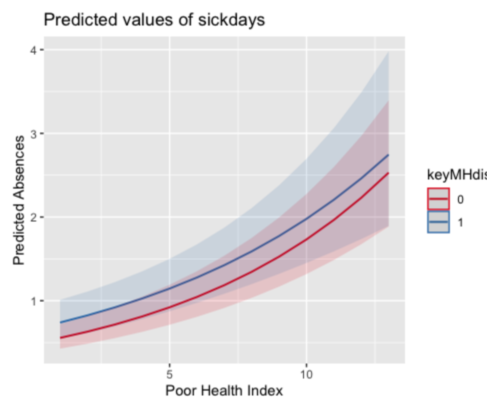

Figure 4: Predicted Absence Given Poor-Health Index and Diagnostic Status, Women

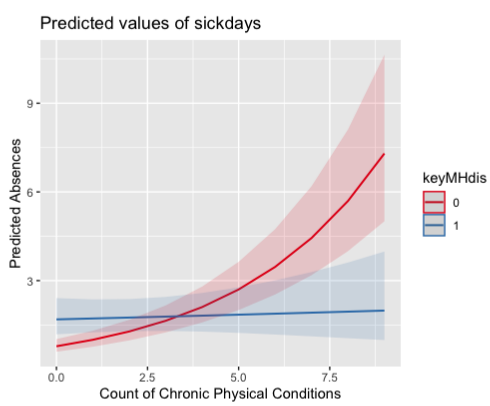

Figure 5: Predicted Absence Given Number of Priority Conditions and Diagnostic Status, Men

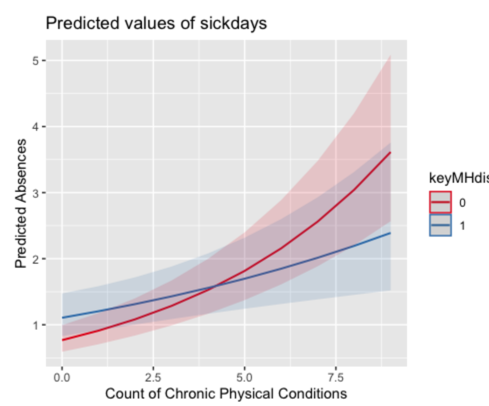

Figure 6: Predicted Absence Given Number of Priority Conditions and Diagnostic Status, Women

Tables

Table 1: Baseline Model AME Estimates

Dependent Variable: sickdays	Men	Women
keyMHdis	1.10 (0.34)***	0.92 (0.31)***
distress	0.11 (0.03)****	0.15 (0.03)****
physhlth	0.37 (0.05)****	0.55 (0.06)****
prtycnds	0.67 (0.09)****	0.69 (0.09)****
Observations	15,713	16,216
Pseudo R^2:	0.164	0.198

Note: *p<0.1; **p<0.05; ***p<0.01; ****p<0.001

Values in parentheses are standard errors computed using the delta method after clustering at individual and family levels.

Table 2: Conditional Average Marginal Effect and Interaction Effects for Model $I1$

Dependent Variable: sickdays Sample	Variable	Conditioning Level	Conditional Average Marginal Effect	Interaction Coefficient
Men	*distress*	5	1.08 (0.35)***	-0.05 (0.02)***
		11	0.18 (0.68)	
		15	-0.70 (1.13)	
	Pseudo R^2: 0.177			
	LR Test (H0: baseline, HA: $I1$): ****			
Women	*distress*	5	0.92 (0.32)***	-0.03 (0.01)**
		15	-0.11 (1.00)	
		20	-0.98 (1.72)	
	Pseudo R^2: 0.208			
	LR Test (H0: baseline, HA: $I1$): ***			

Note: *p<0.1; **p<0.05; ***p<0.01; ****p<0.001

Estimates for males are based on a sample size of $n = 15,713$; estimates for females are based on a sample of size $n = 16,216$. Values in parentheses represent standard errors. CAME standard errors are computed using the delta method after two-way clustering. Standard errors for coefficient estimates are cluster- and heteroskedasticity- robust.

Table 3: Conditional Average Marginal Effect and Interaction Effects for Model $I2$

Dependent Variable: sickdays				
Sample	Variable	Conditioning Level	Conditional Average Marginal Effect	Interaction Coeff.
Men	*physhlth*	3	1.64 (0.38)****	-0.19 (0.04)****
		7	0.21 (0.42)	
		10	-2.11 (1.04)**	
	Pseudo R^2: 0.178			
	LR Test (H0: baseline, HA: $I2$): ****			
Women	*physhlth*	3	0.79 (0.31)**	-0.02 (0.03)
		7	0.94 (0.39)***	
		10	0.98 (0.95)	
	Pseudo R^2: 0.207			
	LR Test (H0: baseline, HA: $I2$): Fail to Reject			

Estimates for males are based on a sample size of $n = 15,713$; estimates for females are based on a sample of size $n = 16,216$. Values in parentheses represent standard errors. CAME standard errors are computed using the delta method after two-way clustering. Standard errors for coefficient estimates are cluster- and heteroskedasticity- robust.

Table 4: Conditional Average Marginal Effect and Interaction Effects for Model $I3$

Dependent Variable: sickdays				
Sample	Variable	Conditioning Level	Conditional Average Marginal Effect	Interaction Coeff.
Men	*prtycnds*	1	1.25 (0.32)****	-0.24 (0.07)****
		3	0.29 (0.44)	
		6	-4.22 (2.06)**	
	Pseudo R^2: 0.178			
	LR Test (H0: baseline, HA: $I3$): ****			
Women	*prtycnds*	1	1.10 (0.31)****	-0.08 (0.04)**
		4	0.12 (0.66)	
		7	-2.39 (2.11)	
	Pseudo R^2: 0.208			
	LR Test (H0: baseline, HA: $I3$): ***			

Note: *p<0.1; **p<0.05; ***p<0.01; ****p<0.001

Estimates for males are based on a sample size of $n = 15,713$; estimates for females are based on a sample of size $n = 16,216$. Values in parentheses represent standard errors. CAME standard errors are computed using the delta method after two-way clustering. Standard errors for coefficient estimates are cluster- and heteroskedasticity- robust.

Appendix

A.1: Total Sample Summary Statistics

Variable	Description	Mean	Min	Max	SD
Dependent Variable (A):					
sickdays	Count of the total days an individual has been absent from work due to physical or mental illness or ailment in the past year.	3.18	0	160	9.97
Mental Health (MH):					
keyMHdis	=1 if individual has a diagnosed mental disorder of interest, =0 otherwise.	0.10	0	1	0.30
distress	Discrete scale from 0 to 24 with higher scores indicating greater emotional distress.	2.60	0	24	3.57
Physical Health (PH):					
physhlth	A discrete scale score with range 1 to 13 with higher scores indicating poorer general health.	4.66	1	13	2.40
prtycnds	The number of priority condition diagnoses.	1.31	0	9	1.45
injury	=1 if individual suffered an injury or illness requiring immediate medical care in the past year.	0.24	0	1	0.43
smoke	=1 if individual smokes cigarettes, =0 otherwise.	0.17	0	1	0.38
exercise	=1 if individual exercises at least 3 times per week, =0 otherwise.	0.53	0	1	0.50
pregnt	Female sub-sample only. =1 if female was pregnant at any point during the year.	0.03	0	1	0.18
Job Traits (J):					
$one4$	Benchmark group. =1 for individuals with tenure between 1 and 4 years, =0 otherwise.	0.37	0	1	0.48
$five14$	=1 for individuals with tenure between 5 and 14 years, =0 otherwise.	0.35	0	1	0.48
$fifteen24$	=1 for individuals with tenure between 15 and 24 years, =0 otherwise.	0.11	0	1	0.32
$25plus$	=1 for individuals with tenure of 25 years or more, =0 otherwise.	0.06	0	1	0.24
temp	=1 if individual has a temporary employment contract, =0 otherwise.	0.05	0	1	0.22
parttime	=1 if individual reports working 35 hours per week or more, =0 otherwise.	0.17	0	1	0.37
Job Traits Continued (J):					
union	=1 if individual is part of a labor union, =0 otherwise.	0.13	0	1	0.33
ssnl	=1 if individual's main job is a seasonal positions, =0 otherwise.	0.05	0	1	0.21
pubsect	=1 if individual works in the public sector, =0 otherwise.	0.18	0	1	0.38
$ind1$	Natural Resources and mining.	0.02	0	1	0.14
$ind2$	Benchmark group. Construction and manufacturing.	0.18	0	1	0.38
$ind3$	Wholesale and retail trade.	0.13	0	1	0.34
$ind4$	Transportation and utilities.	0.05	0	1	0.22
$ind5$	Information.	0.02	0	1	0.14
$ind6$	Financial activities.	0.06	0	1	0.24
$ind7$	Professional and business services.	0.11	0	1	0.31

Table 1 Continued					
Variable	Description	Mean	Min	Max	SD
*ind*8	Education, health, and social services.	0.25	0	1	0.43
*ind*9	Leisure and hospitality.	0.07	0	1	0.28
*ind*10	Other services.	0.04	0	1	0.20
*ind*11	Public administration.	0.06	0	1	0.24
*occ*1	Management, business, and financial operations.	0.13	0	1	0.33
*occ*2	Professional and related occupations.	0.22	0	1	0.41
*occ*3	Service occupations.	0.19	0	1	0.39
*occ*4	Sales and related occupations.	0.08	0	1	0.27
*occ*5	Office and administrative support.	0.14	0	1	0.35
*occ*6	Farming, fishing, and forestry	0.01	0	1	0.10
*occ*7	Construction, extraction, and maintenance.	0.08	0	1	0.27
*occ*8	Benchmark group. Production, transportation, material moving.	0.15	0	1	0.36
*occ*9	Unclassifiable occupation.	0.004	0	1	0.06
1*to*19	=1 if firm has between 1 and 19 workers.	0.30	0	1	0.46
20*to*99	=1 if firm has between 20 and 99 workers.	0.33	0	1	0.47
100*to*499	=1 if firm has between 100 and 499 workers.	0.21	0	1	0.41
500*plus*	Benchmark group. =1 if firm has 500 or more workers.	0.16	0	1	0.37
NR.numemp	=1 if individual did not respond to questions pertaining to firm size.	0.05	0	1	0.22
Demographics (X):					
poor	=1 if household income as a % of poverty line puts them into “poor” or “near poor” groups.	0.14	0	1	0.34
lowinc	=1 if household income as a % of poverty line puts them into “low income” group.	0.16	0	1	0.36
midinc	=1 if household income as a % of poverty line puts them into “middle income” group.	0.34	0	1	0.47
highinc	Benchmark group. =1 if household income as a % of poverty line puts them into “high income” group.	0.37	0	1	0.48
married	=1 if individual is married, =0 otherwise.	0.55	0	1	0.50
famsz	Number of individuals within the surveyed household.	3.02	0	14	1.66
yngchldrn	Number of children aged 6 and under.	0.38	0	6	0.71
age	Age in years.	41.86	18	84	11.82
belowhs	Individuals with less than high school education.	0.13	0	1	0.34
hsdeg	Benchmark group. Individuals with a high school degree or GED.	0.32	0	1	0.47
somecoll	Individuals with some college or an associate’s degree, but no 4-year degree.	0.25	0	1	0.43
bachdeg	Individuals with a bachelor’s degree.	0.20	0	1	0.40
bachplus	Individuals with schooling beyond a bachelor’s degree.	0.10	0	1	0.31
hispanic	=1 if individual is Hispanic, =0 otherwise	0.28	0	1	0.45
black	=1 if individual is Black, =0 otherwise	0.18	0	1	0.38
asian	=1 if individual is Asian, =0 otherwise	0.08	0	1	0.28
bornus	=1 if individual was born in the US.	0.65	0	1	0.48
Other Controls (C):					
unemp.1	=1 if individual reports unemployment at exactly one of the three interviews.	0.07	0	1	0.26
unemp.2	=1 if individual reports unemployment at exactly two of the three interviews.	0.06	0	1	0.23

Table 1 Continued					
Variable	Description	Mean	Min	Max	SD
Other Controls Continued (C):					
$partialemp$	= 1 if an individual is unemployed at one of the three interviews, but worked for at least part of the reference period prior to employment termination.	0.04	0	1	0.20
$empUB$	The total number of days for which an individual was employed during the year.	338	28	365	0.19
$moved\_US$	=1 if individual moved within the US during the calendar year.	0.03	0	1	0.18
$moved\_RU$	=1 if individual joined a new reference unit (household) at any point during the calendar year.	0.01	0	1	0.10
$refpers$	=1 if answering on own behalf.	0.62	0	1	0.49
NE	=1 if individual resides in the Northeastern region for most of the year.	0.16	0	1	0.36
MW	=1 if individual resides in the Midwestern region for most of the year.	0.20	0	1	0.40
W	=1 if individual resides in the Western region for most of the year.	0.28	0	1	0.45
S	Benchmark group. =1 if individual resides in the Southern region for most of the year.	0.37	0	1	0.48
$unemp.rt$	The average annual unemployment rate faced by the individual estimated using monthly regional unemployment reports from the Bureau of Labor Statistics and region of residence reported for each round.	8.05	5.5	10.9	1.29
$yearone$	=1 if individual is only observed for the first year of their designated panel.	0.26	0	1	0.44
$yeartwo$	=1 if individual is only observed for the second year of their designated panel.	0.20	0	1	0.40
$bothyears$	Benchmark. =1 if individual is observed for both years of their panel.	0.54	0	1	0.50

Note: Sample means are derived using a sample of 31,929 observations. It should be noted that statistics reported in Table 1 neglect individual-specific components of the data although there are individuals in the sample observed twice.