Abstract

Along with having a significant impact on a person's quality of life, depression, excessive anxiety, and other mental health disorders could have similar effects on workers in their respective workplaces. Using firm-level surveys on employee mental health and productivity at management consulting firms, our paper estimates the impact of mental health on worker productivity. We utilize established clinical methods to measure mental health and use time-logs and manager satisfaction as productivity measures. We combine the survey data with an instrumental variable approach, using family bereavement, to estimate causal effects.

Introduction

An increasing awareness of mental health as of recent has led to studies analyzing the effects of mental health disorders in patient populations and nationwide surveys to measure quality of life. However, the same curiosity has not reached workplaces and the employees that occupy them. Given the past literature on the effects of mental health on the unprofessional life, questions about its impact on productivity and worker output arise. Namely, how do changes in anxious or depressive symptoms impact firm output?

Studies have shown that poor health, whether mental or physical, leads individuals to take more time off work, a phenomenon known as absenteeism, but also to be less productive on the job, known as presenteeism. These findings suggest that mental health impacts worker output through both of these channels and the main goal of our paper is to quantify the extent of this impact.

The question holds value as it coincides with the population's already hyper-attentiveness to the concept and practicality of wellbeing. With research such as ours, findings can provide guidance to reform workplace policies and benefits to better support employees suffering with mental health disorders. This can allow companies to implement evidence-based remedies to improve the quality of life of employees, lead to reduced absenteeism, disability and lost productivity, and make the preservation of wellbeing a priority in workplace management. The methodology for gathering the data, measuring mental health and productivity, and analyzing the findings provide the validity of the project.

We determined the professional services industry to be the ideal sample population for this study since the job is often high-stress and requires consistent high-performance. Since both work inputs and outputs are intangible there, it is usually difficult to measure productivity. To avoid this measurement problem, this paper obtains productivity data from management consulting firms because, as standard practice, they measure a wide range of required metrics with precision, and at regular intervals.

Our explanatory variable of interest is worker mental health. Though measuring mental health and well-being is a difficult task, the current literature provides a variety of surveys to choose from. This study utilizes the 12-item General Health Questionnaire to generate a mental health score ranging from 0 to 36, according to which a higher score corresponds to higher levels of psychological distress.

The independent variable, productivity, is measured using weekly employee time logs to compute both hours worked and a monetary measure of productivity weighting hours worked by dollar values corresponding to each type of work, e.g. client/internal, as determined by billing rates or fair value estimates. These metrics are complemented by monthly qualitative supervisor ratings on four axes, work quality, client engagement, fulfillment of firm initiatives, and employee output versus expected output, which are additional outcome variables of interest.

To establish a causal relationship between mental health and productivity, we first run a fixed effect model controlling for physical health as measured by the Work Limitations Questionnaire, individual and manager characteristics. Additionally, we supplement this analysis with an instrumental variable approach, using the loss of a loved one as an instrument for mental health. This approach is used as a way to avoid reverse causality and omitted variable bias impacting our estimate of the effect of mental health on productivity.

Literature Review

Happiness and Productivity. The happy/productive worker hypothesis, which is the idea that happy workers are more productive than unhappy ones, is a well-known concept in industrial/organizational psychology, but the different definitions of happiness used across the literature lead to ambiguous findings as noted by Cropanzano and Wright (2001). Using job satisfaction as a proxy for happiness, Argyle (1989) shows that there is a positive correlation between job satisfaction and work performance, particularly among white collar workers, but also notes that little is known about the link between life satisfaction and productivity. Using multiple measures of workers' affect, Wright and Staw (1999) find mixed results when studying the relationship between workers' well-being and their supervisors' performance ratings of them. It is also important to note that most of the psychology literature does not attempt to understand the causal mechanisms behind the relationship between well-being and productivity, and few economics papers have tackled the question.

The paper that is closest to our study, Oswald, Proto, and Sgroi (2015), estimates the causal impact of happiness on productivity in a lab setting. In their first experiment, the researchers randomly assign subjects to either a happiness-enhancing treatment, for which subjects are shown a short comedy clip, or a control treatment, in which they are shown a placebo clip or not shown a clip. After the treatment, they are asked to complete a simple math task under time

pressure with pay based on performance. In their second experiment, the researchers replicate the same experimental design but instead of inducing happiness, they ask subjects to report real-world shocks they have experienced in the last two years, such as family bereavement or serious illness, and compare performance between those who have experienced such shocks and those who have not. The researchers find that the happiness treatment increases the number of correct additions by 12\% and experiencing a bad life event lowers performance on the additions task by approximately 10\%. Our paper extends this analysis to a real-world setting to study whether mental health has an impact on productivity at work.

Productivity. So far, productivity has usually been measured by two principal metrics, absenteeism and presenteeism. Absenteeism, when an employee avoids work entirely, is relatively easy to identify. Kessler et al. (2003) measure absenteeism using the National Comorbidity Survey and find that 59\% of adults who had major depressive disorder with lifetime prevalence were heavily impaired in their ability to perform social roles and missed an average of 35 workdays. Presenteeism, defined as a worker being physically present but not producing at expected/peak output, is more difficult to quantify. Coviello et al. (2017) and Rothbard and Wilk (2011) use phone call volume as a proxy for productivity in call center workers, finding a statistically significant difference and positive correlation with better mood. However, Coviello et al. (2017) notes that a quantity-based approach, while convenient from the data collection perspective, is not ideal since it cannot account for quality.

Rather than using proxy variables, some researchers opt for surveys designed to measure presenteeism. Schultz and Edington (2007) conducted a review of common surveys available at the time, identifying the Work Limitations Questionnaire (WLQ) as a strong instrument. We include this survey in our study design as a control measuring physical wellness. Some other approaches to measure productivity include creativity as defined by peers (Amabile et al. 2005), a composite measure of actual hours worked weighted by self-perceived job performance (Bee et al. 2010), and "financial measures such as revenue or sales per person, growth in revenue or sales over time, quantity per time period, enrollments in programs, labor hours, costs to the budget, cross-sells, or performance ratings" (Krekel et al., 2019).

Mental Health. While the relationship between mental health and temporal focus is well-established in psychological research, fewer economic studies analyze the direct link between mental health and time preferences. Bayer et al. (2019) examine the relationship between clinical depression and time and risk preferences and consumption decisions. They find that individuals with increasingly severe depression are significantly more risk-seeking and present-biased and make less efficient consumption choices than those without depression. This suggests that a potential mechanism through which mental health impacts productivity is through its effect on temporal focus.

Another channel could be through mental health's impact on self-control. Both Bernheim, Ray, and Yeltekin (2015) and Schilbach (2019) indirectly help to establish crucial links between decision-making and economic outcomes, through analyzing the relationship between self-control and savings behaviors. Using a standard model of intertemporal allocation with individuals who exhibit present-bias, Bernheim, Ray, and Yeltekin (2015) show that poverty undermines the ability to exercise self-control, making wealth accumulation impossible below a certain asset threshold. Schilbach (2019) shows that, among low-income cycle-rickshaw drivers in India, receiving sobriety incentives that decreased daytime drinking raised savings by 50 percent. These papers show that an individual's capacity to exhibit self-control has an important impact on their economic outcomes, suggesting that mental health, through its potential effect on self-control, may have similar economic impacts.

Due to growing research interest, medical and economic literature are abundant with methods to measure mental health. A large portion of studies like Chatterji, Alegria, and Takeuchi (2011) use pre-existing government survey data, consisting of health data and professionally verified mental illness diagnoses, to conduct nation-wide longitudinal cohort studies. In smaller settings, other researchers like Clark et al. (2020) use questionnaires or inventories developed to quickly assess (not diagnose) mental illnesses in study participants. One such questionnaire is the General Health Questionnaire, introduced in 1988, which will be described in more detail in the Study Design section.

Contribution to the Literature. We make three main contributions to the literature. Firstly, we study the link between mental health and productivity in a novel environment. Second, we account for wider definitions of productivity. Indeed, previous studies focused on quantitative measures whereas we supplement our hours worked and monetary measures of productivity with a qualitative measure through manager satisfaction surveys. We are also the first study, to our knowledge, to use an instrumental variables approach in a real-world setting to estimate the causal impact of mental health on productivity, instead of implementing a correlational study like most of the previous literature.

Study Design

In this section, we present the study design. First, we present our model to establish the channels through which mental health impacts productivity. Then, we detail our data collection process by explaining our research population, our measures of productivity and of mental health and how we construct our control variables. Finally, we elaborate on our estimation methods using both a fixed effect model to establish the correlational relationship between mental health and productivity and an instrumental variable approach to estimate the causal impact.[1]

[1] This paragraph added in response to a peer review comment asking us to introduce the purpose and relevance of each section.

1. Model

[see presentation]

2. Data

2.1. Research Population

For the purpose of this study, the research population should be of a similar job type, have similar deliverables and responsibilities to control for various definitions of productivity and remove variation in mental health effects due to job type. For this reason, we decided to perform the research on Bain associates in the US since this population gives us a large enough sample for research to be conducted. We chose to only study employees at the associate level to lessen productivity effects from varying experience on the job and selection effects based on performance at higher levels of management. All employees are sampled in the US to avoid comparing payments across different countries and according to recruiting employee metrics, Bain associates travel significantly less than associates from the other MBB/Big Four consulting companies. This means that this population is also less likely to be subject to the mental health effects of traveling occupations. So, our sample consists of all Bain associates across the 11 US offices, leading to a sample size of approximately 1,000 employees.[2]

Our study will take place over a six-month period, and we will collect the mental health and productivity data on a biweekly basis. The mental health data will come from a questionnaire administered via Qualtrics, an online survey platform, to all associates, and we will implement the questionnaire as a part of a partnership with the company, who will set up incentives for their employees to fill it out consistently.[3] All associates will be required to fill out the mental health questionnaire in order to access their central HR program used to enter time logs. We also integrate mechanisms in the mental health questionnaire to encourage truthful responses and verify attentiveness. Both the company and its employees also have an incentive to take this study seriously to improve workplace practices and bolster employee productivity.[4]

2.3. Productivity

[2] Details added in response to peer review comments about unspecified population characteristics. The original design targeted general management consulting groups in the United States. After discussion, we narrowed it down to Bain and Co., for the reasons mentioned above, and estimated that there are approximately 1,000 associate consultants available for the study based on recruiting statistics.

[3] REDCap was initially proposed as the survey platform due to its widespread use in clinical and health research. However, after comparing the features of REDCap and Qualtrics, we decided that Qualtrics best suited our needs.

[4] Incentives are explored further in the Mental Health section.

In this study, we use three measures of productivity. The first is hours worked and is used to study the effect of mental health on absenteeism. The second is based on the monetary value of the tasks performed by employees to account for the fact that an hour of work has varying values to the firm depending on what types of tasks are performed during that time. The third is based on manager satisfaction and allows us to study the impact of mental health on other dimensions, including the capacity to develop work relations and meet expectations. Though it is usually difficult to measure productivity, management consulting firms measure a wide range of required metrics with precision and at regular intervals, making it possible for us to observe the tasks performed by employees, the time spent on them and the monetary value to the firm.

The most important source of data we use are employee time logs, submitted at the end of each week. Companies currently use specialized web tools to track the actual hours worked for each professional, and further classify these into types of work, for example by client or internal task. Reporting accuracy is ensured by stringent human resource supervision, with penalties up to dismissal imposed for false reporting, either by employees on their own volition or for supervisors pressuring them to do so. Hours worked are directly taken from the time-logs and our monetary productivity measure is constructed using the dollar values corresponding to each hour category as weights, such that:

$$Q_{it} = \mathbf{\Sigma} \text{ (Hours Worked}_{it,p} * \text{Hourly Value}_{p}).$$

where $HoursWorked_{itp}$ is the number of hours employee i spent on project p during period t and $HourlyValue_p$ is the monetary value to the firm of spending one hour on project p. The hourly value of a project is determined by how much the consulting firm is charging clients for an hour of work according to their terms of engagement. For internal tasks, like firm initiatives or administrative work, the hourly value is a fair value estimation of the cost of outsourcing each type of job. An example of what a given worker's hourly log from our dataset may look like is provided in Table X.X1[5] below.

Study Week	Logging Category	Hours	Rate/Hour	Productivity (II)
6	Client A	20	$120	$2400
6	Client B	10	$165	$1650
6	Administration	10	$45	$450
6	Selling	15	$70	$1050

[5] Example added in response to peer feedback on clarity of productivity measures. The table helps readers visualize time log and monetary productivity data.

6	Firm Initiatives	5	$45	$225
6	Total	60	—	$5775

Table X.X1: Sample Weekly Log Data, Single Employee

Supplementary data, in the form of supervisor evaluations are collected as additional, qualitative metrics and are a good complement to our other productivity measures as they serve as a non-monetary index of worker value. Ratings are determined on a five point scale for work quality, client engagement, fulfillment of firm initiatives, and employee output versus expected output; this is done by averaging Likert responses to the evaluation questions provided in Table X.X2, using the format in Table X.X3 (these are a combination of specific Bain benchmarks and other, industry standard, criteria).[6]

Though normally filled out quarterly or at the end of each financial year, we would collect these data monthly for increased granularity. We are confident that the specific criteria detailed here can be answered meaningfully despite the greater interval, since they rely on employee deliverables/interactions with a relatively short periodicity. For example, corrections required or peer effects can be gauged per draft of a piece of work (~1 week turnaround); satisfaction can be estimated by managers based on their client interactions, which occur as standard practice, and at minimum, once a month on every account. Because the survey consists of nine very short rating questions, we also believe that it will not significantly increase workload, even if supervisors may have to fill out multiple for multiple managed associates.[7]

The manager rating is constructed as follows:

$$\text{Rating}_{it} = (\text{Rating}_{it,\text{quality}} + \text{Rating}_{it,\text{ engagement}} + \text{Rating}_{it,\text{initiatives}} + \text{Rating}_{it,\text{expectation}})/4$$

Ratings are averaged across five of each employee's direct and indirect managers, to normalize idiosyncrasies, prior to being used to calculate the final index score.

Rating Category	Evaluation Criteria
Work Quality	\% Consumed / \% Complete: For a given project, a comparison of the total client approved billable hours the employee has

[6] Example added in response to peer feedback on clarity of productivity measures. The specific evaluation criteria and method used to construct the ratings productivity measure have now been detailed. We also answer a question on the origin of the system.

[7] Ratings categories and criteria were updated to address Stage 5 concerns about the validity (with respect to effort required for/meaningfulness) of more frequent evaluations. An explanatory paragraph was also added to justify the survey method.

	consumed versus the amount of project work that has been completed against this. [Greatly Over Consumed : Greatly Under Consumed] Corrections Required: A count of obvious errors that a manager has had to point out to the employee. [Great Many : None] Proactivity: How willing is an employee to self-teach? How often do they wait for direction? [Very Unwilling : Very Willing]
Client Engagement	Average Sales / Client: What is the average dollar value of new client sales attributable to an employee? [Benchmarked Very Low : Benchmarked Very High] Client Satisfaction: How satisfied are clients with the employee's project team? [Very Unsatisfied: Very Satisfied] Client Retention \%: How many of an employee's clients choose to continue with the firm or project team for future engagements. [Benchmarked Very Low : Benchmarked Very High] Revenue Enhancement: Has the employee been able to sell new products on existing engagements? [Benchmarked Very Low : Benchmarked Very High]
Firm Initiatives	Leadership: How involved is the employee in future planning for the firm? Do they actively think about and work towards accomplishing team goals? [Very Uninvolved : Very Involved] Peer Effects: What are the employee's effects on peers, e.g. cohesion/culture/motivation? Net positive, negative, or neutral? [Very Negative : Very Positive]
Actual vs. Expected Output	Hours: How many hours a week is the employee working? How do they benchmark against peers? Is the number at least 40

	hours/week? [Benchmarked Very Low : Benchmarked Very High] Staffing: Are staffers satisfied with the employees availability to take on new work when they are believed to have spare capacity? [Very Unsatisfied: Very Satisfied] Work Mix: Is the worker's productivity-II dollar value in line with the firm's expectations for them? Is the employee over-, under- or valued at par? [Very Overvalued: Very Undervalued]

Table X.X2: Sample Evaluation Criteria by Rating Category

Evaluation Criteria	1/5	2/5	3/5	4/5	5/5
WQ: Consumed / Complete	Greatly Over Consumed	Over Consumed	On Par	Under Consumed	Greatly Under Consumed
WQ: Corrections	Great Many	Many	An Average Amount	Few	None
WQ: Proactivity	Very Unwilling	Unwilling	Neutral	Willing	Very Willing
CE: Sales	Benchmarked Very Low	Benchmarked Low	Benchmarked Average	Benchmarked High	Benchmarked Very High
CE: Client Satisfaction	Very Unsatisfied	Unsatisfied	Neutral	Satisfied	Very Unsatisfied
CE: Client Retention	Benchmarked Very Low	Benchmarked Low	Benchmarked Average	Benchmarked High	Benchmarked Very High
CE: Revenue Enhancement	Benchmarked Very Low	Benchmarked Low	Benchmarked Average	Benchmarked High	Benchmarked Very High
FI: Leadership	Very Uninvolved	Uninvolved	Neutral	Involved	Very Involved
FI: Peer Effects	Very Negative	Negative	Neutral	Positive	Very Positive

AvE: Hours	Benchmarked Very Low	Benchmarked Low	Benchmarked Average	Benchmarked High	Benchmarked Very High
AvE: Staffing	Very Unsatisfied	Unsatisfied	Neutral	Satisfied	Very Unsatisfied
AvE: Work Mix	Very Overvalued	Overvalued	On Par	Undervalued	Very Undervalued

Table X.X3: Likert Scoring by Evaluation Criteria

2.2. Mental Health

To measure employee mental health, we implement an established mental health questionnaire from the psychology literature. During the six-month study period, participants will be required to answer the Qualtrics survey every other week to access key features relevant to daily work, such as entering time logs or viewing workflows.[8] This method is based on Coviello et al.'s (2017) method of preventing forgetfulness. Considering the short length of the survey, it should take participants no more than three minutes to respond. The days are randomized to eliminate the possibility that employees will know when to expect the survey and plan their responses ahead of time. Employees will be asked to digitally agree to a statement of truthfulness at the beginning of the survey, similar to an honor code. Additionally, we use Qualtrics for three reasons: security, anonymity, and randomization. Employee mental health is sensitive information that must be treated with care, and Qualtrics is in compliance with various security standards, including those for HIPAA (relevant for patient privacy and health information). Qualtrics allows researchers to assign numerical IDs and create personal links for each employee, ensuring that employers will not be able to identify individuals by name if they were to look at the data. This anonymity seeks to reduce incentives to lie about mental health out of fear of punishment or discrimination by the firm.[9] Finally, Qualtrics's ability to randomize questions is useful for our study; in randomizing the main questions of the survey, it reduces the possibility of employees creating a standard pattern of responses. An attentiveness question will

[8] We received mixed reactions regarding our initial decision to conduct surveys only on Wednesdays to control for potential ways the day of the week could affect mental health. We eventually decided that it would be best to randomize the day of week to prevent people from expecting the survey and preparing false responses ahead of time. Answers are likely to be more candid when people are not expecting to take the survey. We also changed it so that employees needed to take the survey every other week instead of every week to reduce the possibility that they will be annoyed by the survey.

[9] Much consideration went into deciding on participation and truthfulness incentives. Participation can be enforced by the company. After reviewing methodologies in the literature, we decided to adopt Coviello et al.'s method to avoid forgetfulness. Regarding truthfulness, we thought the only reason a person would lie would be due to fear of that information being used against them. Therefore, we also chose a survey platform (Qualtrics) that would allow researchers to match data to an individual while keeping that individual's identity anonymous.

also be included. The prompt will say, “Respond ‘Same as usual’ if you have read this question,” allowing us to filter out surveys that were completed in bad faith.

The main survey consists of the 12-item General Health Questionnaire (GHQ-12).[10] This questionnaire was first proposed as a 60-item survey in 1970 by Goldberg et al. The most popular version used by researchers today is the 12-item version; although brief, it provides data consistent with longer versions, as tested by the World Health Organization (Goldberg et al. 1997). Since its introduction, the GHQ-12 has been translated and validated in 42 different languages and cultures (Hystad and Johnsen 2020). The simplicity of the questions prevents survey fatigue, and two scoring structures (clinical and Likert) allow researchers to account for both the presence and level of psychological distress.

The GHQ-12 consists of the following questions and answers:

In the past two weeks, have you...

1. Been able to concentrate on what you’re doing?	Better than usual	Same as usual	Less than usual	Much less than usual
2. Lost much sleep over worry?	Not at all	No more than usual	Rather more than usual	Much more than usual
3. Felt you were playing a useful part in things?	More so than usual	Same as usual	Less useful than usual	Much less useful
4. Felt capable of making decisions about things?	More so than usual	Same as usual	Less so than usual	Much less capable
5. Felt constantly under strain?	Not at all	No more than usual	Rather more than usual	Much more than usual
6. Felt you couldn’t overcome your difficulties?	Not at all	No more than usual	Rather more than usual	Much more than usual
7. Been able to enjoy your normal day-to-day activities?	More so than usual	Same as usual	Less so than usual	Much less than usual
8. Been able to face up to your problems?	More so than usual	Same as usual	Less so than usual	Much less able

[10] The original study design proposed the use of two different mental health surveys: one to measure wellbeing and another to assess mental health using the Diagnostic and Statistical Manual of Mental Disorders. Upon additional literature review, we discovered the GHQ-12 and believed it to be more effective in capturing the two original aspects. The short length is also a plus.

9. Been feeling unhappy and depressed?	Not at all	No more than usual	Rather more than usual	Much more than usual
10. Been losing confidence in yourself?	Not at all	No more than usual	Rather more than usual	Much more than usual
11. Been thinking of yourself as a worthless person?	Not at all	No more than usual	Rather more than usual	Much more than usual
12. Been feeling reasonably happy, all things considered	More so than usual	Same as usual	Less so than usual	Much less than usual

For this study, we utilize Likert scoring, which assigns values of 0-1-2-3 to responses; a higher score corresponds to higher levels of distress. Studies applied to screening for anxiety and mood disorders find that the threshold for a highly likely presence of a mental disorder is typically ⅔ of the maximum possible score. The Likert score for each participant will be used directly in our analysis in section 3 as our measure of mental health.

2.4. Control variables

In our analysis, we control for physical health, individual characteristics, and office location, individual, manager and time fixed effects. At the beginning of our study, we will ask HR to share an anonymized list of all their associates by location matched to 5 direct and indirect managers. Each associate and manager will be identified by a unique ID number and the map of IDs to actual employees will be unknown to ensure anonymity.

To collect observable characteristics on each employee, we will administer a preliminary survey including age, gender, race, years of experience in management consulting, and if they have experienced the loss of a loved one in the past two years. These characteristics will be used as the individual-level controls in the regressions.

In our analysis, we also control for an employee's physical health given the impact it can have on both mental health and productivity. To account for the relationship between physical health and productivity, we use the Work Limitation Questionnaire (WLQ). Schultz and Edington (2007) recommends this survey on the basis of its reliability, validity, and applicability to various work environments and health risks/conditions. The purpose of the WLQ is to assess the impacts of chronic health conditions on one's ability to work. We use the six questions pertaining specifically to physical health (the Physical Demands scale). Sample items include the following:

In the past month, how much of the time were you **able to do** the following without difficulty caused by physical health problems?

Walk or move around different work locations (for example, go to meetings)					
All of the time (100\%)	Most of the time	Some of the time (About 50\%)	A slight bit of the time	None of the time (0\%)	Does not apply to my job
Use hand-held tools or equipment (for example, a phone, pen, keyboard, computer mouse, drill, hairdryer, or sander)					
All of the time (100\%)	Most of the time	Some of the time (About 50\%)	A slight bit of the time	None of the time (0\%)	Does not apply to my job

The answers from "All of the time" to "None of the time" are scored from 1 to 5 while "Does not apply" receives a score of 0. If all six items are applicable to the participant, then scores range from 6 to 30; a higher score corresponds to greater levels of physical limitations to productivity.

3. Estimation Methods

Estimation models

1. **OLS estimates**

$$Productivity_{it} = \beta_0 + \beta_1 MH_{it} + \beta_2 PH_{it} + X_i + I_i + M_i + O_i + T_t + \epsilon_{it}$$

- MHit: mental health as measured by the GHQ-12 survey
- PHit: physical health as measured by the WLQ T
- Xi: vector of individual characteristics including years of education, years of relevant work experience, age, sex and race
- Ii individual fixed effects
- Oi: office location fixed effects
- Mi: manager fixed effects
- Tt: time fixed effects

Estimation models

2. IV estimates

First stage:

$$MH_{it} = \gamma_0 + \gamma_1 L_i + \gamma_2 PH_{it} + X_i + I_i + M_i + O_i + T_t + \mu_{it}$$

Second stage:

$$Productivity_{it} = \alpha_0 + \alpha_1 \widehat{MH_{it}} + \alpha_2 PH_{it} + X_i + L_i + I_i + M_i + O_i + T_t + \nu_{it}$$

Conclusion

1. Findings

We expect β1 from both the simple OLS and IV regressions to be positive, in line with the results of previous studies in the literature. We hypothesize that these results arise through two main mechanisms: (i) mental health distorts the worker's costs of/utility from work and leisure and by consequence, impacts the worker's allocation of time at work, and (ii) when at work, mental health and well-being impact the amount of mental resources available to do one's job, affecting quality.

We further disaggregate our expected findings by our three different measures of productivity:

i. Hours Worked

   a. Poor mental health increases the cost of going to work relative to staying home and given a worker's time constraint, this could decrease the total number of hours worked via substitution effects.

   b. On the contrary, we may see an increase in hours worked if employees are using their jobs as a coping strategy for poor mental health, for example one may spend

more time being productive if they do not want to think about negative events, like the death of a loved one. This can be said to increase the cost of leisure and cause substitution towards work.

ii. Monetary Value

a. Poor mental health reduces the amount of mental resources and energy left to do work because a larger portion of an individual's mental endowment is spent worrying or thinking about non-work related matters. Employees may also spend a greater proportion of their time on tasks that earn less money for the firm but are easier to do, causing a decline in their computed monetary value of work. Conversely, better mental health could increase the completion of high-value tasks, through the same mechanism.

iii. Manager Satisfaction

a. Work Quality: Mental health may affect the costs of more difficult tasks. Even if employees are working the same amount of hours, and have the same relative proportions of tasks, those with poor mental health may be able to accomplish less in an hour, as each unit of output costs them more to produce. These employees may also make more mistakes, and be less willing or have less time to self-teach, which would decrease their work quality score.

b. Client Engagement/Firm Initiatives: Those with poor mental health may be less able to make connections with co-workers, less willing to take on new projects, less engaged with clients, and less proactive overall, leading to lower scores on both client engagement and firm initiatives.

c. Actual vs. Expected Output: If mental health also distorts time preferences, individuals with poor mental health may procrastinate more or even not be able meet expectations as well as individuals without mental health problems.

2. Limitations

(1) The sample of US workers used in this study, management consultants, may have characteristics that differentiate them from the overall population.

It can be argued that people who self-select into consulting may be more adept at managing stress, since they willingly chose a profession that has a low amount of flexibility on work output. Though this fact, if true, would not directionally invalidate the

results of our paper (since it is unlikely that the sample would garner any benefits from being more stressed/anxious/depressed/et cetera), it may reduce the magnitude of mental health's effect on productivity, and could be something to keep in mind when considering the results' broader applications.[11]

(2) The measure of mental health depends on self-reported information, which may not accurately represent the respondents' true state of mind.

Although we include measures to improve truthful reporting, there is no guarantee that employees will respond truthfully 100\% of the time, whether it be due to a busy schedule and haphazard responding, an inability to fully gauge one's mental state, or the treatment of mental health topics as taboo. This is a common concern regarding mental health and wellbeing studies that do not use formally verified psychiatric evaluations. However, considering how it would be ethically and financially inappropriate to professionally evaluate each employee's mental health, the best we can do is use empirically-validated survey methods. Future research can look into other widely-used methods of measuring mental health if they deem the GHQ-12 to be inadequate.

3. Further Research

To preserve the simplicity of our model, we include the controls we believe to be most relevant to understanding the relationship between mental health and work productivity. The possible directions for future research and analysis are endless. At the individual level, researchers could explore how characteristics like gender and socioeconomic background influence one's resilience and ability to be productive. Whereas our study seeks to illustrate the relationship between productivity and mental health, additional research would be needed to identify the specific mechanisms through which that relationship occurs. This could take the form of a separate, preliminary experiment screening for present bias and self-control tendencies. At the work environment level, it would be interesting to separate the study population into subgroups such as low- and high-performing individuals and see if regression analyses produced significantly different estimates for each subgroup. Additionally, this type of study would need to be replicated across different industries and employment types because of the bias described in the Limitations section. Research across different industries would contribute to the larger discussion of theories like the job-specific stress theory. Finally, at the larger socio-cultural level, future models could account for family values, personal upbringing, and role model effects to explore how environmental factors during development influence one's relationship between work and mental health. Replicating this study in different countries or cultural contexts could also provide insight into how productivity is affected by the greater societal perception of mental-health-related issues. Further research into any of these directions could result in valuable

[11] In response to stage 4 peer feedback about the appropriateness of the choice of professional group due to potential self-selection bias.

implications for multiple settings, be it within a single firm or in the context of a country's work culture, and contribute to the larger conversation concerning the importance of one's mental health.

Reference List

Amabile, Teresa M., Sigal G. Barsade, Jennifer S. Mueller, and Barry M. Staw. 2005. "Affect and Creativity at Work." *Administrative Science Quarterly* 50, no. 3: 367-403. https://doi.org/10.2189%2Fasqu.2005.50.3.367

Argyle, Michael. 1989. "Do Happy Workers Work Harder?" In *How Harmful is Happiness? Consequences of Enjoying Life or Not*, edited by Ruut Veenhoven. Rotterdam, The Netherlands: University Press Rotterdam. https://personal.eur.nl/veenhoven/Pub1980s/89a-C9-full.pdf

Bayer, Ya'akov M., Zeev Shtudiner, Oxsana Suhorukov, and Nimrod Grisaru. 2019. "Time and risk preferences, and consumption decisions of patients with clinical depression." *Journal of Behavioral and Experimental Economics* 78: 138-145. https://doi.org/10.1016/j.socec.2019.01.003

Bee, Penny E., Peter Bower, Simon Gilbody, and Karina Lovell. 2010. "Improving health and productivity of depressed workers: a pilot randomized controlled trial of telephone cognitive behavioral therapy delivery in workplace settings." *General Hospital Psychiatry* 32, no. 3: 337-340. https://doi.org/10.1016/j.genhosppsych.2010.01.006

Bernheim, B. Douglas., Debraj Ray, and Sevin Yeltekin. 2015. "Poverty and Self‐Control." *Econometrica* 83, no. 5: 1877-1911. https://doi.org/10.3982/ECTA11374

Chatterji, Pinka, Margarita Alegria, and David Takeuchi. 2011. "Psychiatric disorders and labor market outcomes: Evidence from the National Comorbidity Survey-Replication." *Journal of Health Economics* 30, no. 5: 858-868. https://doi.org/10.1016/j.jhealeco.2011.06.006

Clark, Ben, Kiron Chatterjee, Adam Martin, and Adrian Davis. 2020. "How commuting affects subjective wellbeing." *Transportation* 47: 2777-2805. https://doi-org.stanford.idm.oclc.org/10.1007/s11116-019-09983-9

Copranzano, Russell and Thomas A. Wright. 2001. "When a "happy" worker is really a "productive" worker: A review and further refinement of the happy-productive worker thesis." *Consulting Psychology Journal: Practice and Research* 53, no. 3: 182–199. https://doi.org/10.1037/1061-4087.53.3.182

Coviello, Decio, Erika Deserranno, Nicola Persico, and Paola Sapienza. 2017. "Effect of Mood on Workplace Productivity*." Kellogg School of Management, Northwestern University. https://nicolapersico.com/files/MoodCallCenter.pdf

Goldberg D.P., R. Gater, N. Sartorius, T.B. Ustun, M. Piccinelli, O. Gureje, and C. Rutter. "The Validity of Two Versions of the GHQ in the WHO Study of Mental Illness in General Health Care." *Psychological Medicine* 27, no.1: 191–97. https://doi.org/10.1017/s0033291796004242

Hystad, Sigurd W. and Bjørn Helge Johnsen. 2020. "The Dimensionality of the 12-Item General Health Questionnaire (GHQ-12): Comparisons of Factor Structures and Invariance Across Samples and Time." *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2020.01300

Kessler, Ronald C., Patricia Berglund, Olga Demler, Robert Jin, Doreen Koretz, Kathleen R. Merikangas, A. John Rush, Ellen E. Walters, and Philip S. Wang. 2003. "The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication (NCS-R)." *JAMA* 289, no. 23: 3095-3105. https://doi.org/10.1001/jama.289.23.3095

Krekel, Christian, George Ward, and Jan-Emmanuel De Neve. 2019. "Employee wellbeing, productivity, and firm performance." Discussion paper, Saïd Business School, University of Oxford. https://cep.lse.ac.uk/pubs/download/dp1605.pdf

Kuroda, Sachiko and Isamu Yamamoto. 2018. "Good boss, bad boss, workers' mental health and productivity: Evidence from Japan." *Japan and the World Economy* 48: 106-118. https://doi.org/10.1016/j.japwor.2018.08.002

Management Research and Services, Incorporated. 2006. "Measurement for Professional Services: Driving Operational Performance." http://www.mrsiconsulting.com/wp-content/uploads/Measurement_Professional_Services.pdf

Oswald, Andrew J., Eugenio Proto, and Daniel Sgroi. 2015. "Happiness and productivity." *Journal of Labor Economics* 33, no. 4: 789-822. https://doi.org/10.1086/681096

Rothbard, Nancy P. and Steffanie L. Wilk. 2011. "Waking Up on the Right or Wrong Side of the Bed: Start-of-Workday Mood, Work Events, Employee Affect, and Performance," *Academy of Management Journal* 54, no. 5: 959-980. https://doi.org/10.5465/amj.2007.0056

Schilbach, Frank. 2019. "Alcohol and Self-Control: A Field Experiment in India." *American Economic Review* 109, no. 4: 1290–1322. https://doi.org/10.1257/aer.20170458

Schultz, Alyssa B. and Dee W. Edington. 2007. "Employee Health and Presenteeism: A Systematic Review." *Journal of Occupational Rehabilitation* 17: 547–579. https://doi.org/10.1007/s10926-007-9096-x

Wright, Thomas. A and Barry M. Staw. "Affect and favorable work outcomes: Two longitudinal tests of the happy-productive worker thesis." *Journal of Organizational Behavior* 20, no. 1:

1–23.
https://doi.org/10.1002/(SICI)1099-1379(199901)20:1%3C1::AID-JOB885%3E3.0.CO;2-W