A NEW TOOL TO MEASURE THE RELATIONSHIP BETWEEN HEALTH-RELATED QUALITY OF LIFE AND WORKFORCE PRODUCTIVITY

A Dissertation
Presented to
The Academic Faculty

By

David T. Huang

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology

August 2008

A NEW TOOL TO MEASURE THE RELATIONSHIP BETWEEN HEALTH-RELATED QUALITY OF LIFE AND WORKFORCE PRODUCTIVITY

-

Approved by:

Dr. Branislav Vidakovic, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paula Edwards, Co-Advisor
Himformatics, LLC

Dr. David Goldsman
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paul Griffin
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Eva Lee
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: May 9, 2008

ACKNOWLEDGEMENTS

Any dissertation is never the result of just one person's work, and my humble composition is no different. I would be remiss if I did not acknowledge first and foremost my Lord and Savior Jesus Christ, without whom my life would not have complete meaning and purpose.

I wish to thank my advisor, Dr. Brani Vidakovic, and co-advisor, Dr. Paula Edwards, for their invaluable support, advice, and encouragement. I would also like to thank my other committee members, Drs. David Goldsman, Paul Griffin, and Eva Lee, for their helpful contributions. Thanks also go to Drs. François Sainfort and Julie Jacko, who provided crucial instrumental support and guidance throughout the pursuit of my Ph.D. My fellow HSI cohorts, especially Brad Jones, Leanne Metcalfe, Sofia Espinoza, Ji Soo Yi, and Kevin Moloney, have provided encouragement and moral support through my graduate studies as well, as did numerous friends at the Atlanta Chinese Christian Church, including Daniel Ng, Mu-Hsin Wei, Michael Lo, Leo Jeng, and Frank Huang.

Lastly, I wish to acknowledge my family for their unconditional love and support. Words could not fully describe how my parents have always pushed me to put forth my very best and sacrificed a great deal so that I could be where I am today. My brother Sam has provided moral support and love, even as he pursued his own graduate studies. I'm especially grateful for my fiancée, Ellen Wu, whose faith in me and love for me continually amazes me, and for her parents, who have supported and encouraged me as they would their own son.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

A30	30-day absenteeism (continuous)
AB30	30-day absenteeism (binary)
ALWQ	Angina-Related Limitations at Work Questionnaire
AM	Whether achievements and mistakes occurred in the past 30 days
BP	Bodily Pain (component of PCS)
EH	Number of extra hours one has worked in the past 30 days (continuous)
EHB	Number of extra hours one has worked in the past 30 days (binary)
EWPS	Endicott Work Productivity Scale
Fa	Faculty
GH	General Health (component of PCS)
GT	Georgia Institute of Technology
HLQ	Health and Labor Questionnaire
HLW	How often health limited work in the past 30 days
HPQ	Health and Work Performance Questionnaire
HQWP	Health-Related Quality of Life and Work Productivity Questionnaire, the survey used in this study
HRQOL	Health-related Quality of Life
HWQ	Health and Work Questionnaire
IRB	Institutional Review Board
MCS	Mental Component Summary (component of SF-36)
MH	Mental Health (component of MCS)
MHPQ	MacArthur Health and Work Performance Questionnaire
MIDUS	MacArthur Foundation Midlife Development – United States
MWPLQ	Migraine Work and Productivity Loss Questionnaire
OT	Osterhaus Technique
P30	30-day presenteeism (self-reported)
P7	7-day presenteeism (self-reported)
PCS	Physical Component Summary (component of SF-36)
PF	Physical Functioning (component of PCS)

PP7	7-day presenteeism, as a peer would rate
PRQ	Presenteeism-related Questions
Prod7	Number of days in the past 7 productivity was lower than expected
PS7	7-day presenteeism, as a supervisor would rate
QALY	Quality-Adjusted Life Year
QQ	Quantity and Quality Instrument
RE	Role Emotional (component of MCS)
RP	Role Physical (component of PCS)
SF	Social Functioning (component of MCS)
SF-36	Short-Form 36 Test
SMPH	Summary Measure of Population Health
SPS-6	Stanford Presenteeism Scale – 6 Question Version
SPS-13	Stanford Presenteeism Scale – 13 Question Version
SSL	Secure Sockets Layer
St	Staff
STD	Standard deviation
UHI	Unnamed Hepatitis Instrument
VT	Vitality (component of MCS)
WLQ	Work Limitations Questionnaire
WPAI-GH	Work Productivity and Activity Impairment General Health Questionnaire
WPAI-SHP	Work Productivity and Activity Impairment Specific Health Problem Questionnaire
WPI	Worker Productivity Index
WPSI	Work Productivity Short Inventory

SUMMARY

Quality of life and productivity are two important measures in health outcomes that usually require the use of self-reported surveys for accurate assessment. Measuring health-related quality of life (HRQOL) has been established as an important field in the past century, and many psychometrically validated instruments exist for both general and specific population use. Another health measure that has attracted a significant amount of attention in recent years, although there is no gold standard, is that of workforce productivity. Most productivity studies measure the amount of work loss incurred by employees in the form of absenteeism and presenteeism.

A new survey that combines questions from existing HRQOL and workforce productivity surveys, the Health-Related Quality of Life and Work Productivity Questionnaire (HQWP), was constructed and tested using a descriptive, cross-sectional study of faculty and staff at a major research university. As expected, HRQOL and work loss were found to be negatively correlated. In addition, staff were found to have statistically higher levels of absenteeism than faculty, but faculty had higher levels of presenteeism. Using multivariate regression models on several measures of productivity, including both absenteeism and presenteeism, we concluded that mental health measures were stronger predictors of productivity than physical health measures for our overall sample, as well as faculty and staff groups separately. In addition, those who work extra hours to make up for lost production had significantly lower social function scores compared to those who do not. Other statistical analyses performed include PCA factor analyses on presenteeism covariates. Lastly, we performed economics analyses on the

cost savings that could be achieved through health management programs to reduce absenteeism and presenteeism levels.

A better understanding of reasons for absenteeism and presenteeism could help inform targeted workplace programs to reduce employer indirect costs related to lost productivity. Moreover, such programs could reduce rates of turnover due to increased employee satisfaction, as well as improve both quantity and quality of life years.

CHAPTER 1

INTRODUCTION

Quality of life and productivity are two important measures in health outcomes that, unlike many health measures, usually require the use of self-reported surveys for accurate assessment because of their subjective nature. Health-related quality of life (HRQOL) measures describe general or domain-specific well-being directly related to physical and mental health aspects of life. Measuring HRQOL has long been established as an important field in the past century because of its implications for chronic disease impact, clinical effectiveness, resource utilization, medication expenditures, and reimbursement for payers.[1]

On the other hand, workforce productivity is a health measure that has attracted more attention in recent years as an emerging field of study. Evaluating workforce productivity is an important but difficult task that helps employers with tasks such as performance assessment, resource allocation, and indirect cost assessment (from lost productivity). Many recent productivity studies focus on measuring the amount of work loss incurred by employees, whether employees are absent from work, measured as absenteeism; or present at work, measured as presenteeism. For example, an employee who was absent from work for one day and a half days in a given week would have incurred 1.5 days of absenteeism. If that employee's job performance for that same week was measured as 8 on a scale from 0 to 10, where 0 is the worst job performance anyone could have at that job 10 is the performance of a top worker, then that person's presenteeism value would be reported as 0.2. While absenteeism can be measured by

combining the number of days and hours spent away from work, presenteeism is a much more difficult construct to measure since it requires estimating how productive employees are while at work. Since there is currently no dominant or gold standard in workforce productivity measurement,[2] any significant study on the subject is an important contribution to the knowledge base.

Measuring both HRQOL and productivity of employees from an employer perspective have become much more significant fields of research in recent years because of the increasingly important roles employers have assumed in healthcare. In fact, employers have been called the new "gatekeepers" of healthcare since they have control over not only influential employee health and insurance programs, but also the very environments – physical, emotional, and, to some extent, social – where people spend a large proportion of their lives. Healthy and satisfied workers contribute greatly to the productivity and efficiency of any company, and employers have a great responsibility, both legal and societal, to provide for the well-being of their employees.[3]

Because HRQOL and productivity are not, generally speaking, objective quantities, survey instruments are normally required estimate these quantities. As with any measurement instrument, surveys designed to measure HRQOL and workforce productivity need to undergo testing to ensure requisite reliability, validity, and responsiveness standards are met. These tests collectively fall into a broad class of assessments known as psychometric tests.

Reliability, which measures the ability of an instrument to generate consistent and reproducible results, can be broken down further into measures of internal consistency and reproducibility. Internal consistency, a measure of how well items in an instrument

correlate with other items in a set, is commonly measured using a coefficient developed by Cronbach, generally known as Cronbach's α and defined to be $\left(\frac{k}{k-1}\right)\left(1-\frac{\sum \sigma_i^2}{\sigma^2}\right)$, where k is the number of items, σ_i^2 is the variance of item i, and σ^2 is the variance of the observed total.[4] In general, instruments with Cronbach's α values of at least 0.5 to 0.7 are said to have acceptable internal consistency reliability.[5]

The other measure of reliability is reproducibility, which measures to what degree an instrument produces stable scores. Specifically, test-retest reproducibility is the degree to which a test performed a stable scores when performed a second time on a clinically stable patient.[2] The most common method used to assess test-retest reproducibility is the interclass correlation coefficient (ICC),[6] which should have a value of at least 0.7 to have acceptable test-retest characteristics.[7] When a test cannot be performed more than once due to time or resource constraints, one can also use split-test reliability measures, which involve splitting the test into two equivalent halves, calculating a Pearson r coefficient between the two halves, and adjusting the correlation using the Spearman-Brown formula. This formula is defined as $r_{SB} = \frac{nr_{xy}}{1+(n-1)r_{xy}}$, where r_{xy} is the original Pearson r coefficient and n is the number of items in the revised version divided by the number of items in the original version (so $n = 2$ for split-half reliability).[8] For surveys that are not self-administered, another reproducibility measure used is inter-rater reproducibility, which measures to what degree a test administered by two different raters on the same patient produces stable scores. The commonly used measure of inter-rater reproducibility is the κ coefficient, which should have a value of at least 0.7 to be acceptable.[7]

The validity of an instrument attempts to describe how well it measures what it was intended to measure. Content validity is a measure of how well the instrument components collectively represent the intended instrument domain. Construct validity assesses the relationship between items and with hypothesized measures, usually using correlation coefficients. Criterion validity describes how well the survey measures against external validated sources like other instruments (concurrent validity) or between groups that have known differences like comparing the health of young adults and the elderly (clinical validity). Assessing criterion validity usually involves calculating Spearman or Pearson correlation coefficients or performing analysis of variance calculations. Finally, predictive validity attempts to measure how well the instrument predicts future outcomes. Responsiveness, a measure of how well an instrument measures changes in outcomes due to an intervention, is performed only for longitudinal studies. Calculations usually involve examining changes in scores and calculation of effect sizes for specified sub-scores.[2, 8] For further information on psychometrics, several reputable resources are available, including Cohen and Swerdlik (2002),[8] Aiken and Groth-Marnat (2005),[9] and Kline (2000).[10]

CHAPTER 2

LITERATURE REVIEW

2.1 Health-Related Quality of Life

Many psychometrically validated HRQOL instruments exist for both general and specific population use, and standards such as QALY (Quality Adjusted Life Years) and Short-Form-36 (SF-36) are commonly used to test and validate new instruments to measure various aspects of health. Instruments to measure HRQOL can be divided into two types, generic and specific (Table 2.1). Generic instruments include all instruments that are not specific to one segment of the population, while specific instruments consider populations in a specific domain such as a certain age or disability group.[1]

Table 2.1: HRQOL Measures

Type	Approach	Strengths	Weaknesses	Examples
Generic	Health profile	One instrument that measures different aspects of health status	May not be responsive	SF-36
	Utility measurement	Single number that represents quantity and quality of life; allows for CUA	Utility values are difficult to calculate; may not be responsive	HUI, QALY, EuroQol, EQ-5D
Specific	Disease specific	Clinically sensible; may be more responsive	Limits populations and interventions; cross-sectional comparisons may not be possible	PDQ-39, Seattle Angina Questionnaire
	Population specific			
	Function specific			
	Problem specific			

2.1.1 Generic Instruments

2.1.1.1 Health Profiles

Health profiles offer the advantage of being able to measure different aspects of health status in any population, regardless of any underlying condition or characteristic, allowing for effective, broad comparisons of general health. Like other utility measurements, however, health profiles may not be as responsive to changes in inputs like specific instruments may be.

One of the largest studies of health profiles ever administered, the Medical Outcomes Study (MOS) was a cross-sectional and longitudinal study designed to evaluate adult patients on health status and treatment in different health care settings. The study was conducted in Los Angeles, Boston, and Chicago from 1986-1987 using a 245-item baseline questionnaire that includes both generic and specific questions. In the cross-sectional study adult patients (n = 22,462) considered health treatment and status, while a sample of those patients (n = 2349) with diabetes, hypertension, heart disease, or depression were surveyed before and after treatment for the longitudinal study. MOS was designed with two purposes in mind: 1) to determine whether differences in patient outcomes can be explained by differences in system of care, clinician specialty, and technical and interpersonal style of the clinician; and 2) to develop practical tools for routine monitoring of patient outcomes, including clinical results; physical, social, and role functioning in daily living; patient perceptions of general health and well-being; and patient satisfaction with treatment received.[11]

Probably the most sensitive, reliable, and validated of the survey measures designed to measure HRQOL, the SF-36 was developed by identifying eight key concepts from MOS which satisfy minimum psychometric standards needed for group comparisons and selecting questions which address those concepts in the MOS instrument to create a new instrument.[12, 13] As a generic QOL measure, the SF-36 is by far the most widely used instrument according to a literature review by Garratt et al. (2002), having almost four times as many publications as the next closest instrument, the sickness impact profile.[14] The popularity of the instrument can be attributed to the delicate balance it has achieved between brevity and comprehensiveness.[15]

The SF-36 is composed of two summary measures, physical component summary (PCS) and mental component summary (MCS), each of which has four scales (Table 2.2). PCS consists of physical functioning (PF), role-physical (RP), bodily pain (BP), and general health (GH). The PF scale consists of ten questions which relate to normal physical activities such as carrying groceries, climbing stairs, and bathing. The RP scale comprises four questions relating to limitations in work or daily activities due to physical health, making this scale of particular interest because these questions are directly related productivity. The BP scale has two questions regarding the magnitude and interference of physical pain. Lastly, the GH scale has five questions related to general health and also has significant correlation with the mental health measure as well as the physical health measure.[12]

MCS also consists of four scales: vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH). The VT scale describes energy and fatigue level using four questions. The SF scale consists of two questions which ask about the

effect of physical and emotional health on normal social activities. Both VT and SF have significant correlations with the physical health measure. The RE scale consists of 3 questions regarding the effect of emotional health problems on work and daily activities, similar to the RP scale but from a mental health perspective. Like RP, this scale is of particular interest because it asks questions related to productivity. Finally, the MH scale comprises 5 questions regarding general mental health.[12]

Table 2.2: SF-36 Summary

Summary Measure	Scale	SF-36 Questions (number)	Number of Levels	Notes
Physical Component Summary (PCS)	Physical Function (PF)	3a – 3j (10)	21	--
	Role Physical (RP)	4a – 4d (4)	5	Contains questions related to productivity
	Bodily Pain (BP)	7, 8 (2)	11	--
	General Health (GH)	1, 11a – 11d (5)	21	Significant correlation with mental summary measure
Mental Component Summary (MCS)	Vitality (VT)	9a, 9e, 9g, 9i (4)	21	Significant correlation with physical summary measure
	Social Function (SF)	6, 10 (2)	9	Significant correlation with physical summary measure
	Role Emotional (RE)	5a – 5c (3)	4	Contains questions related to productivity
	Mental Health (MH)	9b, 9c, 9d, 9f, 9h (5)	26	--

The survey itself is presented in its entirety as Appendix A. Note that question 2 regarding change in health in the past year is not scored, so there are only 35 questions listed in this table. Many published articles have described psychometric and quality testing the survey has undergone, including seminal papers by Ware et al. (1992),[12]

McHorney et al. (1993),[13] and McHorney et al. (1994).[16] And although the SF-36 includes work-related items (the 7 items comprising the RP and RE scales) and is the most common instrument with which productivity instruments are compared, it cannot be used as an effective standalone productivity instrument because the relevant items in the survey are binary and cannot provide a useful measure of productivity.[2]

2.1.1.2 Utility Measurement

Utility measures are derived from economic and decision theory and result in a single number that represents quantity and quality of life. Patient preferences for treatment process and outcome are included as key inputs to utility measures. In addition, utility measures offer the advantage of allowing for CUA, which may be more appropriate than other economic evaluations. On the other hand, utility values can be difficult to calculate and may not be as responsive to changes in inputs as other instruments may be.

Perhaps the mostly commonly used utility measure for HRQOL, QALY is an attempt to combine both quality and quantity of life into a single numeric measure. QALYs are based on utility theory, also known as von Neumann-Morgenstern expected utility theory. Normatively compelling rules for rational decisions under uncertainty, also known as the three axioms of expected utility theory, are as follows:

Given: X is a set of outcomes
$\Delta(X)$ is a set of probability distributions over X
$\succ$ is a person's preference relation over probability distributions
$\sim$ is the indifference relation over probability distributions

1. *Weak order*: $\succ$ is asymmetric ($p \succ q$ implies not $[q \succ p]$), and both $\succ$ and $\sim$ are transitive (if p $\succ$ (~) q and q $\succ$ (~) r, then p $\succ$ (~) r) for all $p, q, r \in \Delta(X)$.

2. *Independence*: For all $p, q, r \in \Delta(X)$ and any $\alpha \in [0,1]$, $p \succ q$ if and only if $\alpha p + (1-\alpha)r \succ \alpha q + (1-\alpha)r$.
3. *Continuity*: For all $p, q, r \in \Delta(X)$ such that $p \succ q \succ r$, there exist α and $\beta \in [0,1]$ such that $\alpha p + (1-\alpha)r \succ \beta p + (1-\beta)r$.[17]

Derivation of health states for QALYs includes a three-step procedure. First, the condition is described using some generic health classification like the Health Utilities Index or EQ-5D. This mapping procedure should, if possible, involve empirical analysis data from patient surveys. Next, the generic data is condensed into one or more one-year descriptions about the health state over time. Finally, the one-year descriptions are converted to a numerical value either directly by a health rater or by applying a formula already available for indices such as the Health Utilities Index or EQ-5D.[18] This last step of deriving the health state weights is itself an intricate task, involving the following three steps: 1. Important aspects of health to be used as health states are decided upon. 2. A classification system is created to map individual health conditions into the discrete set of health states. 3. Some sort of system is devised to assign numbers to health states.[19]

The most commonly used decomposition form of QALY is expressed as a total utility: $U(y,t) = u(y) \cdot m(t) = q \cdot l$, where $u(y)$ is a health utility function as a function of health condition y, $m(t)$ is some increasing function of the duration t with $m(0)=0$, and q and l are the health state and duration, respectively. In practice, $m(t)$ is usually assumed to be the linear function $m(t)=t$, and most of the focus in calculating QALYs is spent determining the utility function $u(y)$ to get values for health states q.

The most important practical concern regarding the use of QALYs is the development and use of utility assessments that can be used to construct utility functions $U(y)$. The time-tradeoff (TTO) procedure involves asking the patient to decide what reduced time duration she would accept in to improve her health from y to y^*, a full

health state. That is, for what value t' is $U(y,t) = U(y^*,t')$, where t is the duration of the current (imperfect) health state and $t' < t$ is a reduced time duration in full health? Assuming a linear QALY model, the equation can be solved for the desired health state as follows:

$$U(y,t) = U(y^*,t')$$
$$u(y) \cdot t = u(y^*) \cdot t'$$
$$u(y) = 1 \cdot t'/t = t'/t$$

Clearly, TTO assumes that QALYs should be positively correlated with longevity. Standard-gamble (SG) techniques involve asking a patient to specify the largest probability of death p she would be willing to accept in order to raise her health state from y to full health state y^*. Since $1 - p$ is the probability of living in the full health state, the desired health state can be solved as follows:

$$U(y,t) = (1-p) \cdot U(y^*,t)$$
$$u(y) \cdot m(t) = (1-p) \cdot u(y^*) \cdot m(t')$$
$$u(y) = (1-p) \cdot u(y^*) = 1 - p$$

Clearly, SG assumes that QALYs are negatively correlated with risk.[20]

A third commonly used technique besides TTO and SG for determining a suitable utility function is the visual analog scale (VAS), which asks respondents to mark on an analog scale their perception of different health states. Although it is simpler to complete and cheaper to implement than TTO and SG (not to mention easier to understand for respondents), VAS is considered by most economists to be inferior to the other two techniques because it does not ask people to make trade-offs in their utility function. Thus, if there were some way to map VAS values onto TTO or SG values, one could come up with theoretically superior utility functions faster and cheaper.[21] Some techniques used less commonly to determine suitable utility functions include willingness

to pay (WTP), where the subject is asked how much she would be willing to pay to have full health restored from a certain state; and person trade-off (PTO), where the subject is asked to decide how many persons in good health she would trade for a number of persons in some other health state.[18]

Another theoretical consideration is that concerning risk with respect to duration function $m(t)$. If the decision-maker wishes to remain risk-neutral regarding future life-year uncertainty, a simpler but less realistic assumption (empirically), she can simply use the standard linear QALY model. On the other hand, a decision-maker who wishes to adjust for risk can use the following (generalized) form: $U(y,t) = u(y) \cdot m(t) = u(y) \cdot t^r$, where the terms are the same as before, except r is defined to be a risk parameter that defines the utility function for life duration. Defining $r = 1$ is the same as using the linear, risk-neutral QALY model, while $r > 1$ denotes the degree of risk seeking and $r < 1$ denotes the degree of risk aversion.[22] In the most basic form of QALYs, it is assumed that the health states are chronic, meaning the health state is the same from onset until death. This assumption, however, is relaxed when considering multi-state profiles.

2.1.2 Specific Instruments

The other approach to measuring HRQOL involves a focus on a specific disease (like Parkinson's), population (like the elderly), function (like sleep), or problem (like pain). For example, several instruments exist to measure the HRQOL of patients with Parkinson's disease, including the Parkinson's disease questionnaire-39 item version (PDQ-39), Parkinson's disease quality of life questionnaire (PDQL) and "Fragebogen Parkinson LebensQualität" (Parkinson Quality of Life questionnaire; PLQ).[23] Specific

instruments are used where clinically sensible and may be more responsive to conditions and characteristics that are specific to the population being studied. However, specific instruments limit the populations and interventions that are tested, and cross-sectional comparisons may not be possible.

2.2 Workforce Productivity

The most significant component of indirect costs, workforce productivity measures, measured as both absenteeism and presenteeism, are unfortunately very difficult to calculate. Traditionally, absenteeism has received much more attention as a field of inquiry, as employers have recognized the benefits of being able to measure and reduce effectively a visible but sometimes overlooked source of costs. To be sure, absenteeism is not always easy to measure because, although missed days due to reported sickness can be compiled rather easily from employer files, employees often take sick days for personal, not health reasons. Also, with increased telecommuting rates, some employers are having an increasingly difficult time measuring absenteeism rates with accuracy. Nonetheless, there is support for a strong correlation between employer-reported and employee-reported absenteeism rates.[24] Some of the seminal and review papers found in the literature address other aspects of absenteeism, including its relationship to job satisfaction,[25] the effect of reporting absenteeism in social contexts,[26, 27] and new ways of valuing absenteeism-related costs.[28]

In recent years, however, presenteeism has become a much more prominent in productivity studies as the extent and severity of its role in workforce productivity has become increasingly apparent. Brouwer et al. (2002) examined workforce productivity

before and after absence to determine the amount of lost productivity that is unaccounted for when considering only actual days absent due to an episode of illness. Although their sample size was small (n = 51), their analysis revealed an increase in production losses by 16% when adding presenteeism costs to absenteeism costs alone. Note, however, that their measure of presenteeism does not include episodes unassociated with absence from work.[29] In fact, a few other studies found in the literature concluded that costs associated with presenteeism can exceed those of absenteeism and medical treatment combined, making quality measures of presenteeism even more crucial.[30-32] For certain chronic health conditions, including allergy, arthritis, asthma, and migraine, Goetzel et al. (2004) found that a whopping 70-80% of combined employer direct (medical and pharmacy) and indirect (productivity) costs come from presenteeism alone.[32]

Measuring overall workforce productivity loss from the employer perspective, including both presenteeism and absenteeism, typically involves creating and implementing self-reported employee surveys. In some cases where performance may be measured using objective measures such as number of items produced or length of customer service calls taken, productivity as measured by presenteeism and absenteeism can be measured using numerical indices that are relevant and easily understood. However, self-reported workforce surveys are the only practical method that can be used in a variety of settings and job functions to measure workforce productivity.[33]

An important characteristic to consider in productivity surveys is the recall period, which describes how far in the past participants must remember events such as episodes of illness or emotional problems. While longer recall periods provide more information for each subject and have greater statistical power, shorter recall periods typically provide

more accurate information. For the most part, productivity instruments use recall periods of the present, 1, 2, or 4 weeks, though some less common recall periods include the most recent episode and one year (for very major events).

2.2.1 Existing Instruments

A thorough literature review over the past 20 years revealed 16 major surveys that have been used to measure workforce productivity. Although most surveys consider the effect of general health conditions on productivity, a few focus on the effect of specific conditions such as angina, hepatitis, or migraine on productivity. The pages that follow summarize the 16 instruments and their value to the knowledge base. Review papers that were especially helpful in the identification and evaluation of these surveys included Lynch and Riedel (2001),[34] Evans et al. (2003),[2] Loeppke et al. (2003),[35] Ozminkowski et al. (2004),[36] Goetzel et al. (2004),[32] Prasad et al. (2004)[37], Chapman (2005),[38] and Lofland et al. (2004).[39] Note that tables summarizing features of the 16 surveys follow in Section 2.2.2.

The Angina-Related Limitations at Work Questionnaire (ALWQ) was developed at the New England Medical Center with funding from Merck and was designed for use within clinical trials and investigations on treatment effectiveness of individuals suffering from chronic stable angina pectoris, or angina. The questionnaire, shown in Appendix B as Figure B.1, consists of 17 questions on a 5-point Likert scale related to work limitations caused by angina and 7 questions designed to measure paid work time loss (both absenteeism and presenteeism). The survey is public domain, and all questions are available in the only publication that was found on the instrument.[40] ALWQ was the only

survey found that focuses on employed individuals with angina, which naturally limits its usefulness in measuring work loss for the general population. In addition, the survey has been the subject of only one significant journal publication (Lerner et al., 1998). On the other hand, it has high internal consistency reliability (Cronbach's $\alpha = 0.97$ and item-to-total correlations mostly >0.75) and good convergent validity. Finally, content and criterion validity were not studied, and construct validity was weak to moderate with a range of -0.35 ($p = 0.04$) to -0.63 ($p = 0.0001$).[40]

The Employer Health Coalition (EHC) is a not-for-profit organization based in Florida that was created to improve health services methods and economies-of-scale, mostly for major corporations in that state. With financial support from Schering-Plough Corporation, EHC created and administered a proprietary Healthy People/Productive Community Survey to 8 employers in 1998, 9 employers in 1999, and 5 employers in 2005-2006 as part of its vision for quality health care on a community-wide basis. Administration was via mail and conducted in two phases. The first phase included about 200 general health questions and the second included about 150 disease-specific questions. Relevant questions asked included disease/condition prevalence (defined as previous 20 working days), number of days absent due to the condition, and the resulting work impairment (including communication, quality of work, and overall productivity) using a Likert scale. Although EHC staff verbally conveyed moderate reliability of the survey and convergent validity testing for some of the diseases (based on correlations with the SF-36 and SF-12), the EHC Assessment remains the only instrument found for which no peer-reviewed published data exist.[35]

The Endicott Work Productivity Scale (EWPS) was developed at the New York State Psychiatric Institute with support from Pfizer. The survey computes a total score from 0 (best) to 100 (worst) and consists of 25 items on a 5-point measured on a Likert scale, covering four productivity areas: attendance (absenteeism and time on task), quality of work, performance capacity, and personal factors (social, mental, physical, and emotional). The instrument is a copyrighted tool charging a fee for commercial use; other researchers need to request permission and complete a license arrangement. Although we were able to purchase a copy of the survey for examination purposes, we are not able to reproduce it here since questionnaire rights are reserved.

Internal consistency reliability (Cronbach's $\alpha = 0.93$) and test-retest reliability (intraclass correlation coefficient of reliability for the total score was 0.92) of EWPS have been found to be high. The survey has demonstrated validity in measuring severity of illness, but psychometric testing has not been performed to test the instrument against performance or productivity standards. In addition, EWPS has been the subject of only one significant journal publication (Endicott and Nee, 1997). Finally, although it was designed for general use, the survey has only been tested on patients with depression.[41]

The Health Limitations Questionnaire (HLQ) was developed at the Erasmus University Rotterdam Institute for Medical Technology (one of two major instruments developed outside the United States) and designed to examine the relationship between illness, treatment, and work performance. The 23-item self-administered paper survey is available free of charge with the signing of a licensing agreement and was obtained for study purposes, but cannot be reproduced here because it is not in the public domain. The four modules that comprise the HLQ, each with a different score, include workplace

absenteeism, workplace presenteeism, unpaid production (e.g., household duties), and impediments to paid and unpaid production. Testing of the instrument has been performed on the general population (both employed and unemployed), migraine patients, and patients with hip or knee problems.[42-44] Limited psychometric testing includes criterion validity of the absenteeism module based on a comparison of non-disease specific absence from work with that of a national registry in the Netherlands.[42] A more recent study compared the feasibility and validity of HLQ with that of the Quantity and Quality instrument (QQ), concluding that the HLQ offered less construct validity than the QQ.[44]

Previously known as the MacArthur Foundation Midlife Development in the US (MIDUS) presenteeism instrument, and later as the MacArthur Health and Work Performance Questionnaire (MHPQ), the Health and Work Performance Questionnaire (HPQ) was developed by a team of researchers led by Ron Kessler at the Harvard Medical School Department of Health Care Policy in partnership with the World Health Organization (WHO) and with support from the John D. and Catherine T. MacArthur Foundation. The MHPQ, is shown in Appendix B as Figure B.2 and is a 30-item questionnaire designed to measure the association between general health and four dimensions of work performance, including presenteeism, absenteeism, work-related accidents and injuries, and work-related successes and failures. The instrument is well suited for general use, as it is relatively short and does not contain disease-specific questions. The HPQ is a much longer survey based on the MHPQ, including 92 questions and going into much more detail about existing chronic conditions and detailed productivity questions such as the number of hours one has worked in the past 30 days.

In addition, the survey contains some questions related to HRQOL, including some that are very similar to those found in the SF-36.[37]

Productivity losses for both MHPQ and HPQ are measured on a Likert scale against workload, health status, and actual performance. Several versions of the instrument are available as public domain. The reliability testing that has been performed on MHPQ/HPQ revealed that internal consistency reliability of the work performance module was moderate (Cronbach's $\alpha = 0.74$ for a reservation agent group and 0.81 for a customer service representative group).[45] In addition, test-retest reliability over two months on a sample of airline reservation agents ($n = 105$) was found to be high (0.89).[46] Validation studies include calibration studies on the relationship between MHPQ/HPQ results and employer archival measures of work performance and absenteeism, which have shown good concordance across four different occupation groups.[45, 47] Interestingly, the newer version of the survey, the HPQ, is the only instrument found that has established a data consortium to facilitate the collection and comparative reporting of HPQ data. Membership is fee-based and open to any organization, public or private, that is willing to contribute HPQ data to the International HPQ Archive.[48]

The Health and Work Questionnaire (HWQ) was developed to assess the relationship between smoking and workplace productivity. The instrument was created with financial support from GlaxoSmithKline and is shown in Appendix B as Figure B.3. HWQ consists of 30 questions, each on a 10-point Likert scale, that have been categorized into six subscales: productivity, supervisor relations, patience, concentration, work satisfaction, and non-work (personal) satisfaction. Of particular note is the fact that one's own productivity was asked from three different perspectives – self, peer, and

supervisor – to minimize social desirability tendencies. Reliability and validity testing was performed on a sample of airline reservations agents for a U.S.-based international airline. Internal consistency reliability for each of the six subscales was found to be moderate to high (Cronbach's α ranged from 0.72 to 0.96). The discriminant validity component of construct validity was measured by comparing the HWQ scores by smoking status. Analysis showed that there was a significant ($p < 0.05$) difference for non-work satisfaction and productivity from a peer's perspective.[49, 50] Criterion validation was performed by comparing correlations between HWQ scores and two objective measures: Total Hours Lost (THL), which measures the average time an individual was unavailable to answer calls without an approved excuse, and Total Performance Points (TPP), which measures employee performance using a proprietary formula based on revenue generated, call waiting time, and ticket delivery service created by the employee. Correlations of THL were statistically significant ($p < 0.05$) for all HWQ subscales except non-work satisfaction, but correlations of TPP were found to be statistically significant ($p < 0.05$) for only the HWQ work satisfaction and supervisor relations subscales.[50]

The Migraine Work and Productivity Loss Questionnaire (MWPLQ) was developed at the New England Medical Center with funding from Merck to measure the impact of migraine headache on employee work performance. The survey, shown in Appendix B as Figure B.4, has been published and consists of 26 questions, including 18 questions on a 5-point Likert scale related to specific work difficulty tasks due to migraine.[51] These 18 work tasks were categorized into 7 domains: time management, work quality, work quantity, bodily effort, interpersonal demands, mental effort, and

environmental factors. Internal consistency reliability for the work difficulty questions was high, with Cronbach's α ranging from 0.86 to 0.95, while correlations between the work domains were also strong, with Cronbach's α ranging from 0.67 to 0.91. Face validity for work performance was established, as subjects claimed high applicability (over 90%) of the questions with regards to work-related migraine attacks for all domains except for the environmental domain (about 67%). Congruent validity was measured by observing the correlation between indicators of migraine burden (severity, disability, 24-hour Migraine Quality of Life Questionnaire[52, 53], and SF-36) and hours of work loss, effectiveness at work with migraine, work difficulty domains, and overall work difficulty. All indicators reported moderate correlation to those factors, with the exception of SF-36, which had low to moderate correlation.[54] Criterion validity of MWPLQ has not yet been assessed.[39]

Developed in 1992, the Osterhaus Technique (OT) was the first method developed to measure work productivity loss due to illness.[39] OT is a 12-question, self-administered survey developed by a research team at Glaxo (now GlaxoSmithKline) and targets employees who suffer from migraine headache. The survey, from which selected questions are shown in Appendix B as Figure B.5, measures both absenteeism, measured by the number of days of work missed due to migraine, and presenteeism, measured as the days worked with migraine symptoms and hours affected by each migraine episode. Reliability and validity of have not yet been reported in the literature.[55]

The Quantity and Quality Instrument (QQ) was developed at the Erasmus University Rotterdam Institute for Medical Technology and is, along with HLQ, one of two major instruments developed outside the United States. The survey was designed to

measure quantity and quality of work performed daily by asking participants to indicate on two separate VAS scales from 1 to 10 the quality and quantity of work performed. QQ, shown in Appendix B as Figure B.6, consists of five questions. Construct validity was performed because it was not possible to keep track of daily individual answers over a period of time due to anonymity.[44, 56] Correlation of the quantity scale of the QQ method, Qt, with the OT survey was found to be strong (Pearson's coefficient = 0.92), though the OT survey itself has not been validated. On the other hand, correlation of Qt with the HLQ survey, which has some basis for criterion validity, was found to be moderate (Pearson's coefficient = 0.40). The quantity (Qt) and quality (Ql) components were found to be strongly correlated (Pearson's coefficient = 0.59), which may signify a strong relationship between the two components if sufficient discrimination between the two can be shown.[44] In addition, it remains unclear how to combine these quantities to obtain a total productivity loss.[56] External validity, measured by comparing self-reported productivity with actual production output, was reported to be moderate for a group of floor layers (Pearson's coefficient = 0.48) and non-existent for a group of road pavers (due to low variation in actual production output). In addition, QQ was more responsive to health indicators and job characteristics than the HLQ, as mentioned above.[44] Perhaps most noteworthy, QQ was the only survey for which a study was performed to explore the relationship between productivity and HRQOL, using EQ-5D.

The Stanford Presenteeism Scale (SPS) was developed by researchers at the Stanford School of Medicine and the American Health Association with funding from Merck. SPS has evolved through three different versions thus far: SPS-32 (32 questions), SPS-6, and SPS-13; published studies have focused on the latter two versions. SPS was

designed to measure how health affects cognitive, emotional, and behavioral functioning at work, especially appropriate for middle- and upper-level management. Perhaps the main limitation of SPS is that it is one of two tools found, along with the Work Limitations Questionnaire (WLQ), that is designed to measure presenteeism but not absenteeism. SPS-6, shown in Appendix B as Figure B.7.A, consists of 6 questions on a five-point Likert scale and was derived by using an item reduction strategy on SPS-32. It is available free of charge to non-commercial users. Internal consistency reliability for SPS-6 was found to be high (Cronbach's $\alpha = 0.8$). Concurrent validity was found to high, with strong to moderate correlations between SPS-6 scores and scores on specific measures of presenteeism: percentage of productive time ($r = 0.53$, $p < 0.001$), proportion of work accomplished ($r = 0.47$, $p < 0.001$), and percentage of time a person is likely to make more mistakes than usual ($r = -0.31$, $p < 0.001$).[30] Criterion validity was established by confirming that presenteeism scores were lower for those who did have a physical disability versus those who did not, a potentially weak argument because a disability does not necessarily imply less productivity.[39] Finally, discriminant validity was measured by looking at the correlation between total scores to job satisfaction ($r = 0.15$, $p < 0.05$) and stress ($r = –0.22$, $p < 0.01$), with both results suggesting presenteeism is distinguishable from job satisfaction and stress.[30]

In contrast to production-based jobs (also commonly referred to as blue-collar jobs), which usually have objective measures of productivity and output, knowledge-based (white-collar) jobs present a unique but important challenge because productivity is more subjective in nature. With this difficulty in mind, researchers developed SPS-13, shown in Appendix B as Figure B.7.B, as an expansion of SPS-6. Specifically, the new

version served two purposes: to add job function (knowledge- versus production-based) as a criterion to examine presenteeism and to help determine health conditions that have the most impact on productivity. Like SPS-6, SPS-13 offers strong excellent internal consistent reliability (Cronbach's $\alpha = 0.82$). Concurrent validity was measured by comparing results with those from the Work Limitations Questionnaire (WLQ), and the two instruments were found to be correlated (Pearson's coefficient = 0.50). Evidence for criterion validity was found to be weak when comparing correlations between SPS-13 measures and health care use and expenditures (Pearson's coefficients < 0.10). Convergent validity, however, was strongly supported when comparing SPS-13 measures to all eight SF-36 scales (range of Pearson's coefficients between –0.25 and –0.62).[57]

Part of a larger survey specifically for hepatitis patients, the Unnamed Hepatitis Instrument (UHI) Survey has 3 items related to the effect of the disease on work functioning and productivity. Although the survey measures both absenteeism and presenteeism, the latter measure cannot be translated into a monetary variable since decreased productivity is measured dichotomously. No overall score is measured, but patients are classified into groups according to whether their productivity has gotten better, gotten worse, or stayed the same since a specified starting point. The UHI content, shown in Appendix B as Figure B.8, is public domain, and reliability and validity have not yet been reported.[58]

The Work and Health Interview (WHI) was developed by AdvancePCS, now known as Caremark, by comparing six candidate questionnaires on duration, participation rate, recall bias, and discriminative validity.[59] The resulting instrument is a 46-question computer-assisted telephone survey designed to measure the relationship between health

and lost productive time in the U.S. workforce. The instrument requires participants to rate performance on a 5-point Likert scale, and the six survey modules include informed consent, employment status, episodes of 21 specific health problems, job visualization, lost productive time (LPT), and demographics. The instrument is available free of charge with the signing of a licensing agreement but cannot be distributed here due to copyright restrictions.[60, 61] Although reliability of WHI has not yet been published, several validity tests have been performed and published on the instrument. Criterion validity was measured by comparing WHI measures with objective workplace measures. Large correlations were reported for measures of overall (full- and partial-day) absenteeism (Pearson's coefficient of 0.84, $p < 0.0001$; Spearman's coefficient of 0.76, $p < 0.0001$) and LPT (Pearson's coefficient of 0.59, $p < 0.0001$; Spearman's coefficient of 0.63, $p < 0.0001$), and moderate correlation was reported for presenteeism (Pearson's coefficient of 0.29, $p < 0.05$; Spearman's coefficient of 0.31, $p < 0.05$). Only the absenteeism component that measured partial days worked had a weak correlation (Pearson's coefficient of 0.13; Spearman's coefficient of 0.19).[62]

The Work Limitations Questionnaire (WLQ) was developed at The Health Institute at the New England Medical Center in 1998 as a psychometric work-productivity questionnaire with funding from the Kaiser Foundation and GlaxoWellcome, now GlaxoSmithKline. Its aim is to quantify the on-the-job impact of chronic or acute health problems and treatment. The survey consists of 25 items on a 5-point Likert scale and grouped into four demand scales: time management, physical demands, mental-interpersonal, and output demands. Along with SPS, WLQ is one of two surveys whose metrics include presenteeism but not absenteeism. The WLQ, sample questions of which

are shown in Appendix B as Figure B.9, is free to non-commercial users; however, obtaining the instrument requires signing a confidentiality disclosure agreement (CDA). Thus, although a copy was obtained for research purposes, it cannot be published here in its entirety. The survey has been tested on many conditions, including osteoarthritis, rheumatoid arthritis, headache, depression, low back pain, and epilepsy; different job types, including production and service; and a multitude of risk factors, including smoking physical activity, satisfaction, stress, blood pressure, cholesterol, and BMI.[63-67] The WLQ has undergone extensive validity and reliability testing. Internal consistency reliability was reported to be high (Cronbach's $\alpha \geq 0.88$)[66, 68] In addition, item-to-total scale correlation was reported to be < 0.40[68] and < 0.72.[66] Construct validity was measured by comparing survey results to that from the SF-36, and weak correlation was found between each WLQ scale and the physical SF-36 components (r^2 range of 0.14 – 0.22) and between three of four WLQ scales and the emotional SF-36 components (r^2 range of 0.07 – 0.17).[68] Further evidence of construct validity includes comparisons of WLQ scores to various measures of health status and osteoarthritis disease burden.[66] Criterion validity has not been published on this instrument.[39] An especially attractive feature of this survey is that it was developed at The Health Institute, the same place where the SF-36 instrument was developed.[34] Finally, an interesting note is that Dr. Debra Kerner was involved in the development of the ALWQ, MWPLQ, and WLQ.

The Work Productivity and Activity Impairment Questionnaire (WPAI) was developed by Reilly Associates in partnership with University of Texas Medical Branch at Galveston and Marion Merrell Dow (now Sanofi-Aventis) in order to assess the effect of general and specific health conditions on productivity losses. One version (WPAI-GH)

covers general health (6 questions, 5 of which relate directly to work productivity) and another (WPAI-SHP) addresses a specific health problem (9 questions, 2 of which relate directly to work productivity and 2 of which relate directly to school productivity); however, validity and responsiveness for the latter version have been established only for a few diseases, including allergies and chronic hand dermatitis. Scores are calculated for four domain areas: percent of work time missed due to health, percent impairment while working due to health, overall work impairment due to health, and percent activity impairment due to health. All survey questions are public domain and available on the website of Reilly Associates.[69] The general health (WPAI-GH) and specific health problem (WPAI-SHP) versions are shown in Appendix B as Figure B.10.A and A.10.B, respectively. Written permission is not required to researchers using the WPAI.[69]

The bulk of WPAI reliability and validity testing has been performed on the general health version of the survey. Test-retest reliability of the instrument was performed by comparing scores taken at different times and was found to be very good (Pearson's coefficient > 0.69). Construct validity was measured by computing correlations between WPAI-GH scores and SF-36 measures, revealing mostly moderate to strong correlations for SF-36 measures.[70] Discriminative validity was measured by examining the correlation between symptom severity and impairment. Weak correlations were reported for work time missed (Spearman's coefficients range from 0.11 to 0.16) and moderate correlations were reported for impairment at work, overall work impairment, and activity impairment (Spearman's range from 0.3 to 0.53).[71] On the other hand, the general health version is too general and short to be psychometrically valid when measuring subjects with a specific condition. The general health version of

the survey was not found to be valid, reliable, and responsive when used in panic disorder, for example.[72]

Of the instruments found in the literature review, WPAI has by far the longest and most active publication history. Some of the more salient publications include studies related to its use in subjects with asthma and allergies,[71] caregivers, Crohn's Disease, COPD (Chronic obstructive pulmonary disease), dermatology, GERD (gastroesophageal reflux disease),[73, 74] general health, headache, hepatitis, hypertension, IBS (irritable bowel syndrome), mental health, nocturia, and restless legs syndrome.[69]

The Work Productivity Index (WPI), created by researchers at Bank One (now JP Morgan Chase), Northwestern University, Depaul University, and the University of Michigan, is the only instrument that is not a survey but is rather an index that combines measures of absenteeism, presenteeism, and short-term disability. The tool is unique in that it combines an objective absenteeism measure, based on absence and disability records which measure actual time away from work, and a subjective presenteeism measure, which electronically measures employee failure to meet a specified productivity standard rather than using an individual self-report. The target population was thus limited to customer service employees at a call center, though the same technique could conceivably be used in another domain where productivity can be measured objectively and inconspicuously. In addition, the tool can be correlated with health risks when used in conjunction with a Health Risk Appraisal (HRA). The complete WPI formula is shown in Appendix B as Figure B.12.[75] Reliability and validity testing of WPI have not been reported,[39] although one published study did report an inverse relationship between health risks and productivity using WPI.[75]

The Work Productivity Short Inventory (WPSI) was developed by a team at the Institute for Health and Productivity Studies at Cornell University and the Medstat Group with funding from Pfizer. The 22-item survey is designed to estimate changes in productivity, both absenteeism and presenteeism, associated with 15 specific health conditions: allergy, respiratory infection, arthritis, asthma, anxiety disorder, depression and bipolar disorder, stress, diabetes, hypertension, migraine, heart disease/high cholesterol, Alzheimer's, pediatric allergies, otitis media (ear infection), and respiratory infections. The survey is available in 12-month, 3-month, 4-week, and 2-week recall versions and is copyrighted.[76, 77] Although the 12-month version was published in Goetzel et al. (2003) and questions were examined for research purposes, the survey cannot be reproduced here.[76]

A distinguishing characteristic of WPSI is that the respondent must give a finite calculation of lost productivity due to a certain condition, resulting in more conservative estimates than most other instruments, which assume that given a health condition, all work productivity changes are a result of that condition.[32] In addition, due to the comprehensive and specific nature of the questions, WPSI is more focused on the health conditions that affect productivity instead of the nature of losses due to health issues.[36]

Traditional scale-based reliability metrics like Cronbach's alpha were not practical for WPSI because each condition had 3 questions, making it impossible to construct detailed scales within which an overall experience with each condition could be measured. Reliability testing was thus performed by randomly splitting subjects into two groups and comparing survey metrics across these groups. Of the 45 possible comparisons, there was only one significant difference for employees whose work was

affected by arthritis/rheumatism (7% versus 1%, *z*-test *p*-value = 0.03), so the instrument was declared to be quite reliable.[76] Based on prevalence comparisons with claims files and a NHIS (National Health Interview Survey) survey, evidence was found to be strong for content and construct validity but weak for predictive validity. Comprehensive methodology and data summaries can be found in Ozminkowski et al. (2003).[77]

2.2.2 Summaries of Existing Instruments

In the tables that follow (2.3, 2.4, 2.5, 2.6), the key features of the 16 productivity instruments found in the literature review are summarized. Table 2.3 provides general information, including the name of the survey, research and funding sponsors, commonly used abbreviation, description of the question set, and recall period for each of the instruments. Table 2.4 continues with the year of first publication, whether the instrument produces a figure suitable for translation into a dollar amount (e.g., a purely binary survey would not provide such a figure), diseases assessed by the survey, productivity metrics measured by the survey, and availability of each instrument. Next, Table 2.5 provides a matrix identifying surveys that are found in each of the workforce productivity instrument review papers mentioned at the beginning of this section. Finally, Table 2.6 lists other references and notes that are helpful to differentiate between the instruments and were used when deciding which instruments would be used for further study.

Table 2.3: General Summary of Productivity Instruments

Name (Research/Funding Sponsor)	Abbrev.	Description of Question Set	Recall Period
Angina-Related Limitations at Work Questionnaire (New England Medical Center/Merck)	**ALWQ**	17 item questionnaire and appendix of 7 other questions; paper, self-administered	4 weeks
Employer Health Coalition Healthy People/Productive Community Survey (Employer Health Coalition)	**EHC**	Phase 1: 200 general questions; Phase 2: 150 disease specific questions; 5 minute completion time	1 month
Endicott Work Productivity Scale (NY State Psychiatric Institute/Pfizer)	**EWPS**	25 item self-scored questionnaire; paper, self-administered; 5 minute completion time	1 week
Health and Labor Questionnaire (Erasmus University Rotterdam Institute for Medical Technology)	**HLQ**	23-item instrument; paper, self-administered; 10-15 minute completion time	2 weeks
Health and Work Performance Questionnaire (WHO, Harvard Medical School - Department of Health Care Policy/John D. and Catherine T. MacArthur Foundation)	**HPQ**	30 item questionnaire; 20-minute completion time via telephone	1 week and 4 weeks
Health and Work Questionnaire (GlaxoSmithKline)	**HWQ**	27 questions in 6 sub-scales; paper, self-administered	1 week
Migraine Work and Productivity Loss Questionnaire (New England Medical Center/Merck)	**MWPLQ**	23 questions; paper, self-administered	most recent episode
Osterhaus Technique (GlaxoSmithKline)	**OT**	12 questions; paper, self-administered	4 weeks
Quantity and Quality Instrument	**QQ**	5 questions	1 day
Stanford Presenteeism Scale (American Health Association)	**SPS**	32 questions (SPS-32) or 6 questions (SPS-6); self-administered	4 weeks
Unnamed Hepatitis Instrument	**UHI**	3 questions; self-administered	4 weeks
American Productivity Audit - Work and Health Interview (Caremark)	**WHI**	46 questions via computer-assisted phone; 15 minute completion time	2 weeks
Work Limitations Questionnaire (New England Medical Center)	**WLQ**	25 questions; self-administered	2 weeks, 4 weeks
Work Productivity and Activity Impairment Questionnaire - General Health (Reilly Associates, UT Medical Branch at Galveston, Sanofi-Aventis)	**WPAI**	6 questions (general), 9 questions (specific); paper, self-administered, interview administered	1 week
Worker Productivity Index (BankOne)	**WPI**	Objective performance index for customer service phone representatives	N/A
Work Productivity Short Inventory (Institute for Health and Productivity Studies at Cornell)	**WPSI**	22-item questionnaire, 66 questions at maximum	12 months, 3 months, 4 weeks, and 2 weeks

Table 2.4: Descriptive Summary of Productivity Instruments

Instrument	Pub Year	$ fig	Diseases Assessed	Productivity metrics	Availability
ALWQ	1998	No	Specific (Angina)	Absenteeism, presenteeism	Public Domain
EHC	2000	Yes	General	Absenteeism, presenteeism	Questions are proprietary and not available for purchase
EWPS	1997	No	General, but intended to be sensitive to the effects of depression and anxiety disorders	Absenteeism, presenteeism	Copyrighted, fee per user for commercial use; purchased a copy
HLQ	1995	Yes	General	Absenteeism, presenteeism, unpaid production, work impediments	Rights are reserved; purchased a copy
HPQ	2003	Yes	General	Absenteeism, presenteeism, critical event information	Public Domain
HWQ	2001	No	General	Absenteeism, presenteeism, work performance	Public Domain
MWPLQ	1999	Yes	Specific (Migraine)	Absenteeism, presenteeism	Public Domain
OT	1992	Yes	Specific (Migraine)	Absenteeism, presenteeism	Selected questions available only
QQ	1999	No	General	Presenteeism	Public Domain
SPS	2002	No	General	Presenteeism	Public Domain
UHI	2001	No	Specific (Hepatitis)	Absenteeism, presenteeism	Public Domain
WHI	2003	No	General	Absenteeism, presenteeism	Proprietary
WLQ	2001	No	General	Presenteeism	Free to non-commercial users but requires a licensing agreement to use
WPAI	1993	Yes	General	Absenteeism, presenteeism	Public Domain
WPI	1999	Yes	General	Absenteeism, presenteeism	Public Domain (Formula)
WPSI	2003	Yes	General	Absenteeism, presenteeism, care giver demands	Copyrighted but 12-month version has been published

Table 2.5: Review Paper Summary of Productivity Instruments

Instrument	Chapman (2005)	Goetzel (2004)	Loeppke (2003)	Lofland (2004)	Lynch, Riedel (2001)	Ozminkowski (2004)	Prasad (2004)
ALWQ	X			X			
EHC		X	X			X	
EWPS	X			X	X	X	X
HLQ	X	X		X	X	X	X
HPQ	X	X	X		X	X	X
HWQ	X			X			X
MWPLQ	X		X	X			X
OT	X			X			
QQ	(not found in any review paper)						
SPS	SPS-6		SPS-6	SPS-6	SPS-6	SPS-6	
UHI	X			X			
WHI	X	X				X	
WLQ	X	X	X	X	X	X	X
WPAI	X	X	X	X	X	X	X
WPI	X	X		X			
WPSI	X	X				X	

Table 2.6: Other References and Notes for Productivity Instruments

Instrument	Other citations (* citation includes instrument)	Notes
ALWQ	Lerner (1998)*	No published articles since 1998 (obsolete)
EHC	EHC (2000)	No peer-reviewed published data exist
EWPS	Endicott (1997)	Minimal psychometric testing
HLQ	van Roijen et al. (1996), van Roijen et al. (1995), Meerding et al. (2005)	No reliability testing
HPQ	http://www.hcp.med.harvard.edu/hpq/info.php, Kessler (2001), Kessler (2003), Wang (2003), Kessler (2004)	Lots of publishing and has an online consortium of users
HWQ	Shikiar (2001), Shikiar (2004)*, Halpern (2001)	Cannot be translated into monetary units
MWPLQ	Mushet (1996), Davies (1999), Lerner (1999)*	Limited to migraine
OT	Osterhaus (1992)*	Limited to migraine, not available in full form, no published articles since 1992
QQ	Brouwer (1999), Meerding (2005)*	Not in any published review paper
SPS	Koopman (2002), Turpin (2004)	Only measures presenteeism
UHI	McHutchinson (2001)*	Limited to hepatitis
WHI	Stewart (2003), Stewart (2004)	Does not translate into monetary units, no reliability testing
WLQ	Lerner (2001,2002,2003), Burton (2004,2005,2006), Ozminkowski (2003)	Only measures presenteeism but has lots of publications
WPAI	Reilly (1993, 1996, 2003, 2004, 2005, 2006), http://www.reillyassociates.net/, Wahlqvist (2002), Andreasson (2003), Gawlicki (2005, 2006), Chen (2006), Bushnell (2006), Chirban (1997), http://www.reillyassociates.net/WPAI_References.html	Lots of publishing and free usage, but too short
WPI	Burton (1999)	Designed for customer service employees only; not a survey/self-report
WPSI	Goetzel (2003)*, Ozminkowski (2003)	Fewer publications and validation

2.3 Relationship between HRQOL and Productivity

Though several papers have explored the statistical relationship between productivity and HRQOL (using the SF-36 survey) and found direct correlations,[57, 68, 70] none have implemented HRQOL and productivity instruments simultaneously for a general population. Two aspects of the theoretical relationship between HRQOL and productivity that must be addressed include the effect of productivity considerations on health state valuations and the effect of HRQOL on observed work-related productivity. First, the effect of productivity on health state valuations will be considered.

There are two effects on health state valuation that need to be considered when it is determined that a person's productivity will not be normal (i.e., lower than usual). First, role functioning in paid or unpaid work, which describes an ability to perform work-related functions and be a productive member of society, is an aspect of health that is affected by changes in productivity. In general, existing HRQOL instruments already capture this element of reduced productivity as one of its dimensions.[78] In the SF-36, for example, role functioning is captured in the RP and RE scales. Second, loss of productivity could lead to loss of income and unpaid production, which could have an effect on health state according to the US Panel on Cost-Effectiveness in Health and Medicine.[79] This phenomenon leads to a valuation of imperfect HRQOL (e.g., disability) based on lost income rather than based directly on lost productivity. Generic HRQOL questionnaires such as SF-36 and EuroQol intentionally do not ask questions specifically related to income effects. On the other hand, lost income could also lead to an increase in social benefits or level of private insurance, which may actually result in a net increase in HRQOL valuation.[80] Suggested areas of future research in this area that have been

identified include altering existing or creating new HRQOL questionnaires that are more sensitive to changes in productivity.[78]

Even if productivity could be measured using non-obtrusive, objective methods (which would severely limit the job functions that could be studied) and only HRQOL measurement required regular measurement, successfully implementing a longitudinal study on the relationship between HRQOL and productivity would be logistically complex and of questionable reliability and validity due to respondent burden and fatigue error from asking the same set of questions regularly over a long period of time.

Instead, a more reasonable and empirically testable relationship is a cross-sectional study of productivity and HRQOL, which by itself presents a great challenge. In one such study found in the literature, Lamers et al. (2005) concluded that using HRQOL to model productivity costs was not recommended. However, their study was based on a short, general questionnaire (EQ-5D) on people with a specific condition (lower back pain). In addition, their study is a secondary analysis on existing data, which weakens their argument and underscores the need for a study specifically designed to measure both workforce productivity and HRQOL.[81]

Another, more recent study (Allgulander et al., 2007) looked at the impact of the anti-depressant Escitalopram for patients with Generalized Anxiety Disorder (GAD) on HRQOL and work productivity. Although the authors found a correlation between treatment and increased HRQOL, and between treatment and increased work productivity, the authors did not look at the direct relationship between HRQOL and work productivity.[82]

2.4 Summary of Key Research Needs

Overall, the literature review exposed several gaps in the measurement of HRQOL and workforce productivity: 1. There is no accepted gold standard for measuring workforce productivity among a general employee population. 2. The relationship between HRQOL and workforce productivity has not been clearly defined, theoretically and empirically. 3. There are no studies comparing the effect of physical and mental HRQOL components on productivity. 4. No existing studies have looked at differences between production- and knowledge-based jobs (traditionally referred to as blue- and white-collar jobs) or between regular and flexible time jobs in measuring HRQOL and productivity. 5. Studies have not examined the impact of missed work due to health on increased productivity (i.e., working extra hours to make up for time lost due to health) and the resulting combined impact on social functioning.

CHAPTER 3

METHODOLOGY

3.1 Problem Statement

The literature review exposed a glaring dearth of studies exploring the relationship between general HRQOL and workforce productivity. More significantly, according to the review of literature, there is no existing study that tries to examine the relationship between general HRQOL and workforce productivity by including an empirical study. In addition, presenteeism is a relatively young and undeveloped field of study, so any new study that includes it as a parameter is a significant contribution to that knowledge base. A secondary question, also a relatively unexplored area, is that concerning the relationship of job function to HRQOL effects on productivity.

3.2 Research Questions

The following are research questions have been formulated for this study based on needs that were identified in the review of literature:

1. Is there a relationship between HRQOL and productivity, and, if so, how can it be described?
2. Is there a difference between the effect of physical and mental HRQOL components on productivity? How does this change depending on job type (e.g., knowledge-based versus production-based jobs, traditional vs. flexible time)?
3. Do the role physical and emotional scales from the SF-36 have a stronger correlation to productivity than the other SF-36 scales?

4. Is there a correlation between increased productivity to make up for productivity loss and social functioning?
5. What implications does this relationship, both theoretical and practical, have to policy for program decision makers?

3.3 Hypotheses

Based on the review of literature, the following hypotheses are proposed:

Hypothesis 1: There is a large, positive correlation between HRQOL, as measured by the SF-36 physical and mental summary measure scores; and workforce productivity, as measured by absenteeism and presenteeism. This hypothesis is based on existing evidence that physical and mental health have a direct correlation with one's productivity. That is, the more physically and mentally healthy one is, the more productive one is.

Hypothesis 2: There is a stronger correlation between MCS than PCS scores of the SF-36 and workforce productivity for the sample population as a whole. This hypothesis is based on the supposition that work productivity in an academic environment is more directly tied to one's mental than physical health, as measured by the SF-36. Based on the review of the literature, thus far, there have not been any similar studies conducted.

Hypothesis 3: For production-based jobs, there will be a stronger correlation between PCS than MCS scores of the SF-36 and workforce productivity. For knowledge-based jobs, the opposite result will hold. This hypothesis is based on the fact that production-based jobs typically involve more physical activity than knowledge-based

jobs, so the productivity of people employed in the former should be more directly tied to their physical than mental health compared to those employed in the latter.

Hypothesis 4: There is a statistically stronger correlation between productivity and the role physical (RP) and role emotional (RE) scales than the other SF-36 scales. This hypothesis is based on the supposition that the RP and RE questions in the SF-36, which attempt to describe the physical and mental health impact on work and daily activities, contain questions that are more directly related to productivity, and thus should have a higher correlation to productivity than the other scales.

Hypothesis 5: There is a statistically stronger correlation between productivity and the vitality (VT) scale than the other SF-36 scales. This hypothesis is based on the supposition that the vitality scale, which describes one's energy and fatigue level, contains questions that are more closely related to productivity than the other scales.

Hypothesis 6: Respondents who work extra hours to catch up on work have statistically lower social function (SF) scales than those who do not. This hypothesis is based on the presumption that those who work extra hours to catch up on work may not have as much time or energy to spend on social activities, which would affect their SF scale score negatively.

3.4 Methodology

A descriptive, cross-sectional and group-comparison study of the working population was implemented via an online survey distributed via email among randomly selected faculty and staff at the Georgia Institute of Technology, a major research university in Atlanta. This target population was chosen based on the fact that it offers a

wide variety of job positions and job types within a single employer. In addition, the setting offered the ability to get IRB (institutional review board) approval and carry out a study at a low cost and in a relatively short amount of time. The survey questionnaire used, named the Health-Related Quality of Life and Work Productivity Questionnaire (HQWP), is presented as Appendix D. HQWP attempts to examine the relationship between HRQOL and workforce productivity in a general (not disease-specific) setting by including components of both types of questions in a single instrument.

The HRQOL component of HQWP comprise the 36 questions of the standard (4-week recall) SF-36 survey, version 1. The SF-36 was chosen because it is the most well-known and best validated generic instrument to measure HRQOL, as described earlier. Moreover, as a generic instrument, the SF-36 measures different aspects of health status simultaneously without merging results into a single measure. In the same way, productivity is a multi-faceted measure that provides value when its components can be distinctly identified, so using a generic HRQOL is a more natural fit. Research on the SF-36 community website revealed that order effects when combining the SF-36 with another survey are minimized when the SF-36 questions are placed first,[15] so once the decision was made to use the SF-36, it was determined that those questions should be the first to appear in HQWP. In addition, other studies have been published regarding the advantages of supplementing the SF-36 with more precise general and specific measures.[83-85] Usage and scoring of the SF-36 required a current license and user registration with QualityMetric, Inc., which HSI had obtained previously for other studies.

The workforce productivity component of HQWP consists of the MHPQ survey plus a few additional questions to help improve validity of the results. To select

questions to use to measure productivity, we examined the 16 surveys found in the literature and sought to identify one to two salient surveys from which to select the majority of these questions. The MHPQ and WPAI surveys were the only productivity instruments found in the literature that meet all of the following properties: measures productivity in relation to general health (not specific to any condition), measures both absenteeism and presenteeism, includes concrete and comprehensive existing evidence of validity and reliability in peer-reviewed literature, is not limited by type of profession, and is available for use as public domain or through a free licensing agreement. In addition, the MHPQ was the only survey found which includes a question on extra hours spent to catch up on work, which is relevant to Hypothesis 6. Other unique features of the MHPQ that made them more attractive than WPAI (and the other productivity surveys) included questions regarding work-related accidents, injuries, successes, and failures, which, though relatively rare, are relevant measures because they have high indirect costs (or rewards) which could be translated directly to cost-savings related to treatment.[45]

MHPQ was an ideal length compared to the HPQ and WPAI (20 versus 92 and 6 questions), since its length was more comparable to that of the SF-36. In addition, the HPQ asked questions about specific disease conditions and already included questions related to HRQOL, so MHPQ was determined to be a better option than HPQ. Thus, to increase content validity by drawing questions from as few surveys as possible, all productivity questions were selected from the MHPQ. Moreover, each of the 6 questions found in the WPAI survey was found to be already covered by MHPQ or unnecessary for

purposes of this study. Specifically, Table 3.1 below describes each question from WPAI and addresses how each question is or is not covered by HQWP.

Table 3.1: Coverage of WPAI Questions

WPAI Question (Figure B.12.A)	**How Covered by HQWP Survey (Appendix D)**
1. Are you currently employed (working for pay)?	Not covered, but included in HQWP Part 2 Instructions (recipient was requested not to fill out productivity questions if not working)
2. During the past seven days, how many hours did you miss from work because of your health problems?	MHPQ 1-8 (included)
3. During the past seven days, how many hours did you miss from work because of any other reason, such as vacation, holidays, time off to participate in this study?	Not covered (not relevant to study)
4. During the past seven days, how many hours did you actually work?	Not covered specifically, but presenteeism is measured using Likert scale questions in MHPQ 23
5. During the past seven days, how much did your health problems affect your productivity while you were working?	Covered in the SF-36 4, 5, 6, 8
6. During the past seven days, how much did your health problems affect your ability to do your regular daily activities, other than work at a job?	

A few minor changes were made to the MHPQ in incorporating it into HQWP. MHPQ was included in its entirety except for question 16 ("How would you rate the usual job performance of most workers on your job?"), which was replaced by two questions regarding the perceptions of peers and supervisors ("Using the same 0 to 10 scale, how would your peers rate your job performance during the past 7 days?" and "Using the same 0 to 10 scale, how would your supervisor rate your job performance during the past

7 days?"). These questions were added specifically to help reduce social desirability bias, which would tend to cause over-reporting or overestimation of productivity. Arguments found in Shikiar et al. (2004) supported the use of similarly-worded questions as a method to help alleviate such errors.[50] For these added questions, a 7-day recall period was chosen over a 30-day recall period to help reduce recall bias.

Finally, the HQWP survey included standard, non-identifying demographic questions for gender, age group, ethnicity, and job classification. Although this study was designed primarily to study characteristics of a population at large as well as to compare the productivity levels by job type, classification information was used to provide further relevant statistical analysis and discussion. Table 3.2 below summarizes how the HQWP survey was created.

Table 3.2: HQWP Formulation

HQWP Question	**Source**	**Notes**
1-11	SF-36 1-11 (entire survey)	No changes were made
12-25a	MHPQ 1-14a	No changes were made
25b, 25c	New Questions	Added to reduce social desirability bias
26-29	MHPQ 17-20	No changes were made
30-36	New Questions	Demographic questions

To deal with the concerns of the accuracy of self-reported productivity estimates, several issues were addressed. First of all, the recall period for the instrument was set for most questions to be 4 weeks, to be consistent with that found in much of the literature and the recall period used for the standard SF-36 (i.e., the questions used in the first half of HQWP). The most significant deviation from the 30-day recall period was with some

of the extra questions asked related to presenteeism to reduce social desirability bias. Specifically, a 7-day recall period was used to ask questions regarding peer and superior perceptions of work productivity.

Web surveys were selected for implementation because they allow respondents the opportunity to take the survey at their leisure and do not suffer from interviewer effects. In addition, web-based surveys offer statistically shorter response times and faster data collection,[86] lower monetary and time costs,[86] and better data accuracy.[87] Also, though the population at a university is a generally computer literate, survey recipients were given the option to complete an equivalent paper-based survey, alleviating accessibility issues related to web-based surveys.[33]

3.5 Parameters of Interest

In order to test the hypotheses, several parameters were identified. The input parameters include HRQOL parameters from the SF-36: PCS and MCS, as well as the PCS scale scores of Role Physical (RP), Bodily Pain (BP), and General Health (GH); and the MCS scale scores of Vitality (VT), Social Function (SF), Role Emotional (RE), and Mental Health (MH). Demographic information used as inputs includes gender, age group, race, occupation, and job type. Each of these inputs is a categorical variable except age group, which was measured as an ordinal variable.

The output parameters identified for this study include both absenteeism- and presenteeism-related parameters. Absenteeism was measured as a continuous variable in terms of total days lost (a sum of whole days and portion of days) in the past 30 days (A30) and as a binary variable describing whether a person has missed any work in the

past 30 days (AB30). Presenteeism was measured in several different forms. The primary form was a self-reported evaluation of job performance in the past 30 days on a scale of 0 to 10, where 0 is the worst job performance anyone could have at one's job and 10 is the performance of a top worker. This whole number was then translated into a decimal between 0 and 1 by dividing by 10 to arrive at our final measure of presenteeism (P30). Other secondary measures included a 7-day version of the primary (self-assessed) form (P7), as well as measures of how the respondent thought a peer (PP7) and supervisor (PS7) would evaluate her job performance in the past 7 days. Note that P30, P7, PP7, and PS7 will sometimes be referred to as direct presenteeism measures in this thesis.

We also created a summary measure, the average of the scores for the 9 parts of question 23, which were denoted presenteeism-related questions (PRQ). Note that PRQ was a distinct measure from the other productivity variables and included questions such as, "How much of the time did you have trouble getting along with others at work?" and "How much of the time did you do no work at times when you were supposed to be working?". Finally, separate measures were created to describe other presenteeism related concepts in the survey: how often productivity was lower than expected in the past 7 days (Prod7); whether or not a special success or achievement, major work failure, mistake, or missed deadline occurred in the past 30 days (AM); how often in past 30 days health problems limited work (HLW); number of extra hours worked in past 30 days, measured as a continuous variable (EH) and a binary variable (EHB). Table 3.3 below summarizes the input and output parameters used in this study and where they can be found in the HQWP survey, where the main parameters used in the analysis are in bold.

Table 3.3: Summary of Parameters

Type	Name (Abbreviation)	Type (Range)	Description (HQWP Question Number)
Input (HRQOL)	**Physical Component Summary (PCS)**	**Continuous (0–100)**	**Physical health related components of SF-36**
	Physical Function (PF)	Continuous (0-100)	Scale of PCS (3a-3j)
	Role Physical (RP)	Continuous (0–100)	Scale of PCS (4a-4d)
	Bodily Pain (BP)	Continuous (0–100)	Scale of PCS (7, 8)
	General Health (GH)	Continuous (0–100)	Scale of PCS (1, 11a-11d)
	Mental Component Summary (MCS)	**Continuous (0–100)**	**Mental health related components of SF-36**
	Vitality (VT)	Continuous (0–100)	Scale of MCS (9a, 9e, 9g, 9i)
	Social Function (SF)	Continuous (0–100)	Scale of MCS (6, 10)
	Role Emotional (RE)	Continuous (0–100)	Scale of MCS (5a-5c)
	Mental Health (MH)	Continuous (0–100)	Scale of MCS (9b-9d, 9f, 9h)
Input (Demogr.)	Gender	Categorical	Male or female (30)
	Age Group	Ordinal (6 groups)	Age group (31)
	Hispanic	Binary (0 or 1)	Whether Hispanic or not (32)
	Race	Categorical	Race category (33)
	Occupation	Categorical	Occupation type (34)
	Job Type	Categorical	Faculty or staff (35)
Output (Absenteeism)	**Continuous Absenteeism (A30)**	**Continuous (0–30)**	**How much work respondent has missed in past 30 days (12, 16, 17)**
	Binary Absenteeism (AB30)	Binary	Whether respondent has missed work in past 30 days (12, 16)
Output (Presenteeism)	**30-Day Presenteeism (P30)**	**Ordinal (0–100)**	**Self-assessment of job performance in past 30 days (25a)**
	7-Day Self Presenteeism (P7)	Ordinal (0–100)	Self-assessment of job performance in past 7 days (25b)
	7-Day Peer Presenteeism (PP7)	Ordinal (0–100)	How a peer would assess job performance in past 7 days (25c)
	7-Day Supervisor Presenteeism (PS7)	Ordinal (0–100)	How a supervisor would assess job performance in past 7 days (25d)
	Presenteeism-Related Questions (PRQ)	**Continuous (0–100)**	**Average of 9 questions related to presenteeism (23a-23i)**
	7-Day Productivity (Prod7)	Continuous (0–100)	How often in past 7 days productivity lower than expected (26)
	Achievements or Mistakes (AM)	Ordinal (0–100)	Whether a special success or failure was experienced in past 30 days (27-29)
	Health Limits Work (HLW)	Ordinal (0–100)	How often in past 30 days health problems limited work (24)
	Extra Hours (EH)	Continuous (0–140)	Number of extra hours worked in past 30 days (20-22)
	Binary Extra Hours (EHB)	Binary	Whether respondent has worked extra hours in past 30 days (20-22)

Finally, Table 3.4 below explains which parameters are used to test each of the 6 hypotheses.

Table 3.4: Mapping of Hypotheses to Parameters

Hypothesis	**Parameters Used**
Hypothesis 1	SF-36 physical and mental summary measure scores (PCS, MCS), absenteeism (A30), presenteeism (P30, PRQ)
Hypothesis 2	SF-36 physical and mental summary measure scores (PCS, MCS), absenteeism (A30), presenteeism (P30, PRQ)
Hypothesis 3	SF-36 physical and mental summary measure scores (PCS, MCS), absenteeism (A30), presenteeism (P30, PRQ), job type
Hypothesis 4	SF-36 role physical (RP) and role emotional (RE) scales, absenteeism (A30), presenteeism (P30, PRQ)
Hypothesis 5	SF-36 vitality (VT) scale, absenteeism (A30), presenteeism (P30, PRQ)
Hypothesis 6	SF-36 social function (SF) scale, extra time (EH)

3.6 Data Collection

Originally, the survey was designed to compare the results between production- and knowledge-based employees at Georgia Tech (GT). However, due to the extremely low survey response of the production-based workers in the first batch of surveys, the survey design and sample pool were changed to reflect a group comparison between GT faculty and staff who were knowledge-based (white-collar) employees. Batches 1 and 2 were completed within a week of each other, before the response rate of the survey could be estimated accurately. We soon realized that our estimated response rate of 50% was a significant overestimation of the actual response rate, which was closer to 30%, and that we would not receive enough responses to reach our target amount of 200 responses in the faculty and staff groups. Thus, we calculated the estimated number of additional

faculty and staff we needed to survey for Batch 3 based on the response rates of Batches 1 and 2 as explained below in the Results chapter.

Survey recipients were randomly drawn from a faculty and staff population of 5,005 faculty and staff who had email addresses (as of June, 2007). We noted that 167 GT faculty and staff had unlisted email addresses and 10 members of the faculty and staff were either GT Health Systems or GT IRB faculty or staff who may have presented a conflict of interest with the study, so these GT employees were not included in the population of 5,005 from which recipients were chosen.

The survey was approved by GT IRB in early June 2007 and was later sent out in three batches of emails with a link to an online survey. Recipients of the email were given the opportunity to receive an equivalent paper copy of the survey to be returned via campus or U.S. mail instead of completing the survey online. For each of the 3 batches, a single email reminder was sent out 3 to 4 days before the survey deadline of 20 days after the initial email invitation. The survey itself was hosted on a third-party website, Surveykey.com, which offers 128-bit SSL (Secure Sockets Layer) data protection. In the survey invitation, both online and paper survey participants were advised that filling out and submitting the survey denoted providing informed consent.

As an incentive for recipients to complete the survey, a lottery was held for four 30 GB 5th Generation Video iPods (MSRP $250 each). To comply with Georgia state laws and GT IRB guidelines, all GT employees who received the email were eligible to enter the drawing, irrespective of completion of the survey, by submitting their name and contact information at the end of the survey. Likewise, employees who requested a paper version of the survey were eligible to enter the same drawing, whether or not the survey

was completed, by submitting their name and contact information on the last page of the paper-based survey. Four lottery winners were chosen randomly from all participants who submitted their name and at least one method of contact (email address, phone number, or both) by the appropriate survey deadline.

3.7 Data Cleaning

Upon survey closure, survey responses, anonymous feedback, and personal information submitted for lottery purposes were imported into three separate data files. In addition, the feedback and lottery data files were randomly sorted to ensure complete severance of personal data from survey data. All survey data collected were cleaned and imported into a format suitable for statistical analysis. To ensure good quality data, respondents who did not complete at least 20 of the 25 survey questions were excluded from analysis. Since demographic questions were asked at the survey conclusion, it was not possible to draw conclusions about the types of respondents who began but did not complete the survey.

SF-36 results were compiled and scored under the QualityMetric user license mentioned earlier. In addition, productivity covariates, with the exception of 30-day absenteeism (days absent from work) and extra hours worked, were calculated and normalized to a scale from 0 to 100. Since we were using absenteeism and presenteeism as baseline covariates, 0 was coded as perfect productivity (no work loss or impairment) and higher numbers signified less productive measures for all productivity measurements.

Pairwise exclusion of missing or ambiguous data was used for statistical analysis. An example of an ambiguous response to “How many days in the past 30 days did you

either come in late for work or leave early?" is "doctor's appointment." In addition, any user-entered response of "n/a" to a question asking for a number (e.g., "How many days in the past 30 days did you either come in late for work or leave early?") was assumed to be 0. For entries that consisted of a range (e.g., 3 to 4), the average of the range (e.g., 3.5) was used for the data value. If a user wrote an answer like "at least 10", the smallest value given (e.g., 10) was used. In a few cases, values were calculated based on the user-entered text (e.g., "once to twice a week" or "five") rather than numerical values.

Finally, several responses for race entered manually were changed to fit the categories offered. For example, a response that read "Southeast Asian" was changed to the "Asian" category, while responses that included multiple races were placed into a new "Multiracial" category. If race was not clear from the response given, the response was simply ignored. To ensure privacy and confidentiality, personal information entered for purposes of the drawing was separated from survey responses, and this information was completely destroyed upon completion of the drawing.

CHAPTER 4

RESULTS

4.1 Descriptive Demographic Statistics

All statistical analyses were performed in SPSS 16.0 (Chicago, IL). A total of 434 responses were collected over 3 batches, each of which spanned 20 consecutive days during the summer months (June, July, and August) of 2007. The overall survey completion rate was about 27.1% (out of 1600 emails sent), with Batches 2 and 3 having a much higher response rate than Batch 1, as explained below. A total of three paper surveys were requested from the email recipients, but only one was returned. Thus, of the 434 total surveys completed, 433 were completed online, while 1 paper survey was completed and returned via campus mail. Table 4.1 below summarizes the original survey response rate for the 3 batches.

Table 4.1: Original Survey Response Rate

Batch	Survey completed	Survey began, not completed	Survey not completed	Total
Batch 1	91 (18.2%)	23 (4.6%)	386 (77.2%)	**500**
Batch 2	62 (31%)	16 (8%)	122 (61%)	**200**
Batch 3	281 (31.2%)	64 (7.1%)	555 (61.7%)	**900**
Total	**434 (27.1%)**	**103 (6.4%)**	**1063 (66.4%)**	**1600**

Originally, the study was intended to compare knowledge-based (white-collar) with production-based (blue-collar) workers, where types of workers were sampled based

on official job titles. Thus, for Batch 1, production- and knowledge-based workers were sampled relatively evenly. However, production-based workers had a markedly lower survey response rate (2.2%) than knowledge-based workers (31.0%), as shown in Table 4.2 below. While the response rate for knowledge-based workers seems fairly reasonable given the method and timing of the survey, the low response rate for production-based workers could be due to factors such as less familiarity with the internet, minimal use of computers and the internet while on the job, and distrust of the security of web-based surveys.

Table 4.2: Batch 1 Survey Response Rate

Type of employee	Survey completed	Survey not completed	Total
Production-based	5 (2.2%)	218 (97.8%)	**223**
Knowledge-based	86 (31.0%)	191 (61.0%)	**277**
Total	**91 (18.2%)**	**409 (81.8%)**	**500**

Due to the vast differences in survey completion rates between the two worker types, it was determined that an online survey would not be an effective means of collecting data for production-based workers and knowledge-based workers simultaneously. Since the other option, administering a new, paper-based survey, would not have been possible on a large scale given the limited timeframe and budget of our study, the survey design was changed from a production- versus knowledge-based study to a faculty versus staff group-comparison study within the knowledge-based employees. Thus, subsequent samples (Batches 2 and 3) did not include any production-based

workers. In addition, further analyses excluded the production-based workers surveyed in Batch 1. In doing so, the survey completion rates across all 3 batches were, in fact, found to be nearly identical (around 30%). A table summarizing the revised survey response rates is shown below in Table 4.3.

Table 4.3: Revised Survey Response Rate

Batch	Survey completed	Survey not completed	Total
Batch 1	86 (31.0%)	191 (69.0%)	**277**
Batch 2	62 (31.0%)	138 (69.0%)	**200**
Batch 3	281 (29.6%)	669 (70.4%)	**950**
Total	**429 (30.1%)**	**998 (69.9%)**	**1427**

As mentioned earlier in the Methodology chapter, a third batch was necessary in order to balance the number of faculty and staff respondents and to achieve our goal of 200 responses in each group. We were able to estimate the number of additional faculty and staff we needed to survey by calculating the number of additional responses necessary and by calculating the average response rates for the faculty and staff groups in the first two batches.

For the faculty, we received 53 responses in Batches 1 and 2, meaning that we wanted to get 144 responses in Batch 3. In addition, the estimated faculty response rate from Batches 1 and 2 was 47% (53 out of 113). As a conservative estimate and to account for the fact that the third batch would take place around the end of the end of the

summer semester, we estimated the response rate to be the overall response rate of 30%, resulting in a target population of 490 faculty to survey, which we rounded up to 500. For the staff, we received 94 responses in Batches 1 and 2, meaning that we wanted to get 106 responses in Batch 3. The estimated staff response rate from Batches 1 and 2 was 25.8% (94 out of 364). Since the end of the summer semester was not predicted to affect the staff response rate for Batch 3 as much, we estimated the response rate to be 23.5%, resulting in a target population of 450. Thus, we sent out Batch 3 to approximately 900 faculty and staff.

Demographic information collected from survey participants included gender (male or female); age group (18 – 24, 25 – 34, 35 – 44, 45 – 54, 55 – 64, or 65+); ethnicity (Hispanic or non-Hispanic); race (White, Black or African American, Asian, American Indian, or Multiracial); occupational classification (professional and technical; managerial; administrative and clerical; service; production, construction, operations, maintenance, material handling; or sales and related); and job type (faculty or staff). Note that the "production, construction, operations, maintenance, material handling" occupational classification (which includes all production-based workers) was removed after Batch 1 since only knowledge-based faculty and staff were included in the final results. Table 4.4 below summarizes all demographic information obtained from the surveys.

Table 4.4: Summary of Demographics

	Number (%) of Responses			
	Batch 1	**Batch 2**	**Batch 3**	**Total**
Gender				
Female	49 (57.0%)	22 (35.5%)	106 (37.7%)	**177 (41.3%)**
Male	37 (43.0%)	39 (62.9%)	174 (61.9%)	**250 (58.3%)**
Missing	0 (0%)	1 (1.6%)	1 (0.4%)	**2 (0.5%)**
Age Group				
18 - 24	4 (4.7%)	0 (0%)	2 (0.7%)	**6 (1.4%)**
25 – 34	17 (19.8%)	16 (25.8%)	41 (14.6%)	**74 (17.2%)**
35 - 44	21 (24.4%)	21 (33.9%)	73 (26.0%)	**115 (26.8%)**
45 - 54	24 (27.9%)	16 (25.8%)	88 (31.3%)	**128 (29.8%)**
55 - 64	17 (19.8%)	7 (11.3%)	69 (24.6%)	**93 (21.7%)**
65+	3 (3.5%)	1 (1.6%)	8 (2.8%)	**12 (2.8%)**
Missing	0 (0%)	1 (1.6%)	0 (0%)	**1 (0.2%)**
Hispanic				
Yes	2 (2.3%)	1 (1.6%)	4 (1.4%)	**7 (1.6%)**
No	83 (96.5%)	60 (96.8%)	273 (97.2%)	**416 (98.6%)**
Missing	1 (1.2%)	1 (1.6%)	4 (1.4%)	**6 (1.4%)**
Race				
White	51 (59.3%)	41 (66.1%)	201 (71.5%)	**293 (68.3%)**
Black or African American	22 (25.6%)	12 (19.4%)	35 (12.5%)	**69 (16.1%)**
Asian	7 (8.1%)	7 (11.3%)	31 (11.0%)	**45 (10.5%)**
Multiracial	5 (5.86%)	0 (0%)	5 (1.8%)	**10 (2.3%)**
Missing	1 (1.2%)	2 3.2%)	9 (3.2%)	**12 (2.8%)**
Occupation				
Professional and technical	42 (48.8%)	40 (64.5%)	189 (67.3%)	**271 (63.2%)**
Managerial	17 (19.8%)	13 (21.7%)	54 (19.2%)	**84 (19.6%)**
Administrative and clerical	23 (26.7%)	6 (10.0%)	33 (11.7%)	**62 (14.5%)**
Service	3 (3.5%)	0 (0%)	3 (1.1%)	**6 (1.4%)**
Sales and related	1 (1.2%)	1 (1.6%)	0 (0%)	**2 (0.5%)**
Missing	0 (0%)	2 (3.2%)	2 (0.7%)	**4 (0.9%)**
Job Type				
Faculty	25 (29.1%)	28 (45.2%)	177 (63.0%)	**230 (53.6%)**
Staff	61 (70.9%)	33 (53.2%)	104 (37.0%)	**198 (46.2%)**
Missing	0 (0%)	1 (1.6%)	0 (0%)	**1 (0.0%)**

Many of the differences between the three batches can be explained by the fact that batches 2 and 3 include a higher proportion of faculty, which by nature have a higher percentage of males, whites, and people in professional and managerial positions. In addition, batch 1 had a significantly higher proportion of staff, who are predominantly female. The only other significant difference between the batches was the date of survey administration, as batches took place during different dates spanning 20 days each during the months of June, July, and August, 2007. Under the assumption that the dates of the batches did not have any significant effect on the results, we combined the results for these 3 batches for purposes of further analysis and study.

4.2 Descriptive Health and Productivity Statistics

The first step in our analysis was to look at the histograms for the 2 main input and 3 main output parameters identified: PCS (Figure 4.1), MCS (Figure 4.2), 30-day continuous absenteeism (Figure 4.3), 30-day presenteeism (Figure 4.4), and presenteeism-related questions (Figure 4.5). As expected, histograms for the 2 input measures had a negative skew, while histograms for all 3 output measures had a positive skew. Approximately 3% of respondents had prefect PCS measures, while 0.7% of respondents had perfect MCS measures. The lowest PCS score was 23.125, while the lowest MCS score was 11.875.

30-day absenteeism had the highest amount of positive skew with approximately 56.8% of respondents reporting no 30-day absenteeism, compared to 18.6% reporting no 30-day presenteeism and 1.42% reporting no presenteeism using presenteeism-related questions as an index. We also observed that, while 30-day absenteeism was not as

common as 30-day presenteeism, the former scale offers much finer granularity, as our 30-day presenteeism measure only offers 11 possible values. Finally, it is interesting to note that, though not a commonly accepted measure of productivity, the index using presenteeism-related questions, offers both good discretization and a normal-looking curve, unlike 30-day presenteeism and 30-day absenteeism.

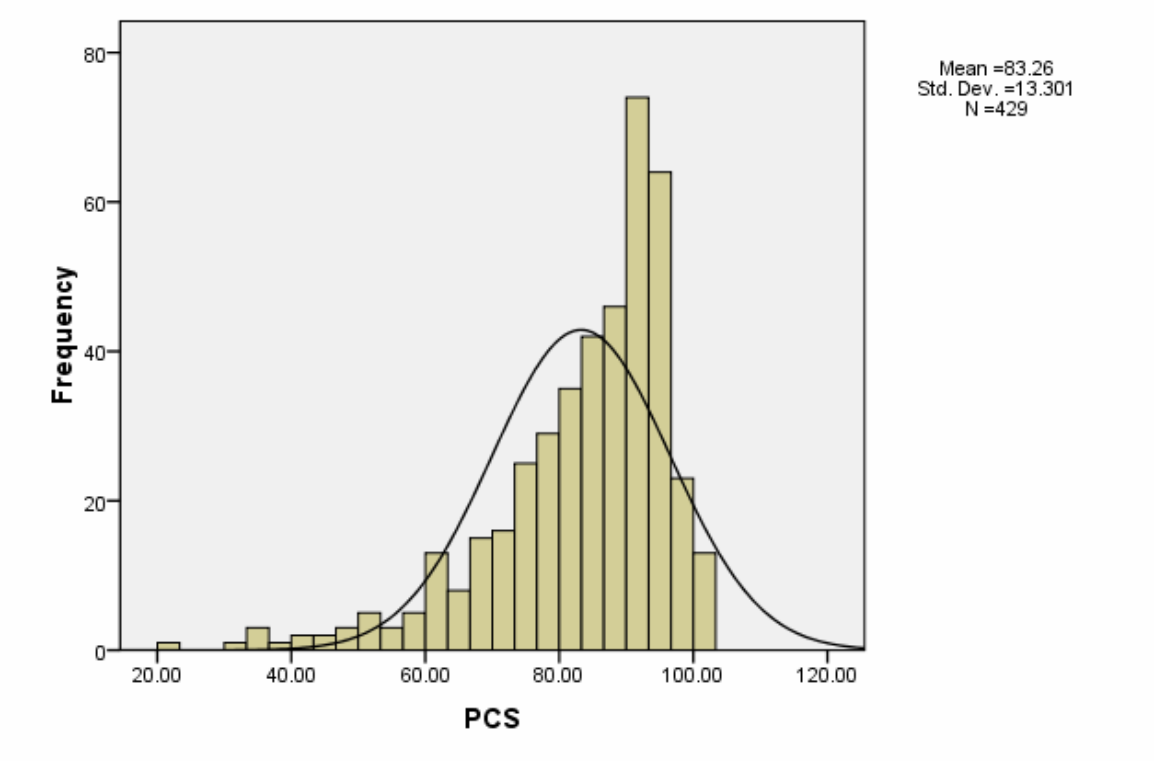

Figure 4.1: PCS Histogram

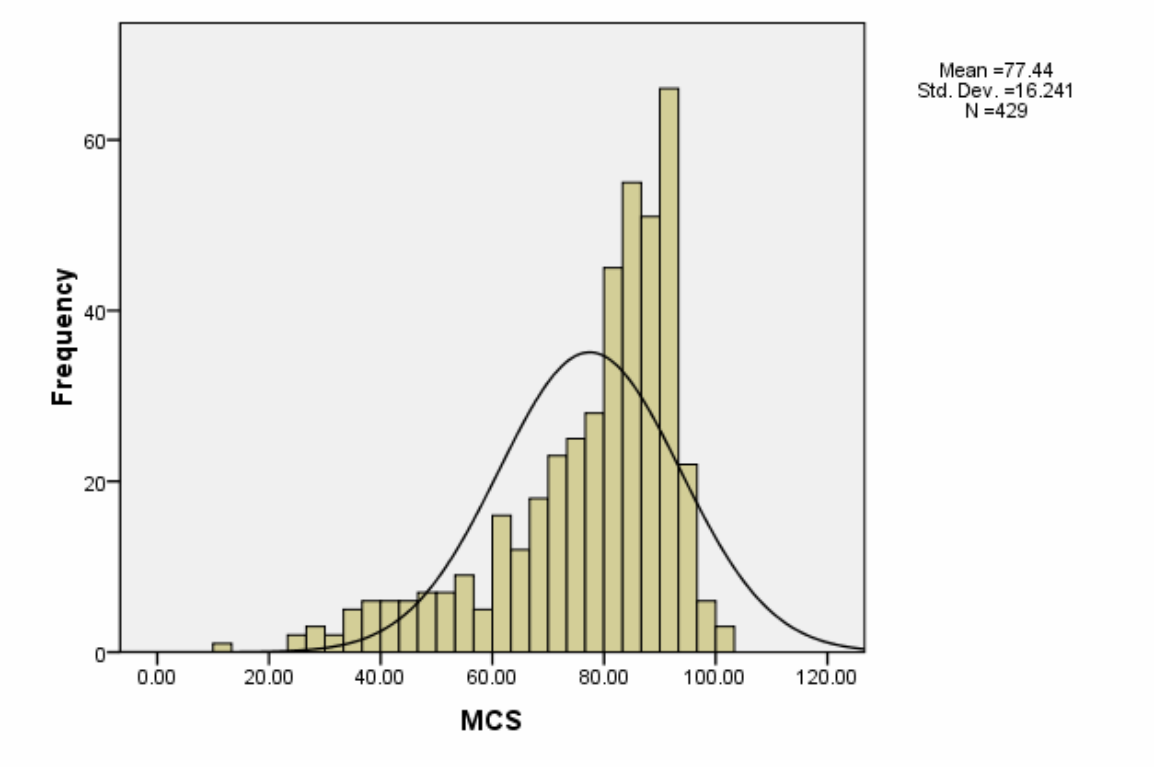

Figure 4.2: MCS Histogram

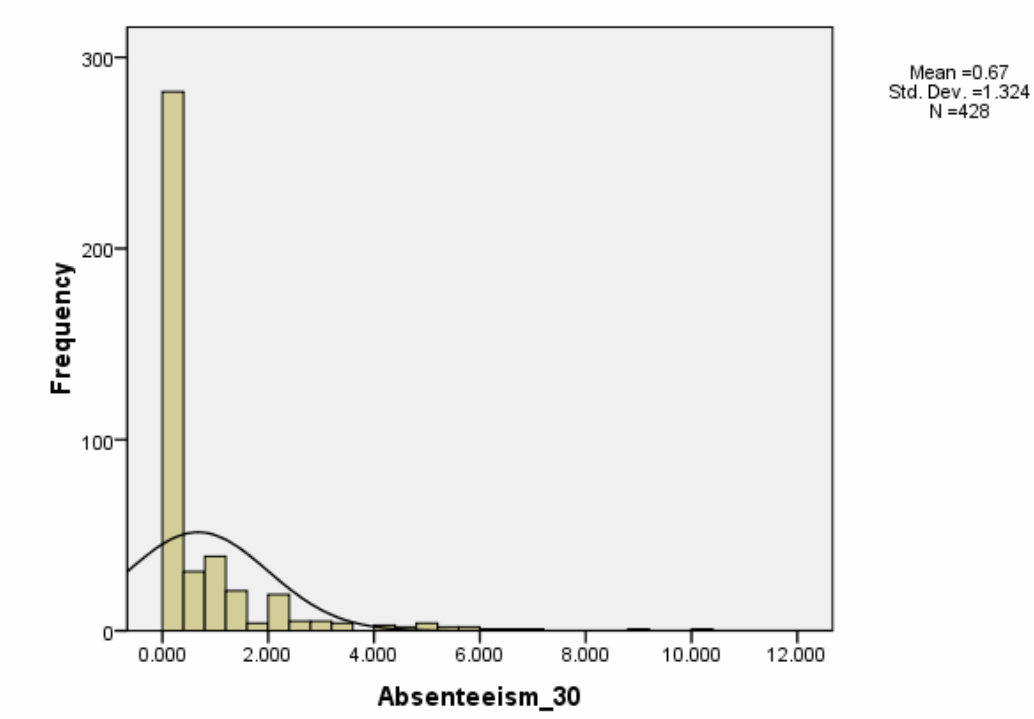

Figure 4.3: 30-Day Absenteeism Histogram

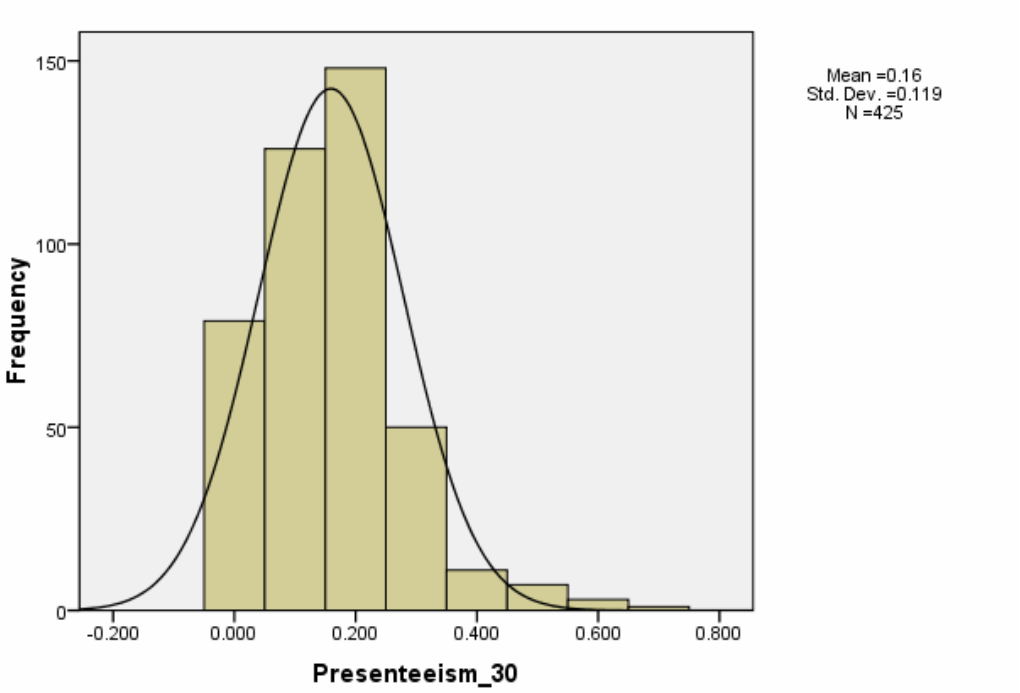

Figure 4.4: 30-Day Presenteeism Histogram

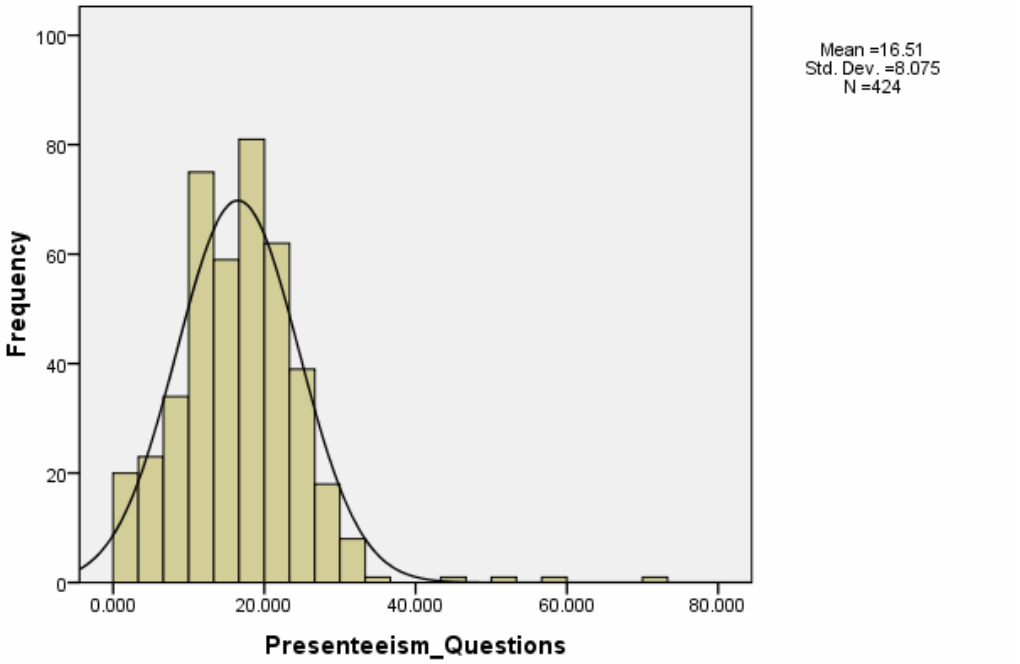

Figure 4.5: Presenteeism-Related Questions Histogram

Next, we ran some preliminary analysis to compare faculty and staff groups. For each of the five major covariates, we compared the results for two major job groups (all faculty versus all staff). We first present the results of the raw data in a summary table, followed by boxplots of each covariate. The raw data summarizing these results are presented in Table 4.5 below.

Table 4.5: Major Covariate Summary

	Average (Standard Deviation) of Responses				
	PCS	**MCS**	**A30**	**P30**	**PRQ**
Faculty (n = 230)	84.57 (12.15)	78.38 (14.72)	0.49 (1.08)	0.17 (0.13)	17.43 (7.97)
Staff (n = 198)	81.67 (14.51)	76.23 (17.82)	0.90 (1.55)	0.14 (0.11)	15.50 (8.10)

To further observe differences between faculty and staff, we looked at the boxplots for each of the two employee groups for all five major covariates, shown below in Figures 4.6 – 4.10 below. When comparing the boxplots for faculty versus staff, there seemed to be only slight differences between the two groups. The most significant difference was that staff seemed to have higher rates of absenteeism (A30) than faculty. Interestingly, however, faculty and staff had similar rates of presenteeism on both scales (measured on a self-reported scale, P30; and using presenteeism-related questions, PRQ). More formal statistical tests comparing faculty and staff are performed in Section 4.5 below.

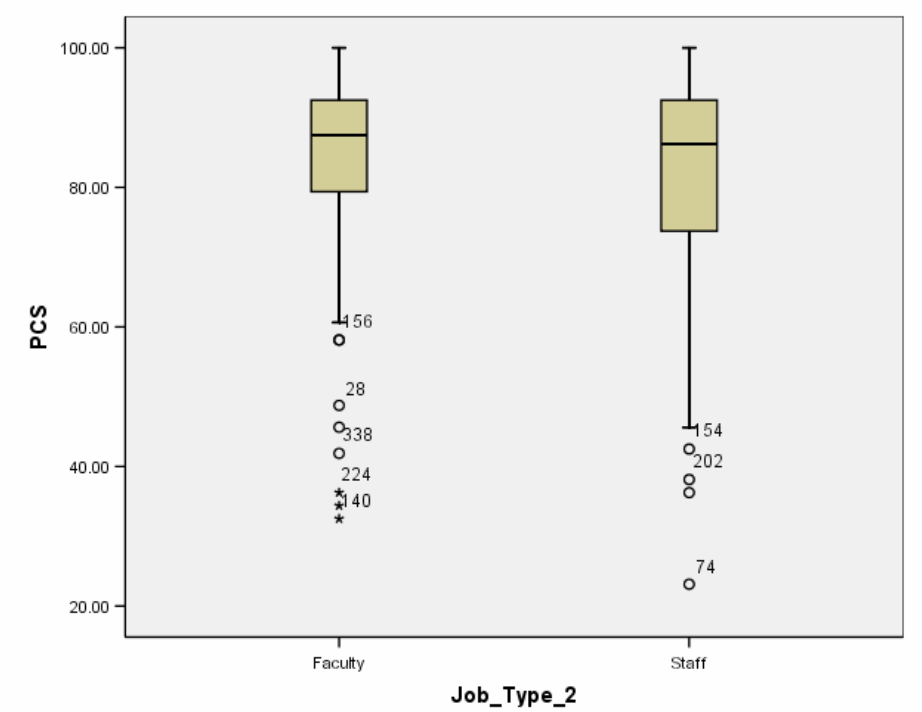

Figure 4.6: PCS Boxplots

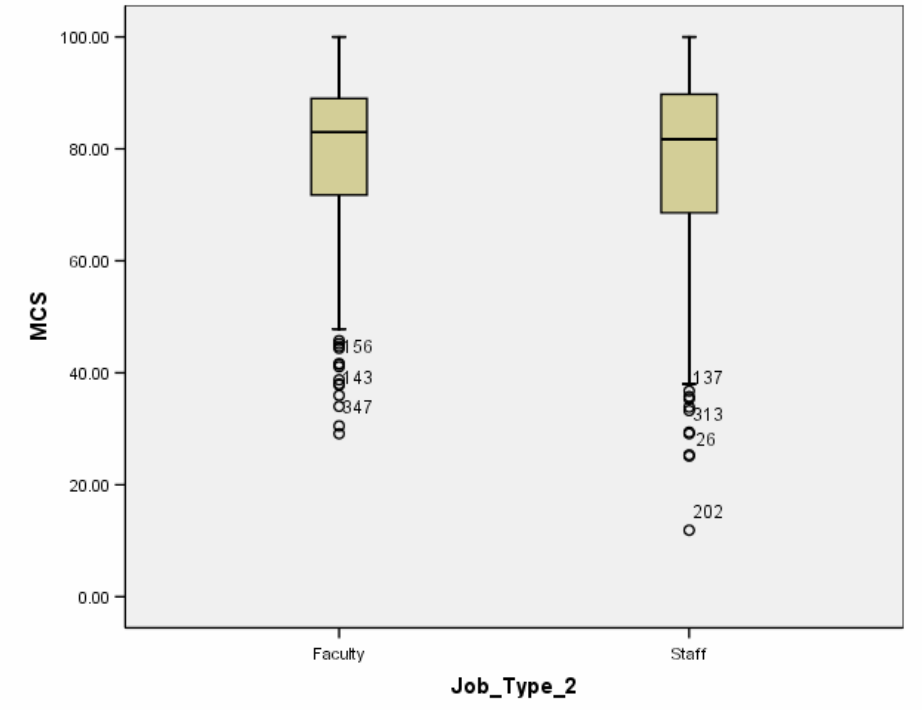

Figure 4.7: MCS Boxplots

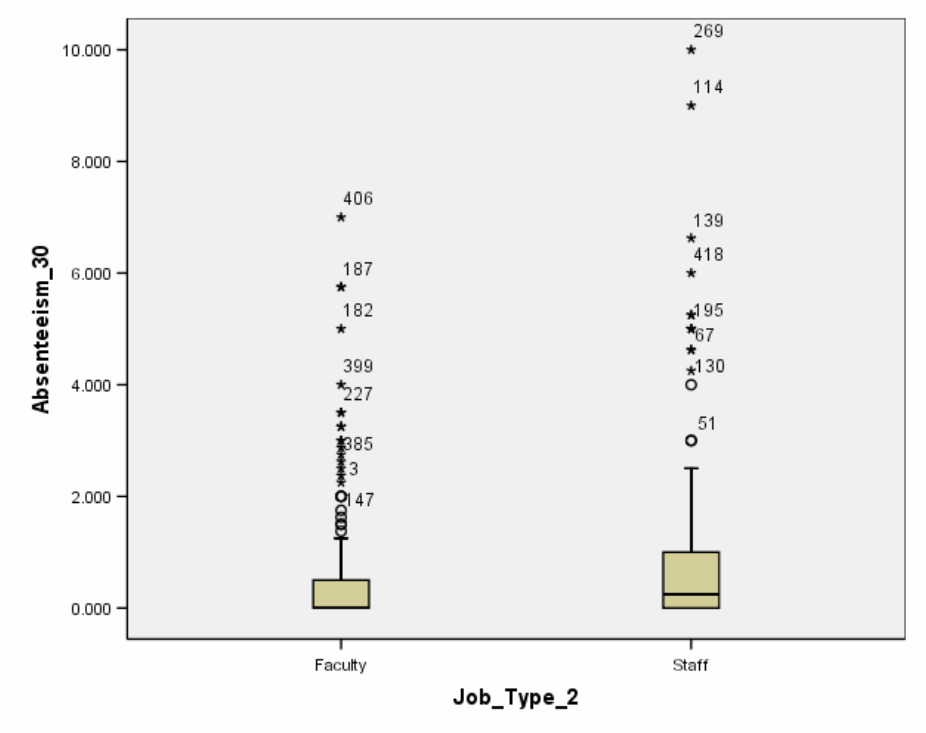

Figure 4.8: 30-Day Absenteeism Boxplots

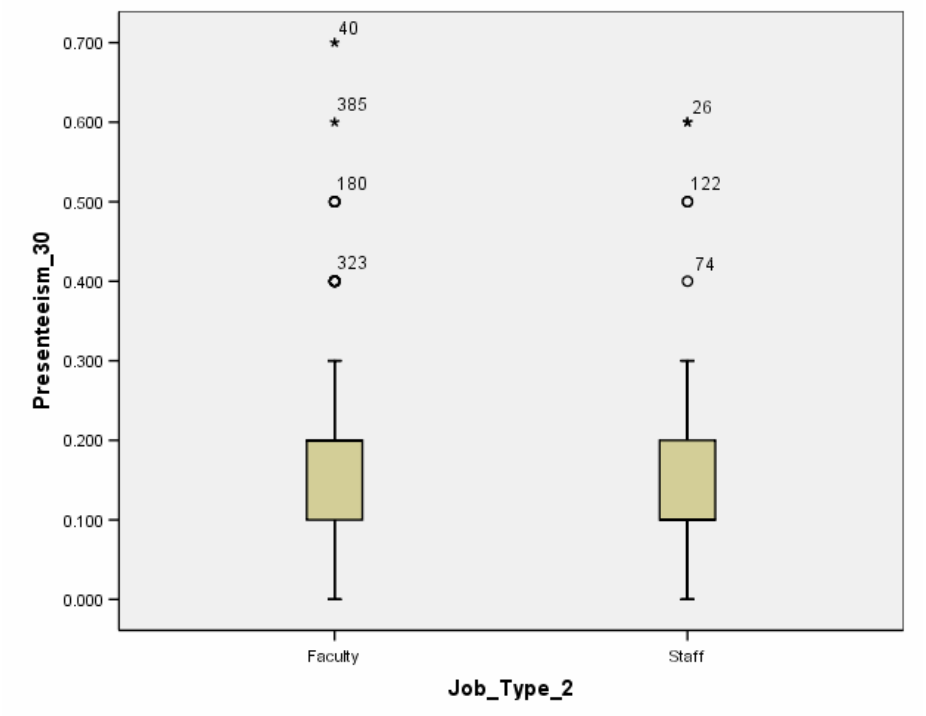

Figure 4.9: 30-Day Presenteeism Boxplots

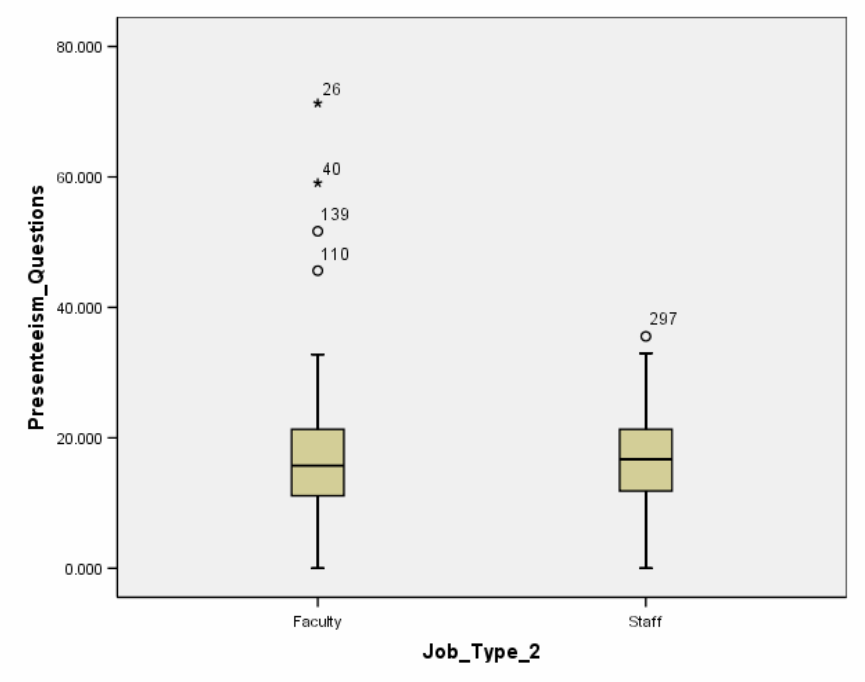

Figure 4.10: Presenteeism-Related Questions Boxplots

Next, we examined scatterplots for our main input parameters, PCS and MCS, versus our main output parameters, 30-day absenteeism, 30-day presenteeism, and presenteeism-related questions. For these scatterplots, we wanted to see not only the relationship between HRQOL and productivity, but also whether there was any difference in the relationship when comparing faculty and staff. For absenteeism, there seemed to be a weak or minimal inverse relationship to PCS and MCS for faculty and staff. On the other hand, presenteeism had a stronger inverse relationship with PCS and MCS, and this relationship was stronger for staff than faculty. Presenteeism-related questions had an even stronger inverse relationship with PCS and MCS scores for faculty and staff. Finally, we noted that staff had a wider range and variance for absenteeism, while faculty had a wider range and variance for presenteeism and presenteeism-related questions. Scatterplots are shown below as Figures 4.11, 4.12, and 4.13.

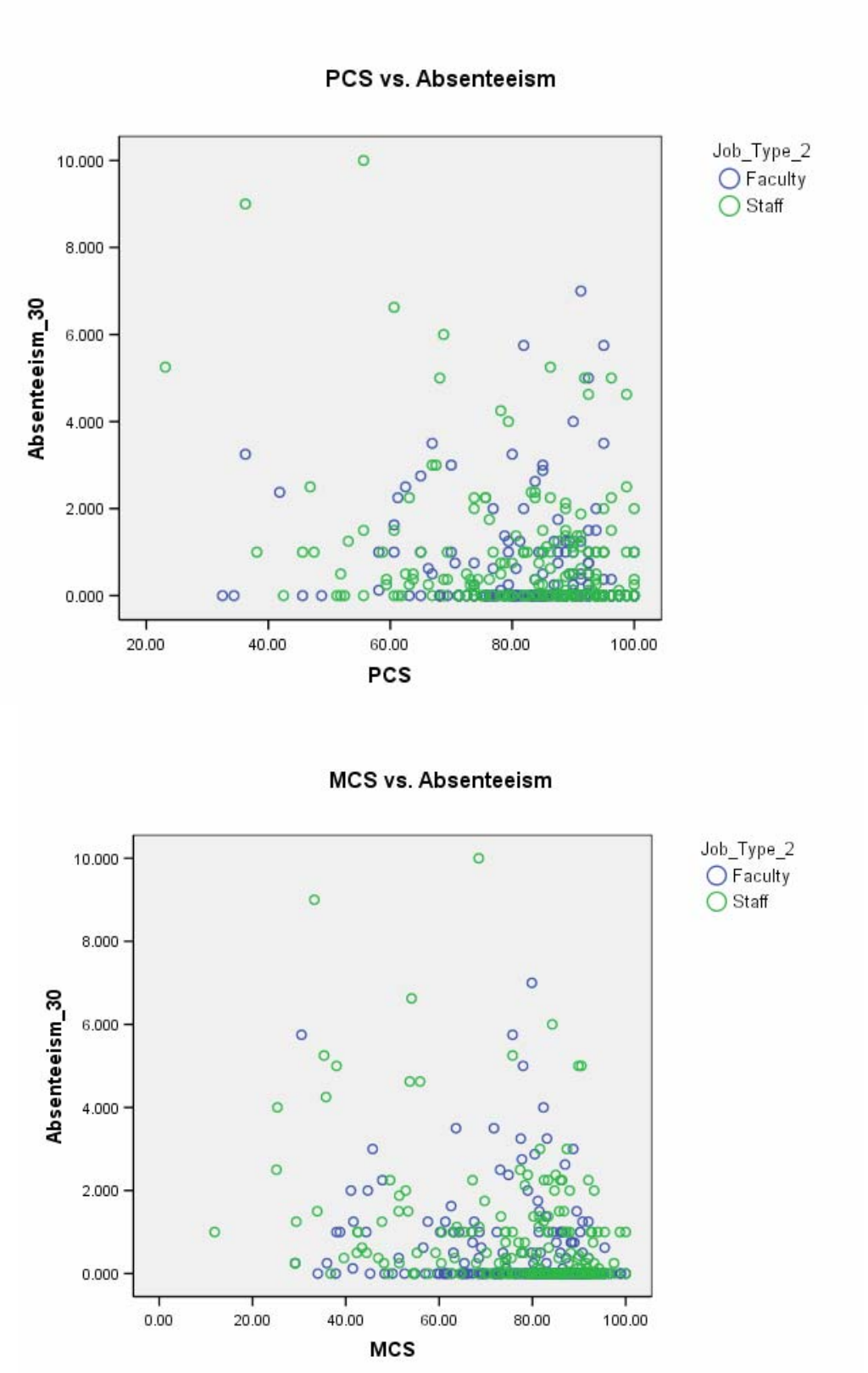

Figure 4.11: PCS and MCS versus 30-Day Absenteeism

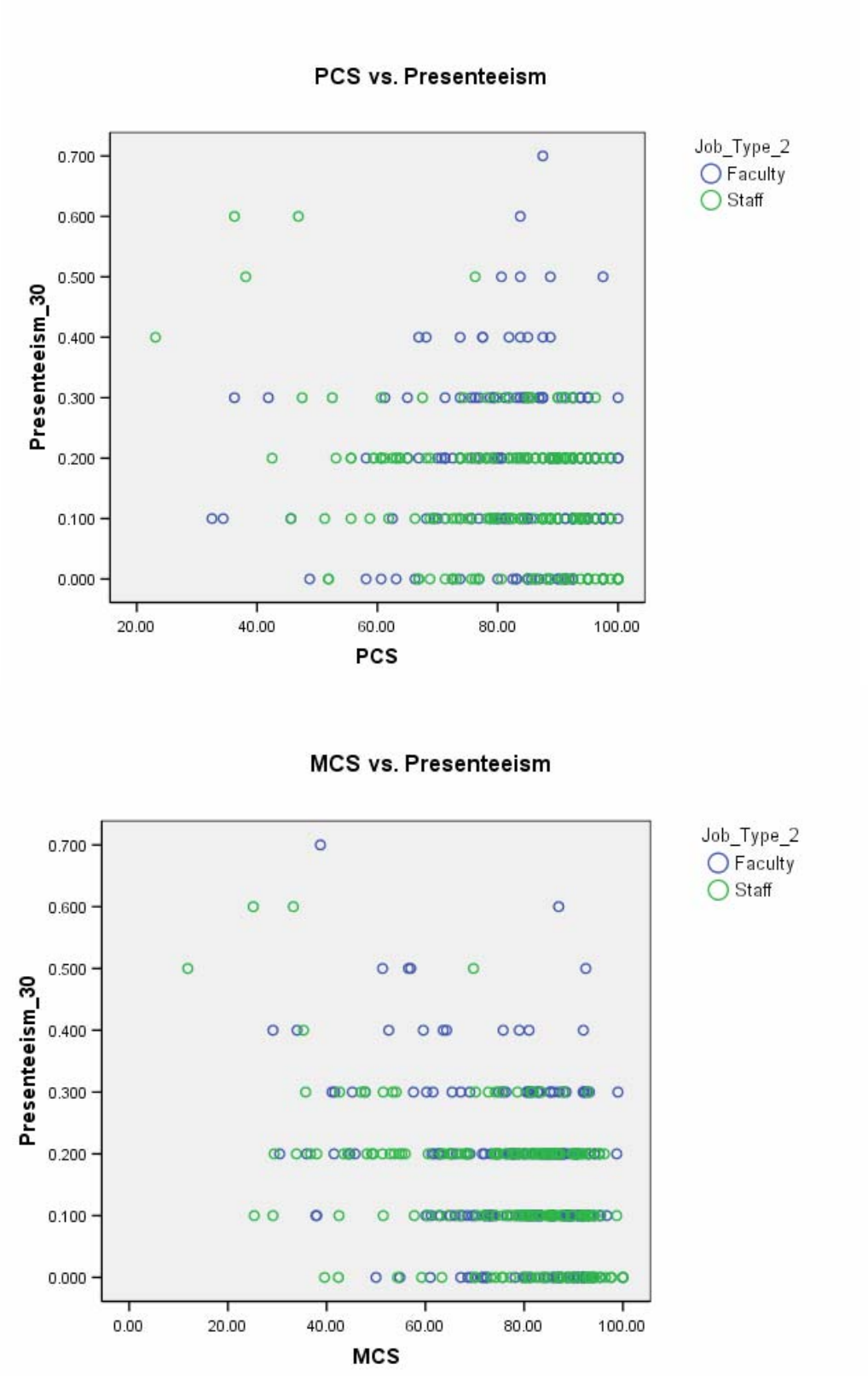

Figure 4.12: PCS and MCS versus 30-Day Presenteeism

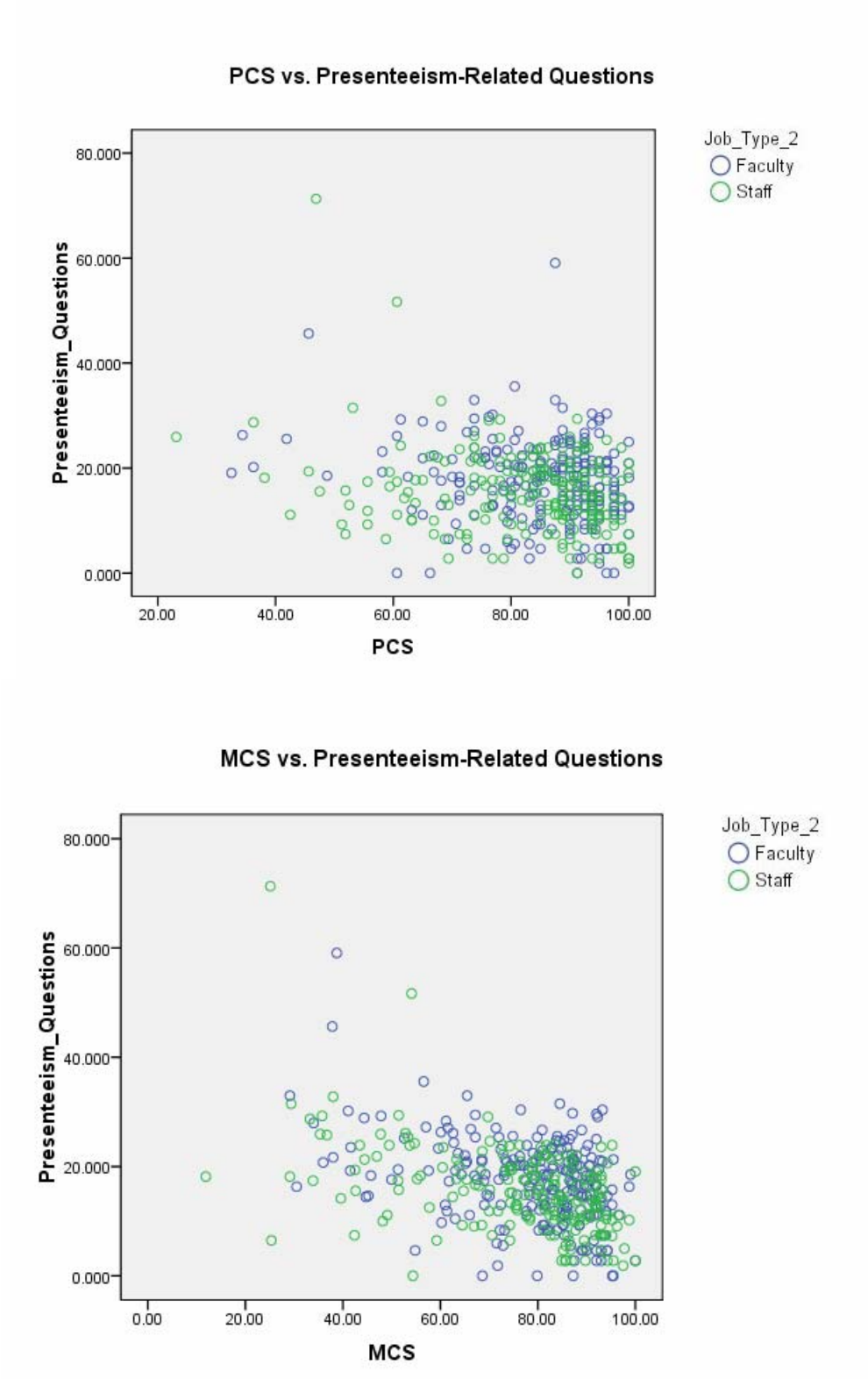

Figure 4.13: PCS and MCS versus Presenteeism-Related Questions

The final preliminary test we performed was to compare differences in input parameters for the faculty and staff groups using an independent samples t-test. We compared the means of not only the major scores, PCS and MCS, but also the 8 scales that comprise the SF-36. Of the 8 scales and 2 summary scores, only bodily pain (BP) and PCS showed a significant difference, with faculty having significantly higher BP and PCS values (i.e., less bodily pain and higher levels of physical health) than staff, as shown below in Table 4.6 (significant differences in bold). Interestingly, none of the MCS scales had significant differences, meaning that the mental health components of faculty and staff were not statistically different.

Table 4.6: Faculty vs. Staff HRQOL Differences

Type	Variable	Mean Value		t	p-value
		Staff	Faculty		
Physical	**Physical Component Summary (PCS)**	**81.7077**	**84.5435**	**-2.179**	**0.030**
	Physical Function (PF)	89.520	90.717	-0.797	0.426
	Role Physical (RP)	85.480	90.109	-1.759	0.079
	Bodily Pain (BP)	**80.290**	**84.348**	**-2.494**	**0.013**
	General Health (GH)	71.540	73.000	-0.859	0.391
Mental	Mental Component Summary (MCS)	76.2988	78.4431	-1.343	0.180
	Vitality (VT)	59.949	62.870	-1.646	0.101
	Social Function (SF)	87.374	89.674	-1.449	0.148
	Role Emotional (RE)	80.135	84.638	-1.441	0.150
	Mental Health (MH)	77.737	76.591	0.761	0.447

4.3 Psychometric Testing

After the survey was implemented, psychometric testing was performed to test reliability and validity of the instrument. Only three questions were added to HQWP that

have not already undergone reliability testing, namely self, peer, and supervisor perceptions of 7-day presenteeism. For these questions, we determined reliability by calculating Cronbach's α coefficient with the closest existing presenteeism measure in MHPQ, 30-day presenteeism (P30). This coefficient was measured to be 0.911, well above the largest minimum accepted threshold of 0.7, so we determined that these additional questions have satisfactory reliability as measures of presenteeism. For PRQ, a productivity construct composed of nine presenteeism-related concepts, the Cronbach's α was measured to be 0.705, just above the minimum accepted threshold. As reported earlier, the MHPQ has already reported consistent reliability across four different occupation groups, so we did not pursue further reliability testing for the productivity section of the HQWP. The SF-36, of course, has also undergone significant reliability testing, so we did not pursue reliability testing for the HRQOL section of the HQWP either. Since HQWP was self-administered, we did not have to worry about inter-rater reproducibility. In addition, instrument responsiveness is not applicable to this study because no type of intervention is being implemented as part of this study.

We now address three categories of validity: content, criterion, and construct validity. Of the 16 productivity instruments found in the literature, only ALWQ and WPSI had any published reports of content validity. Thus, although content validity, the most subjective of these validity measures, was not performed due to time and resource constraints, the fact that all HRQOL and most of the productivity questions came from an existing survey gives a great amount of credibility that the questions in the instrument are reasonable and relevant.

In terms of criterion validity, we first considered whether to pursue measures of clinical validity. Previous studies reported mixed results about how HRQOL and productivity would vary by age group[88, 89] and by gender.[90, 91] In addition, we found no studies that looked at differences in HRQOL and productivity due to ethnicity and job function. Due to time and resource constraints, it was not feasible to perform concurrent validity. Thus, criterion validity measures were not pursued, which is not uncommon among productivity instruments.

Construct validity was assessed by looking at correlations between health (input) and productivity (output) parameters of the questionnaire. As is the standard in HRQOL literature, we used Pearson correlation coefficients for all HRQOL correlations. However, 30-day absenteeism (A30) and extra hours of work (EH) violated the normality assumption given their distributions, and 30-day presenteeism (P30) is an ordinal variable. Thus, we used the Spearman's correlation coefficient for all productivity correlations.

As expected, PCS and MCS correlated strongly with their respective scales, with all Pearson correlations above 0.6 and p-values < 0.001. Also as expected, vitality and social function were the scales of MCS that had the strongest correlations with PCS, and general health was the scale of PCS that had the strongest correlation with MCS. Furthermore, three of the four PCS scales (all except physical function) had moderate correlations with MCS, and three of the four MCS scales (all except mental health) had moderate correlations with PCS. Several of the scales also had moderate correlations with each other, both within and across scales.

Productivity covariates also had many interesting correlations. Specifically, the following covariates had fairly strong Spearman correlations to each other (all pairwise

correlations > 0.65 and p-values < 0.001): 30-day self presenteeism, 7-day self presenteeism, 7-day peer presenteeism, and 7-day supervisor presenteeism. In addition, the following covariates had moderate to strong Spearman correlations to each other (all pairwise correlations between 0.3 and 0.6 and p-values < 0.001): 30-day self presenteeism, 7-day self presenteeism, 7-day peer presenteeism, and 7-day supervisor presenteeism, presenteeism-related questions, and 7-day productivity. Tables 4.7 and 4.8 below show the Spearman correlations for HRQOL (SF-36) and productivity measures, respectively, where the correlations greater than 0.3 (as noted above) are displayed in bold.

Table 4.7: SF-36 Pearson Correlations

Spearman Corr. p-value	PF	RP	BP	GH	PCS	VT	SF	RE	MH
RP	0.229 <0.001								
BP	**0.311** **<0.001**	**0.380** **<0.001**							
GH	**0.329** **<0.001**	0.248 <0.001	**0.331** **<0.001**						
PCS	**0.610** **<0.001**	**0.773** **<0.001**	**0.705** **<0.001**	**0.649** **<0.001**					
VT	0.257 <0.001	0.264 <0.001	**0.394** **<0.001**	**0.510** **<0.001**	**0.498** **<0.001**				
SF	0.179 <0.001	**0.317** **<0.001**	**0.400** **<0.001**	**0.306** **<0.001**	**0.437** **<0.001**	**0.471** **<0.001**			
RE	0.174 <0.001	0.423 <0.001	0.174 <0.001	0.240 <0.001	**0.397** **<0.001**	**0.366** **<0.001**	**0.497** **<0.001**		
MH	0.099 <0.001	0.159 <0.001	0.209 <0.001	**0.397** **<0.001**	**0.304** **<0.001**	**0.637** **<0.001**	**0.546** **<0.001**	**0.477** **<0.001**	
MCS	0.227 <0.001	**0.401** **<0.001**	**0.349** **<0.001**	**0.434** **<0.001**	**0.519** **<0.001**	**0.734** **<0.001**	**0.762** **<0.001**	**0.835** **<0.001**	**0.792** **<0.001**

Table 4.8: Productivity Spearman Correlations

Spearman Corr. p-value	A30	P30	P7	PP7	PS7	PRQ	Prod7	AM	HLW
P30	0.194 <0.001								
P7	0.156 0.001	**0.782** **<0.001**							
PP7	0.117 0.018	**0.753** **<0.001**	**0.767** **<0.001**						
PS7	0.120 0.015	**0.712** **<0.001**	**0.704** **<0.001**	**0.808** **<0.001**					
PRQ	0.124 0.011	**0.524** **<0.001**	**0.461** **<0.001**	**0.461** **<0.001**	**0.462** **<0.001**				
Prod7	0.196 <0.001	**0.433** **<0.001**	**0.454** **<0.001**	**0.358** **<0.001**	**0.358** **<0.001**	**0.461** **<0.001**			
AM	0.164 0.001	0.181 <0.001	0.196 <0.001	0.163 0.001	0.151 0.002	0.088 0.071	0.001 0.978		
HLW	0.312 <0.001	0.207 <0.001	0.175 <0.001	0.105 0.033	0.110 0.026	0.242 <0.001	0.288 <0.001	0.061 0.212	
EH	0.087 0.078	0.038 0.442	0.033 0.503	0.006 0.912	0.031 0.535	0.134 0.007	0.127 0.013	0.101 0.042	0.103 0.038

For further exploration and to get a sense of relationships between our input and output covariates, the last set of correlations we examined were the SF-36 measures versus our productivity measures. Health limits work (HLW) correlated moderately with PCS and MCS, as well as seven of the eight SF-36 subscales. Otherwise, only MCS (with A30, PRQ, and Prod7) as well as MCS subscales (VT with A30, PRQ, and Prod7; SF with A30; RE with Prod7, and MH with PRQ and Prod7) had moderate correlations with any of the productivity measures. Based on correlations alone, these results imply that productivity is more closely tied to mental than physical measures of health, in general. Table 4.9 below shows the Spearman correlations for HRQOL versus productivity measures, where the correlations greater than 0.3 (as noted above) are shown in bold.

Table 4.9: SF-36 vs. Productivity Spearman Correlations

Corr. p-val.	PF	RP	BP	GH	PCS	VT	SF	RE	MH	MCS
A30	-0.065 0.183	-0.204 <0.001	-0.169 <0.001	-0.165 0.001	-0.196 <0.001	**-0.310** **<0.001**	**-0.315** **<0.001**	-0.233 <0.001	-0.215 <0.001	**-0.317** **<0.001**
P30	-0.150 0.002	-0.097 0.047	-0.085 0.082	-0.226 <0.001	-0.189 <0.001	-0.261 <0.001	-0.189 <0.001	-0.238 <0.001	-0.265 <0.001	-0.283 <0.001
P7	-0.152 0.002	-0.037 0.442	-0.072 0.141	-0.205 <0.001	-0.159 0.001	-0.251 <0.001	-0.120 0.013	-0.184 <0.001	-0.258 <0.001	-0.234 <0.001
PP7	-0.121 0.014	-0.013 0.789	-0.060 0.224	-0.204 <0.001	-0.131 0.008	-0.229 <0.001	-0.124 0.011	-0.214 <0.001	-0.233 <0.001	-0.230 <0.001
PS7	-0.076 0.125	-0.031 0.530	-0.005 0.918	-0.144 0.003	-0.074 0.135	-0.189 <0.001	-0.102 0.039	-0.149 0.003	-0.173 <0.001	-0.184 <0.001
PRQ	-0.102 0.036	-0.128 0.008	-0.058 0.234	-0.218 <0.001	-0.168 <0.001	**-0.314** **<0.001**	-0.226 <0.001	-0.283 <0.001	**-0.386** **<0.001**	**-0.371** **<0.001**
Prod7	-0.089 0.076	-0.133 0.008	-0.154 0.002	-0.214 <0.001	-0.212 <0.001	**-0.304** **<0.001**	-0.233 <0.001	**-0.321** **<0.001**	**-0.353** **<0.001**	**-0.377** **<0.001**
AM	-0.055 0.259	-0.022 0.653	0.037 0.448	-0.080 0.100	-0.041 0.405	-0.157 0.001	-0.071 0.147	-0.076 0.116	-0.071 0.146	-0.109 0.025
HLW	-0.281 <0.001	**-0.448** **<0.001**	**-0.411** **<0.001**	**-0.330** **<0.001**	**-0.498** **<0.001**	**-0.338** **<0.001**	**-0.504** **<0.001**	**-0.306** **<0.001**	**-0.306** **<0.001**	**-0.436** **<0.001**
EH	0.014 0.783	-0.060 0.221	-0.016 0.750	-0.069 0.165	-0.050 0.314	-0.042 0.392	-0.052 0.294	-0.074 0.135	-0.107 0.030	-0.095 0.056

4.4 Hypothesis Testing

The next step was to test the hypotheses posed in Section 3.3. Since our study was changed from a group comparison study of production- vs. knowledge-based workers to faculty vs. staff, we changed Hypothesis 3 to reflect no differences in HRQOL vs. productivity correlation strengths between faculty and staff. Otherwise, the remaining hypotheses remained unchanged. The results of our hypothesis tests follow below.

Hypothesis 1: There is a large, positive correlation between HRQOL, as measured by the SF-36 physical and mental summary measure scores; and workforce productivity, as measured by absenteeism and presenteeism.

This hypothesis was tested using standard correlations between PCS and MCS scores, and the 3 major productivity measures, A30, P30, and PRQ. For absenteeism and

PRQ, we used Pearson correlations, and for presenteeism, we used Spearman correlations (due to the ordinal nature of the measure). All 3 PCS correlations had small correlations (between 0.18 and 0.23 with p-values <0.001), while the 3 MCS correlations had small to medium correlations (between 0.28 and 0.42 with p-values <0.001) with productivity measures. These correlations were not as strong as hypothesized, so Hypothesis 1 is only partially supported. Intuitively, it would seem that there is at least a moderate correlation between HRQOL and workforce productivity, but perhaps certain covariates need to be included in or excluded from the instrument in order to make this relationship stronger.

Hypothesis 2: There is a stronger correlation between MCS than PCS scores of the SF-36 and workforce productivity for the sample population as a whole.

This hypothesis was tested by performing a correlation analysis comparing MCS and PCS scores calculated for Hypothesis 1 for two parametric outcome variables A30 and PRQ. The correlation analysis we performed involved transforming the correlation coefficients with the Fisher Z-transform and using the z-value to determine significance levels.[92] Looking at the raw numbers, the MCS correlations were higher than PCS correlations for each of the major outcome variables. PRQ was found to have correlations with statistically different strengths at the $\alpha = 0.05$ level, although 30-day absenteeism was close, with a p-value of 0.059. Since PRQ is not a standard measure of productivity (unlike absenteeism and presenteeism), we conclude that this hypothesis has somewhat limited support.

Hypothesis 3: For correlations between HRQOL and productivity measures, there will be no difference in correlation strength between faculty and staff.

This hypothesis was tested by performing a correlation analysis on the Pearson correlations using the Fisher Z-transform. Spearman correlations were also calculated for P30 for comparison purposes, although there is no established method for comparing the strength of these correlations. For the faculty and staff groups, we made four comparisons corresponding to the possible combinations of two main input parameters (PCS and MCS) and two parametric output parameters (A30 and PRQ). Of these combinations, the staff did have a significantly higher correlation than faculty between PCS and absenteeism at the $\alpha = 0.05$ level. None of the other comparisons yielded statistically different correlation strengths, however.

A table summarizing the results of the statistical tests used for Hypotheses 1 – 3 is shown below as Table 4.10. Note that all correlations are assumed to be two-tailed.

Table 4.10: Hypothesis 1 – 3 Test Results

Hypothesis	Variables	Test used	Value	p-value	Conclusion
Hypothesis 1	PCS, A30	Pearson	-0.227	<0.001	Small corr.
	PCS, P30	Spearman	-0.189	<0.001	Small corr.
	PCS, PRQ	Pearson	-0.207	<0.001	Small corr.
	MCS, A30	Pearson	-0.290	<0.001	Small corr.
	MCS, P30	Spearman	-0.283	<0.001	Small corr.
	MCS, PRQ	Pearson	-0.418	<0.001	Medium corr.
Hypothesis 2	PCS/MCS, A30	Fisher Z trans.	N/A	0.325	No
	PCS/MCS, PRQ	Fisher Z trans.	N/A	<0.001	Yes
Hypothesis 3	St PCS, A30	Pearson	-0.266	<0.001	Small corr.
	St PCS, P30	Spearman	-0.261	<0.001	Small corr.
	St PCS, PRQ	Pearson	-0.260	<0.001	Small corr.
	St MCS, A30	Pearson	-0.321	<0.001	Medium. corr.
	St MCS, P30	Spearman	-0.359	<0.001	Medium Corr.
	St MCS, PRQ	Pearson	-0.467	<0.001	Medium corr.
	Fa PCS, A30	Pearson	-0.137	0.039	Small corr.
	Fa PCS, P30	Spearman	-0.162	0.015	Small corr.
	Fa PCS, PRQ	Pearson	-0.184	0.005	Small corr.
	Fa MCS, A30	Pearson	-0.234	<0.001	Medium corr.
	Fa MCS, P30	Spearman	-0.228	0.001	Small corr.
	Fa MCS, PRQ	Pearson	-0.393	<0.001	Medium corr.
	St/Fa PCS, A30	Fisher Z trans.	N/A	0.050	Yes
	St/Fa PCS, PRQ	Fisher Z trans.	N/A	0.246	No
	St/Fa MCS, A30	Fisher Z trans.	N/A	0.169	No
	St/Fa MCS, PRQ	Fisher Z trans.	N/A	0.187	No

Hypothesis 4: There is a statistically stronger correlation between productivity and the role physical (RP) and role emotional (RE) than the other SF-36 scales.

This hypothesis was tested by performing a correlation analysis on the Pearson correlations for A30 and PRQ. Again, Spearman correlations were also calculated for P30 for comparison purposes. Compared to the other scales, RP had a statistically stronger correlation strength than PF (physical function), but statistically weaker correlation strength than SF (social function) for 30-day absenteeism at the $\alpha = 0.05$ level. In addition, RP had statistically weaker correlations than VT (vitality), SF, RE, and MH (mental health) for presenteeism-related questions. All other correlations between RP

and the other scales for the main output parameters were non-significant. When compared to other scales, RE had a statistically stronger correlation than PF and weaker correlation than SF for 30-day absenteeism. For PRQ, however, RE had stronger correlation strength than PF, RP, and BP for PRQ. All other correlations between RE and the other scales for the main output parameters were non-significant. Thus, this hypothesis has weak support for RP and some support for RE.

It was interesting to note that, when comparing the correlations between RE and RP, RE had a statistically higher correlation for presenteeism-related questions. This finding is consistent with our general finding that mental health is a stronger predictor of productivity than physical health.

Hypothesis 5: There is a statistically stronger correlation between productivity and vitality (VT) scales than the other SF-36 scales.

This hypothesis was tested by performing a correlation analysis on Spearman correlations using the Fisher Z transform. Compared to the other scales, VT had statistically stronger correlation strength than PF for 30-day absenteeism; and PF, RP, and BP for presenteeism-related questions. Thus, this hypothesis has moderate statistical support, and certainly more support than RE and RP from the previous hypothesis.

Hypothesis 6: Respondents who work extra hours to catch up on work have statistically lower social function (SF) scales than those who do not. This hypothesis was tested using a Spearman correlation for the continuous scale of EH and Wilcoxon-Mann Whitney for the binary scale for EH (EHB). Using a Spearman correlation, SF and EH do not have a significant correlation. Using the Wilcoxon-Mann Whitney test for EHB, however, we do get a significant difference in the social functions of those who do work

extra hours to make up for lost production compared to those who don't. Thus, this hypothesis has significant statistical support when extra hours are considered on a binary but not continuous scale. In other words, whether or not one works extra hours is correlated with social functioning, but the amount of extra hours worked is not.

Results of statistical tests used for Hypotheses 4 – 6 are shown below in Table 4.11, with particular parameters of interest for Hypotheses 4 and 5 shown in bold. In addition, results of the correlation analyses calculated to test Hypotheses 4 and 5 are shown in Tables 4.12 and 4.13 below for 30-day absenteeism and presenteeism-related questions, respectively.

Table 4.11: Hypotheses 4 – 6 Test Results

Hypothesis	Variables	Test used	Value	p-value	Conclusion
Hypotheses 4 and 5	PF, A30	Pearson	-0.038	0.437	No correlation
	PF, P30	Spearman	-0.150	0.002	Small corr.
	PF, PRQ	Pearson	-0.112	0.021	Small corr.
	RP, A30	**Pearson**	**-0.209**	**<0.001**	**Small corr**.
	RP, P30	**Spearman**	**-0.097**	**0.047**	**No correlation**
	RP, PRQ	**Pearson**	**-0.132**	**0.007**	**Small corr**.
	BP, A30	Pearson	-0.225	<0.001	Small corr.
	BP, P30	Spearman	-0.085	0.082	No correlation
	BP, PRQ	Pearson	-0.117	0.016	Small corr.
	GH, A30	Pearson	-0.121	0.012	Small corr.
	GH, P30	Spearman	-0.226	<0.001	Small corr.
	GH, PRQ	Pearson	-0.218	<0.001	Small corr.
	VT, A30	**Pearson**	**-0.233**	**<0.001**	**Small corr**.
	VT, P30	**Spearman**	**-0.261**	**<0.001**	**Small corr.**
	VT, PRQ	**Pearson**	**-0.338**	**<0.001**	**Medium corr**.
	SF, A30	Pearson	-0.349	<0.001	Med. corr.
	SF, A30	Spearman	-0.189	<0.001	Small corr.
	SF, PRQ	Pearson	-0.285	<0.001	Small corr.
	RE, A30	**Pearson**	**-0.191**	**<0.001**	**Small corr**.
	RE, P30	**Spearman**	**-0.238**	**<0.001**	**Small corr**.
	RE, PRQ	**Pearson**	**-0.333**	**<0.001**	**Medium corr**.
	MH, A30	Pearson	-0.182	<0.001	Small corr.
	MH, P30	Spearman	-0.265	<0.001	Small corr.
	MH, PRQ	Pearson	-0.369	<0.001	Medium corr.
Hypothesis 6	SF, EH	Pearson	0.044	0.370	Not significant
	SF, EHB	Wilcoxon-MW	-2.274	0.011	Significant

Table 4.12: 30-Day Absenteeism Correlation Analysis

Fisher Z transform p-value	PF	RP	BP	GH	VT	SF	RE
RP	**0.011**						
BP	**0.005**	0.807					
GH	0.223	0.187	0.118				
VT	**0.004**	0.713	0.902	0.092			
SF	**<0.001**	**0.027**	**0.048**	**<0.001**	0.064		
RE	**0.024**	0.785	0.604	0.295	0.521	**0.013**	
MH	**0.033**	0.682	0.513	0.363	0.437	**0.009**	0.892

Table 4.13: PRQ Correlation Analysis

Fisher Z transform p-value	PF	RP	BP	GH	VT	SF	RE
RP	0.768						
BP	0.941	0.825					
GH	0.114	0.198	0.131				
VT	**<0.001**	**0.001**	**<0.001**	0.059			
SF	**0.009**	**0.020**	**0.011**	0.299	0.394		
RE	**<0.001**	**0.002**	**<0.001**	0.071	0.935	0.441	
MH	**<0.001**	**<0.001**	**<0.001**	**0.016**	0.607	0.172	0.551

4.5 Faculty and Staff Differences

To further explore other differences between faculty and staff, we conducted statistical testing on all output parameters using the Wilcoxon-Mann-Whitney test, which revealed several measures with statistically significant differences. When measuring absenteeism as a continuous and binary variable, staff had significantly higher levels, meaning they had statistically higher rates of being absent from work. However, when measuring presenteeism using 30-day self (P30), 7-day self (P7), 7-day peer (PP7), 7-day supervisor (PS7), and presenteeism-related questions (PRQ), faculty had significantly

higher levels than staff, meaning that faculty were statistically less productive than staff members while at work. Staff also reported significantly higher 7-day productivity (Prod7) and lower levels of extra hours worked when measured on a continuous scale (EH). Productivity covariates that did not exhibit statistically significant differences between the two groups include significant achievements or mistakes (AM), whether or not health limited work in the past 30 days (HLW), and whether or not extra hours were worked in the past 30 days as measured on a binary scale (EHB). The results of these statistical tests are shown below in Table 4.14, with significant differences in bold.

Table 4.14: Faculty and Staff Productivity Differences

Variable	Mean Value		Mean Rank		p-value
	Faculty	Staff	Faculty	Staff	
Continuous Absenteeism (A30)	**0.4842**	**0.8962**	**190.22**	**241.50**	**<0.001**
Binary Absenteeism (AB30)	**0.32**	**0.57**	**189.56**	**242.27**	**<0.001**
30-Day Presenteeism (P30)	**0.1704**	**0.1444**	**223.88**	**199.51**	**0.034**
7-Day Self Presenteeism (P7)	**0.1805**	**0.1439**	**226.00**	**194.78**	**0.006**
7-Day Peer Presenteeism (PP7)	**0.1719**	**0.1311**	**220.93**	**191.58**	**0.009**
7-Day Supervisor Presenteeism (PS7)	**0.1673**	**0.1174**	**221.11**	**187.32**	**0.003**
Presenteeism-Related Questions (PRQ)	**17.3780**	**15.5010**	**229.66**	**192.73**	**0.002**
7-Day Productivity (Prod7)	**15.1628**	**12.8772**	**186.28**	**219.62**	**0.002**
Achievements or Mistakes (AM)	11.1111	14.4781	220.09	202.81	0.082
Health Limits Work (HLW)	4.6843	6.0998	215.45	209.13	0.477
Extra Hours (EH)	**19.6193**	**8.8620**	**226.37**	**181.81**	**<0.001**
Binary Extra Hours (EHB)	0.67	0.58	213.35	196.58	0.088

4.6 Gender Differences

Since we also had a good group sizes for males and females, we conducted a group comparison between the two genders for all output parameters using the Wilcoxon-Mann-Whitney test. Interestingly, these tests yielded the same parameters that had statistically significant differences as the faculty and staff group comparison did.

Specifically, when measuring absenteeism as a continuous and binary variable, males had significantly higher levels, meaning they had statistically higher rates of being absent from work. However, when measuring presenteeism using 30-day self (P30), 7-day self (P7), 7-day peer (PP7), 7-day supervisor (PS7), and presenteeism-related questions (PRQ), males had significantly higher levels than females, meaning that males were statistically less productive than females members while at work. Females also reported significantly higher 7-day productivity (Prod7) and lower levels of extra hours worked when measured on a continuous scale (EH). Productivity covariates that did not exhibit statistically significant differences between the two groups include significant achievements or mistakes (AM), whether or not health limited work in the past 30 days (HLW), and whether or not extra hours were worked in the past 30 days as measured on a binary scale (EHB). These results are not surprising given the previous analysis on faculty vs. staff and the fact that a significant proportion of faculty members in our sample were male (172 of 230, or 74.8%) and a smaller but still considerable proportion of staff members in our sample were female (119 of 197, or 60.4%). Complete results of the statistical tests are shown below in Table 4.15.

Table 4.15: Gender Productivity Differences

Variable	Mean Value		Mean Rank		p-value	Sig.
	Male	Female	Male	Female		
Continuous Absenteeism (A30)	**0.5439**	**0.8637**	**196.53**	**237.38**	**<0.001**	**Yes**
Binary Absenteeism (AB30)	**0.36**	**0.54**	**197.13**	**236.53**	**<0.001**	**Yes**
30-Day Presenteeism (P30)	**0.1721**	**0.1398**	**227.26**	**190.58**	**0.002**	**Yes**
7-Day Self Presenteeism (P7)	**0.1781**	**0.1437**	**226.62**	**188.82**	**0.001**	**Yes**
7-Day Peer Presenteeism (PP7)	**0.1710**	**0.1282**	**225.10**	**181.06**	**<0.001**	**Yes**
7-Day Supervisor Presenteeism (PS7)	**0.1653**	**0.1129**	**223.43**	**179.09**	**<0.001**	**Yes**
Presenteeism-Related Questions (PRQ)	**17.5383**	**15.1508**	**231.39**	**184.78**	**<0.001**	**Yes**
7-Day Productivity (Prod7)	**15.2493**	**12.5784**	**214.37**	**186.16**	**0.010**	**Yes**
Achievements or Mistakes (AM)	12.3306	13.2576	209.82	213.85	0.688	No
Health Limits Work (HLW)	4.3548	6.7714	205.53	221.17	0.084	No
Extra Hours (EH)	**17.5500**	**10.4527**	**216.13**	**189.20**	**0.020**	**Yes**
Binary Extra Hours (EHB)	0.65	0.60	208.57	199.93	0.384	No

4.7 Regression Modeling

In this section, we describe regression modeling approaches used to further analyze relationships between our major input and output variables. Based on the histogram for 30-day continuous absenteeism (A30), binary logistic regression was deemed to be a logical choice for modeling due to the high number of respondents reporting zero absenteeism. For 30-day continuous presenteeism (P30), ordinal regression seemed like a logical choice because we only had eight output values, and there was not a heavy concentration at zero like there was for absenteeism. Lastly, presenteeism-related questions (PRQ), was the only major output covariate for which linear regression was a realistically feasible candidate based on the normal-looking histogram.

For modeling purposes, all ordinal and categorical variables were recoded into indicator variables. These variables included: race (White, Black, Asian, Multicultural, and Other), age group (18-24, 25-34, 35-44, 45-54, 55-64, 65+), and job type

(professional and technical, managerial, administrative and clerical, service, and sales and related). In addition, due to possible collinearity with other terms, MCS and PCS were not included, but all eight scales of the SF-36 were included as possible covariates in our models. Lastly, we adjusted for gender by including it in all regression models.

4.7.1 30-Day Absenteeism: Binary Logistic Regression

Since approximately 56.8% of all respondents had no 30-day absenteeism and there was a wide range of values for this covariate, we decided to model A30 using binary logistic regression. Binary logistic regression does not require the restrictive assumptions of linear regression,[93] which made it an attractive option for modeling this variable in particular. Our logistic regressions were performed using a forward stepwise regression, with an entry criterion of 0.05 and exit criterion of 0.10. For the overall sample, variables entered the model in the following order: VT, administrative and clerical jobs, RE, faculty, PF, and Black. This model had a Cox and Snell r^2 of 0.211 and Nagelkerke r^2 of 0.283, which are rather low values since they are well below 0.5; and a classification percentage of 71.7% and ROC curve area of 0.702, which is good.[93] Of the binary variables in the model, administrative and clerical jobs had the highest odds ratio, with people in those jobs (many of which are likely staff jobs) having 2.794 times the odds of reporting absenteeism than those in other jobs. On the other hand, faculty had 0.615 times the odds of reporting absenteeism than staff, consistent with our result in Table 4.14 that staff had significantly higher levels of absenteeism than staff. Two mental health measures, VT and RE, were found to be significant negative predictors of absenteeism, while one physical health measure, physical function, was found to be a

significant positive predictor of absenteeism. Since the latter was an unexpected result, this suggests possible collinearity with another variable such as race or age group.

Next, we ran the same models for staff and faculty groups separately. For the faculty model, our initial run with all variables yielded four significant predictors, which entered the model in the following order: VT, administrative and clerical jobs, RE, and service jobs. This model had low values for Cox and Snell r^2 of 0.159 and Nagelkerke r^2 of 0.221; and a very good classification percentage of 72.0% and ROC curve area of 0.629. Faculty members in administrative and clerical positions had 11.933 times greater odds of reporting 30-day absenteeism, while those in service positions had 9.497 times greater odds. VT and RE were again found to be significant negative predictors of absenteeism.

For the staff group, variables entered the model in the following order: VT, Black, RE, PF, and professional and technical jobs. This model had values for Cox and Snell r^2 of 0.237 and Nagelkerke r^2 at 0.318, but very good classification percentage of 71.9% and ROC curve area of 0.709. Of the binary variables in the model, Blacks had the highest odds ratio, with 2.770 times the odds of reporting absenteeism than those of non-Black races. Staff members in professional and technical positions had 0.411 times the odds of 30-day absenteeism compared to staff members in other positions. As in the overall sample, VT and RE were found to be significant negative predictors and PF a significant positive predictor of 30-day absenteeism. The results for all three models, including Wald χ^2, 95% confidence intervals for the odds ratios, and p-values, are presented below in Table 4.16.

Table 4.16: Predictors of 30-Day Absenteeism (Binary Logistic Regression)

Model	All	Staff	Faculty
ROC curve area	0.702*	0.709*	0.629
Cox and Snell r^2	0.211	0.237	0.159
Nagelkerke r^2	0.283	0.318	0.221
Classification Percentage	71.7%	71.9%	72.0%
Predictor Wald χ^2 OR (95% CI)	Constant 3.458 4.448 (N/A)*	Constant 1.134 3.390 (N/A)*	Constant 11.865 11.207 (N/A)
	VT 23.339 0.966 (0.952, 0.980)*	RE 13.529 0.975 (0.961, 0.988)*	VT 12.875 0.965 (0.946, 0.984)*
	RE 16.425 0.984 (0.976, 0.992)*	VT 13.521 0.961 (0.940, 0.981)*	RE 5.337 0.987 (0.977, 0.998)
	Administrative and clerical 7.880 2.794 (1.364, 5.725)	PF 9.030 1.040 (1.014, 1.067)	Administrative and clerical 4.716 11.933 (1.273, 111.834)
	PF 4.490 1.018 (1.001, 1.036)	Black 6.464 2.770 (1.263, 6.077)	Service 3.300 9.497 (0.837, 107.732)
	Black 4.128 1.997 (1.025, 3.893)	Professional and technical 5.101 0.411 (0.190, 0.889)	
	Faculty 3.657 0.615 (0.374, 1.012)		

* $p < 0.001$

4.7.2 30-Day Presenteeism: Ordinal Regression

The model chosen for 30-day presenteeism was ordinal regression. A major assumption for ordinal regression is that relationships between dependent variables and the levels of the independent variable are the same for all levels, which we can check for each model by using the test of parallel lines. For the group at large, our Cox and Snell,

Nagelkerke, and McFadden r^2 values were low at 0.204, 0.215, 0.077, respectively, although their interpretation is not as straightforward as the r^2 term is for linear regression.[94] Only two predictors were significant at the 0.05 level, VT and MH. The staff model yielded similar results, with larger but still low Cox and Snell, Nagelkerke, and McFadden r^2 values of 0.268, 0.287, and 0.115, respectively. Again, RE was the only predictor that was significant at the 0.05 level, with an odds ratio of 0.988. The faculty model yielded similar r^2 values but more predictors. The model Cox and Snell, Nagelkerke, and McFadden r^2 values were low at 0.205, 0.215, and 0.073, respectively. There were three significant predictors at the 0.05 level: RE, with an odds ratio of 0.985; female, with an odds ratio of 2.032; and PF, with an odds ratio of 0.980.

For all three models, the test of parallel lines was not significant, so we are justified in using ordinal regression. The significant predictors for all three models, including Wald χ^2, 95% confidence intervals for the odds ratios, and p-values, are presented below in Table 4.17.

Table 4.17: Predictors of 30-Day Presenteeism (Ordinal Regression)

Model	All	Staff	Faculty
Cox and Snell r^2	0.187	0.268	0.205
Nagelkerke r^2	0.197	0.287	0.215
McFadden r^2	0.070	0.115	0.073
Predictor Wald χ^2 OR (95% CI)	RE 8.327 0.989 (0.981, 0.996)	RE 3.884 0.988 (0.975, 1)	RE 7.675 0.985 (0.975, 0.996)
			Female 5.469 2.032 (1.122, 3.684)
			PF 5.356 0.980 (0.963, 0.997)

4.7.3 PRQ: Linear Regression

Linear regression models require four assumptions to be met. First, these models assume linearity of the relationship between inputs and output. We can test for a linear relationship by examining the unstandardized residual plot for PRQ, looking for an average residual around 0, and no curves or clumps of points but rather points that roughly form an ellipsoidal shape.[93] The unstandardized residual plot for PRQ, shown below as Figure 4.14, looks much like an ideal residual plot, except that there are a few outliers in the upper right corner of the plot.

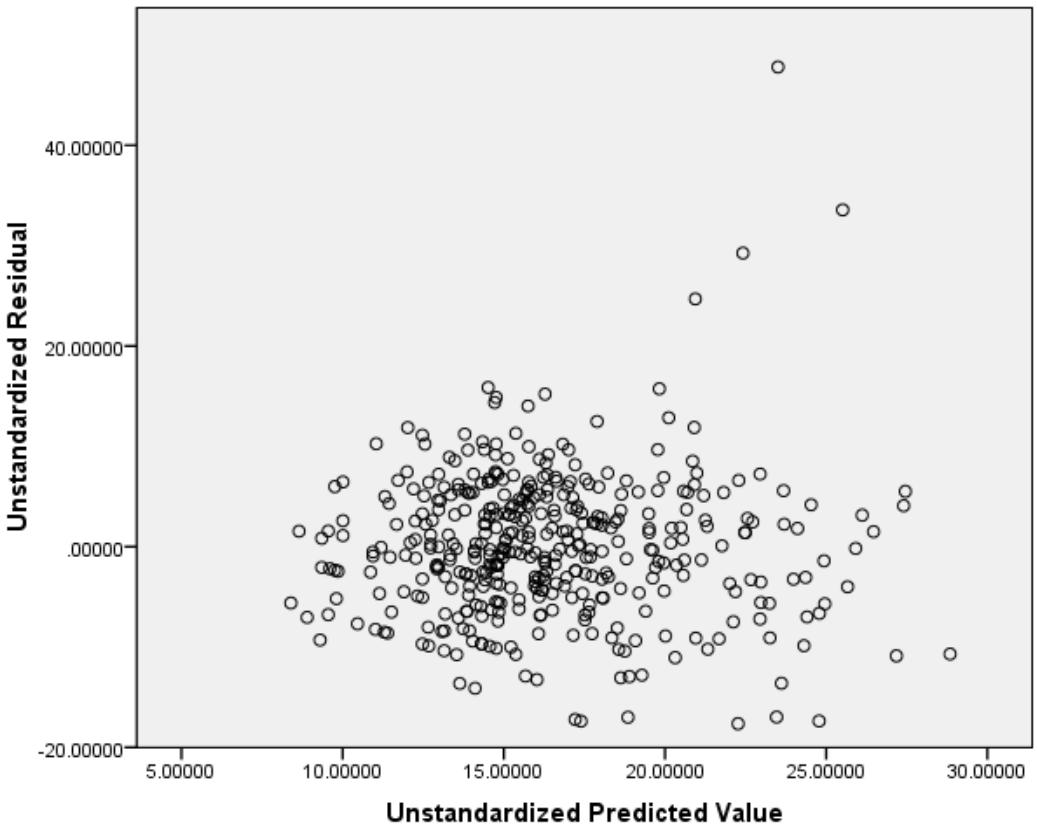

Figure 4.14: Original Unstandardized PRQ Residual Plot

The second assumption for linear regression models is that of constant error term variance, also known as homoscedasticity. We can test for homoscedasticity by examining the studentized residual plot for PRQ, looking for an ellipsoidal shape rather than a diamond- or triangle-shaped plot.[93] The studentized residual plot for PRQ, shown below as Figure 4.15, again looks much like an ideal residual plot, except that there are a few outliers in the upper right corner of the plot.

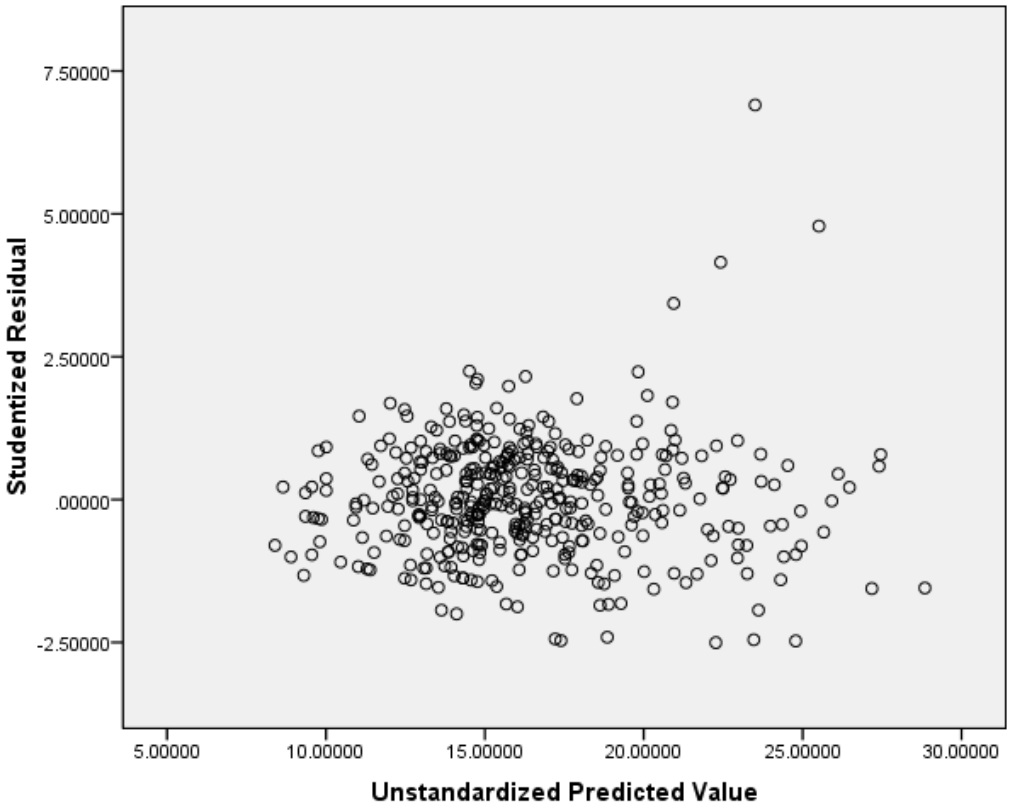

Figure 4.15: Original Studentized PRQ Residual Plot

Thirdly, linear regression models assume independence of error terms. This assumption means that predicted values are not related to other variables such as time or

events.[93] This assumption can be tested using the Durbin-Watson test, with a value of 2 meaning no correlation.[7, 95] Our Durbin-Watson value was 2.048, so this assumption is met also.

Lastly, linear regression models assume normality of error terms, which we can test by examining normal probability plots, or P-P plots. For normality, the standardized residual plot should roughly follow a diagonal line, which represents the normal distribution.[7] The standardized P-P plot for PRQ, shown below as Figure 4.16, fits this last assumption nicely.

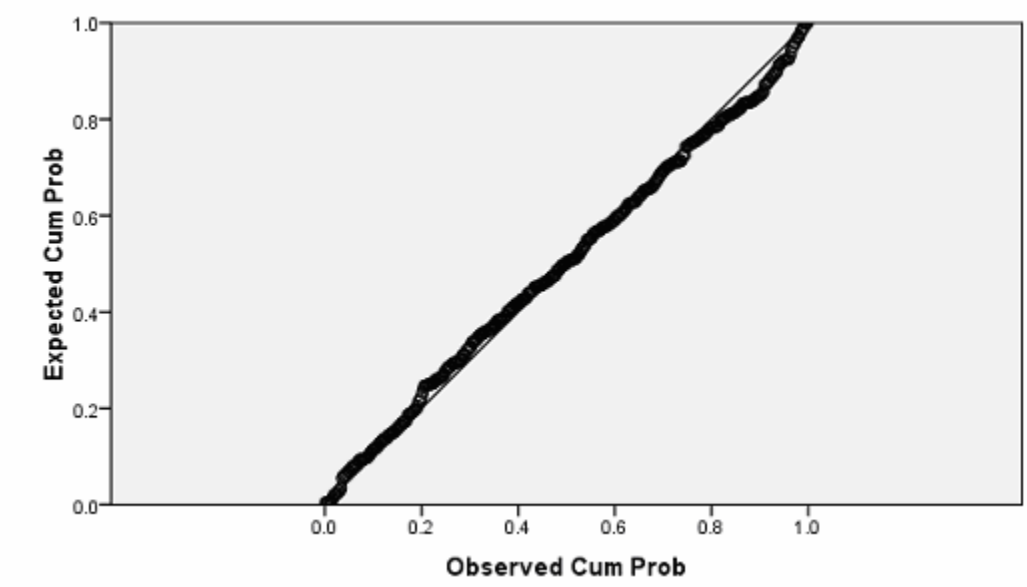

Figure 4.16: Original PRQ P-P Plot

For comparison purposes, we ran the same four tests with the four outliers, the individuals with the four highest PRQ scores (least productivity), removed. Our unstandardized and studentized residual plots (Figures 4.17 and 4.18) more strongly supported the use of linear regression for PRQ and our input covariates.

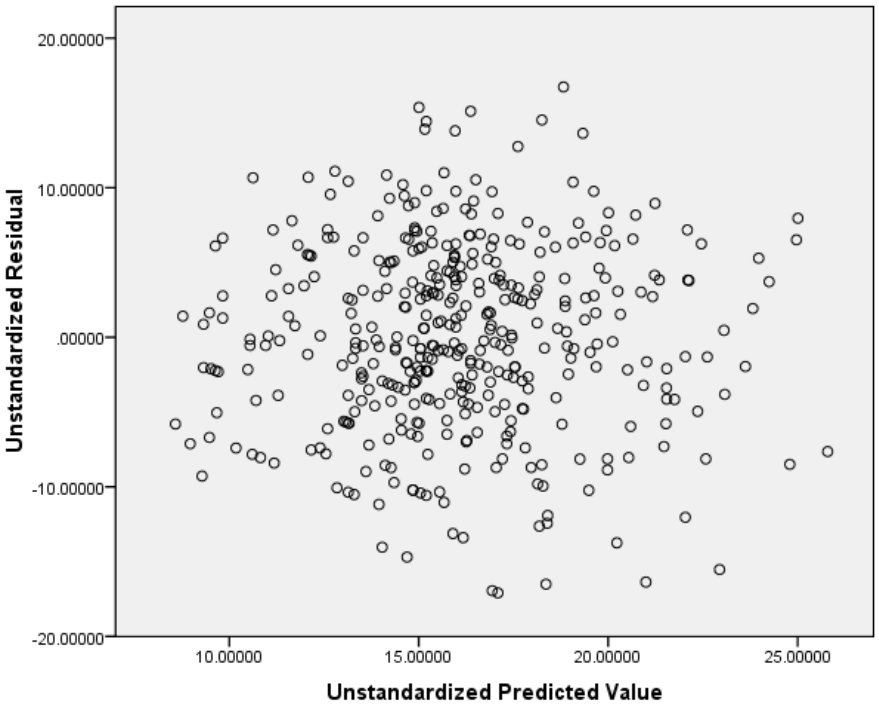

Figure 4.17 Revised Unstandardized PRQ Residual Plot

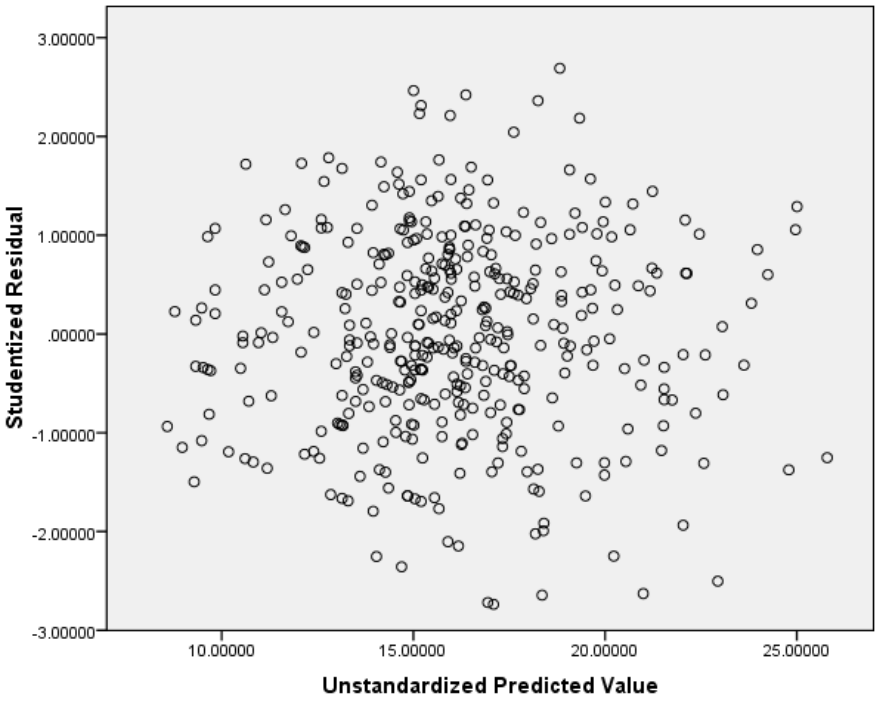

Figure 4.18 Revised Studentized PRQ Residual Plot

In addition, the revised Durbin-Watson value was 2.021, and the revised P-P plot, shown below as Figure 4.19, supports the last assumption as well as the original model.

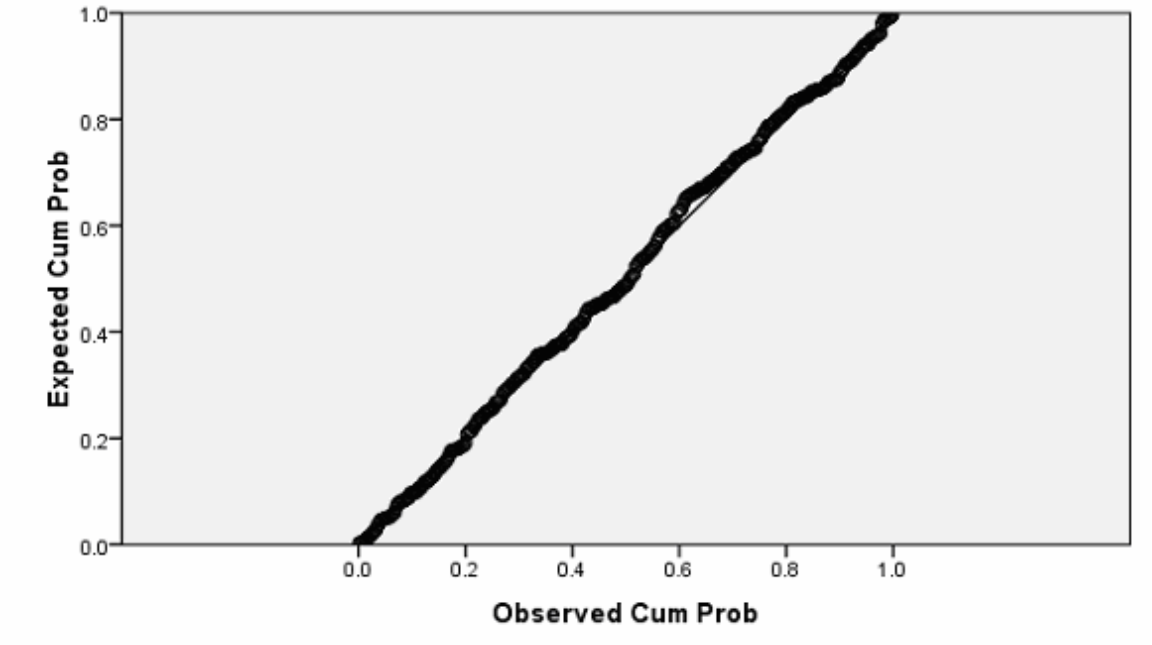

Figure 4.19 Revised PRQ P-P Plot

We report our regression models with and without these outliers as original and revised models, respectively. For both the original and revised data, we modeled the overall sample and then separately looked at faculty and staff groups, which as separate groups also met the linear regression assumptions. To adjust for gender, we first forced the covariate female to enter the model. Then, each model was constructed using a forward stepwise regression, with an entry criterion of 0.05 and exit criterion of 0.10.

For the original data on the overall sample, we adjusted for gender by forcing it into the model, and then other variables entered the model in the following order: MH, RE, and VT. This model had an r^2 of 0.205, and gender was found to be significant even

though we forced it to enter the model. For female employees, the model predicted a 2.629 lower PRQ index (i.e., less work loss in the form of presenteeism) compared to males. Lastly, three mental health measures, MH, RE, and VT, were negative predictors of work loss as measured by PRQ. According to the condition index used for collinearity diagnostics, none of the variables in the model were in danger of being collinear.

Next, we ran the model for staff. After adjusting for gender, variables entered into the model in the following order: SF, Black, GH, and RE. This model had an r^2 of 0.250, and gender was not found to be a significant predictor. For black employees, the model predicted a 2.923 lower PRQ index (i.e., less work loss) compared to non-blacks. Two mental health measures, SF and RE, and one physical health measure, GH, were negative predictors of PRQ. Again, no variables in the model were in danger of being collinear.

Lastly, we ran the model for faculty. After adjusting for gender, predictors entered the model in the following order: VT, RE, and professional and technical jobs. This model had an r^2 of 0.197, and gender was found to be a significant predictor, as it was for the overall sample. For female employees, the model predicted a 2.489 lower PRQ index (i.e., less work loss in the form of presenteeism) compared to males. Two mental health measures, RE and VT, were negative predictors of work loss as measured by PRQ. In addition, those in professional and technical jobs were predicted to have 2.793 higher PRQ index (more work loss) than faculty members who were not. Again, no variables in the model were in danger of being collinear.

The results for all three original models, including 95% confidence intervals for the B coefficients and p-values < 0.001, are presented below in Table 4.18. The revised

models, which excluded the outliers identified earlier, yielded similar predictors. Results for the revised data set are presented below in Table 4.19.

Table 4.18: Original Predictors of PRQ (Linear Regression)

Model	Overall	Staff	Faculty
r^2	0.205	0.250	0.197
Predictor B (95% CI)	Constant 33.054 (29.438, 36.671)*	Constant 35.438 (29.316, 41.560)*	Constant 29.422 (24.813, 34.032)*
	Female -2.629 (-4.057, -1.201)*	SF -0.125 (-0.198, -0.052)	Female -2.489 (-4.702, -0.276)
	MH -0.092 (-0.155, -0.029)	Black -2.923 (-5.239, -0.606)	VT -0.137 (-0.197, -0.077)*
	RE -0.045 (-0.071, -0.020)*	GH -0.062 (-0.119, -0.006)	RE -0.061 (-0.097, -0.026)
	VT -0.075 (-0.126, -0.024)	RE -0.038 (-0.074, 0.000)	Professional and technical 2.793 (0.056, 5.530)

* $p < 0.001$

Table 4.19: Revised Predictors of PRQ (Linear Regression)

Model	Overall	Staff	Faculty
r^2	0.217	0.265	0.176
Predictor B (95% CI)	Constant 29.928 (26.728, 33.128)*	Constant 24.313 (19.610, 29.017)*	Constant 30.966 (26.159, 35.774)*
	Female -1.644 (-3.010, -0.277)	MH -0.062 (-0.135, 0.011)	Female -2.936 (-5.019, -0.854)
	MH -0.092 (-0.147, -0.037)	White 3.285 (1.527, 5.044)*	MH -0.172 (-0.232, -0.111)*
	Black -2.662 (-4.462, -0.862)	RE -0.030 (-0.059, -0.001)	Multiracial 7.182 (1.605, 12.760)
	VT -0.054 (-0.098, -0.009)	VT -0.056 (-0.111, -0.001)	Administrative and clerical jobs -6.961 (-12.627, -1.296)
	RE -0.024 (-0.046, -0.001)		
	Administrative and clerical jobs -2.009 (-3.941, -0.077)		

* $p < 0.001$

4.8 Factor Analysis

Since presenteeism was found in the literature to be an especially difficult construct to measure, we next decided to perform factor analyses to further explore the underlying relationship between individual questions for presenteeism constructs in HQWP that have been normalized (i.e., P30, P7, PP7, PS7, PRQ, Prod7, AM, and HLW). This analysis involved five steps:

1. Determine whether analysis is exploratory or confirmatory and select objectives of the analysis.
2. Determine whether variables or cases are being grouped.
3. Meet statistical and conceptual assumptions regarding appropriateness of factor analysis.
4. Determine whether component factor analysis or common factor analysis should be used and specify the how the number of factors will be determined.
5. Select a rotation method based on whether factors are assumed to be correlated (oblique) or uncorrelated (orthogonal), run the analysis, and interpret the factor matrix.[93]

First we determined that our analysis was exploratory and that our objective in running factor analyses was to summarize data (rather than to reduce data). Next, we determined that we were looking for correlations among variables, which meant that we would be pursuing an R-type factor analysis rather than a Q-type factor analysis or cluster analysis. The next assumption involved meeting conceptual and statistical assumptions

regarding the appropriateness of factor analysis. Knowing there were moderate to strong correlations between some presenteeism covariates (Table 4.8), we had evidence that some underlying structure does exist in our covariates.[93] Statistically, two tests are typically performed to determine the appropriateness of factor analysis. First, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy estimates the degree of distinct and reliable factors in factor analysis, where values of 0.5 – 0.7 are mediocre, 0.7 – 0.8 are good, 0.8 – 0.9 are very good, and 0.9 – 1.0 are superb. In addition, the Barlett test of sphericy is used to find out if there are statistically significant correlations among at least some of the variables, as shown by significant p-values (< 0.05).[96]

Next, we determined that we would use component factor analysis, which is more common than and usually yields results similar to common factor analysis. In addition, we decided first to use the latent root (eigenvalue) criterion to determine the number of factors and then to use the Scree test criterion if latent root criterion did not yield a satisfactory solution. Since there are theoretical grounds for a correlation between HRQOL and productivity factors, we decided to run our factor analyses using oblique rotation models, which, unlike orthogonal rotation models, do permit cross-factor loadings. However, a few sources recommend running both orthogonal and oblique models and comparing them if an oblique model is chosen, so we ran an orthogonal model at the end for comparison purposes. Our chosen software package, SPSS, only offers one oblique rotation method, Oblimin with Kaiser normalization. For the orthogonal model, we decided to use the most commonly used method, Varimax with Kaiser normalization. After running each analysis, we checked for reliability of these factors by calculating Cronbach's α for each component. If any of these were less than

the commonly accepted value of 0.7, we examined the Scree plot to see if a different number of factors should be used, based on the point in the plot where there is a noticeable straightening of the curve.[7] For the presenteeism covariates, the KMO measure of sampling adequacy was 0.861, which is very good. In addition, the Barlett test of sphericy was significant with a *p*-value of <0.001, which supported the use of factor analysis to analyze our data.

4.8.1 Oblique Rotation

We first ran an oblique rotation on the overall data set for the questions that comprise the productivity covariates of P30, P7, PP7, PS7, PRQ, Prod7, AM, and HLW. For oblique rotations, two matrices are output, a pattern matrix and structure matrix. The pattern matrix describes the unique relationship between each component and indicator, while the structure matrix is calculated by multiplying the pattern matrix by the factor correlation matrix. Although the pattern matrix is used more frequently, we present both matrices here for comparison purposes.[97] Tables 4.20 and 4.21 below present the oblique rotation pattern and structure matrix, respectively, using the latent root criterion. Factor loadings above 0.4 are highlighted, and the five components explained 60.049% of the variance.

Table 4.20: Oblique Rotation Factor Pattern Matrix

HQWP Question	Component				
	1	2	3	4	5
PP7.25c	0.937	0.089	-0.033	-0.092	-0.141
PS7.25d	0.899	0.012	-0.031	-0.043	-0.070
P7.25b	0.821	0.040	0.034	0.062	0.019
P30.25a	0.808	0.079	-0.033	0.099	0.050
PRQ.23a	0.380	-0.147	0.161	0.093	0.266
PRQ.23h	-0.029	0.738	-0.011	-0.006	0.109
PRQ.23g	0.079	0.716	0.102	0.124	-0.011
PRQ.23i	0.070	0.669	-0.009	-0.029	0.098
PRQ.23f	0.347	0.382	-0.091	0.263	-0.150
AM.28	<0.001	0.023	0.819	-0.079	-0.047
AM.29	-0.036	0.023	0.804	0.026	-0.018
PRQ.23d	-0.190	0.119	0.027	0.807	-0.192
PRQ.23e	0.123	0.129	-0.032	0.711	-0.075
PRQ.23b	0.175	-0.186	-0.095	0.653	0.173
PRQ.23c	0.024	0.080	0.016	0.629	0.060
Prod7.26	0.243	-0.143	0.032	0.520	0.320
AM.27	-0.139	0.100	-0.101	-0.136	0.750
HLW.24	0.034	0.218	0.064	0.169	0.517

Table 4.21: Oblique Rotation Factor Structure Matrix

HQWP Question	Component				
	1	2	3	4	5
PP7.25c	0.884	0.201	0.006	0.346	-0.002
P30.25a	0.872	0.234	0.014	0.500	0.194
PS7.25d	0.869	0.135	0.010	0.366	0.065
P7.25b	0.861	0.185	0.079	0.457	0.161
PRQ.23a	0.451	-0.050	0.195	0.270	0.338
PRQ.23g	0.250	0.763	0.108	0.368	0.057
PRQ.23h	0.099	0.737	-0.012	0.206	0.139
PRQ.23i	0.175	0.676	-0.006	0.208	0.137
AM.28	0.001	-0.005	0.814	-0.044	-0.028
AM.29	0.018	0.020	0.802	0.048	0.010
PRQ.23e	0.458	0.348	0.002	0.793	0.046
PRQ.23d	0.173	0.311	0.045	0.729	-0.105
PRQ.23b	0.470	0.036	-0.051	0.700	0.277
PRQ.23c	0.338	0.267	0.046	0.672	0.153
Prod7.26	0.513	0.058	0.079	0.636	0.423
PRQ.23f	0.499	0.504	-0.069	0.508	-0.044
AM.27	-0.073	0.077	-0.088	-0.074	0.711
HLW.24	0.231	0.296	0.090	0.321	0.558

Note that the questions are in the form XXX.QQ, where XXX is the abbreviation for the presenteeism construct, and QQ is the question number in HQWP. Also note that some presenteeism constructs spanned multiple questions (e.g., PRQ consisted of 9 questions), while several only consisted of one question. As expected, the structure matrix components are larger than those of the pattern matrix. However, unlike the structure matrix, the pattern matrix does have one indicator, PRQ.23c, without a component. Looking at the component correlations, we were not surprised to see that components 1 and 4 had a moderate correlation of 0.462 because of the significant cross-loadings between those components in the structure matrix. We noted that component 1

also had a small correlation with components 2 and 4. The component correlation matrix is shown below in Table 4.22

Table 4.22: Oblique Component Correlation Matrix

Component	1	2	3	4	5
1	1.000	0.154	0.051	0.462	0.158
2	0.154	1.000	-0.004	0.286	0.048
3	0.051	-0.004	1.000	0.043	0.036
4	0.462	0.286	0.043	1.000	0.136
5	0.158	0.048	0.036	0.136	1.000

When comparing Cronbach α values of the pattern and structure matrices, the pattern matrix had somewhat mediocre values of 0.912, 0.618, 0.481, 0.764, and 0.089; while the structure matrix had somewhat better values of 0.868, 0.650, 0.481, 0.843, and 0.288. Given some low Cronbach α values, we considered changing the extraction criteria to a larger number of factors by examining the Scree plot, shown below in Figure 4.20. Since there was no noticeable straightening of the curve after 2 components, we did not pursue performing a further oblique factor analysis on a larger number of components.

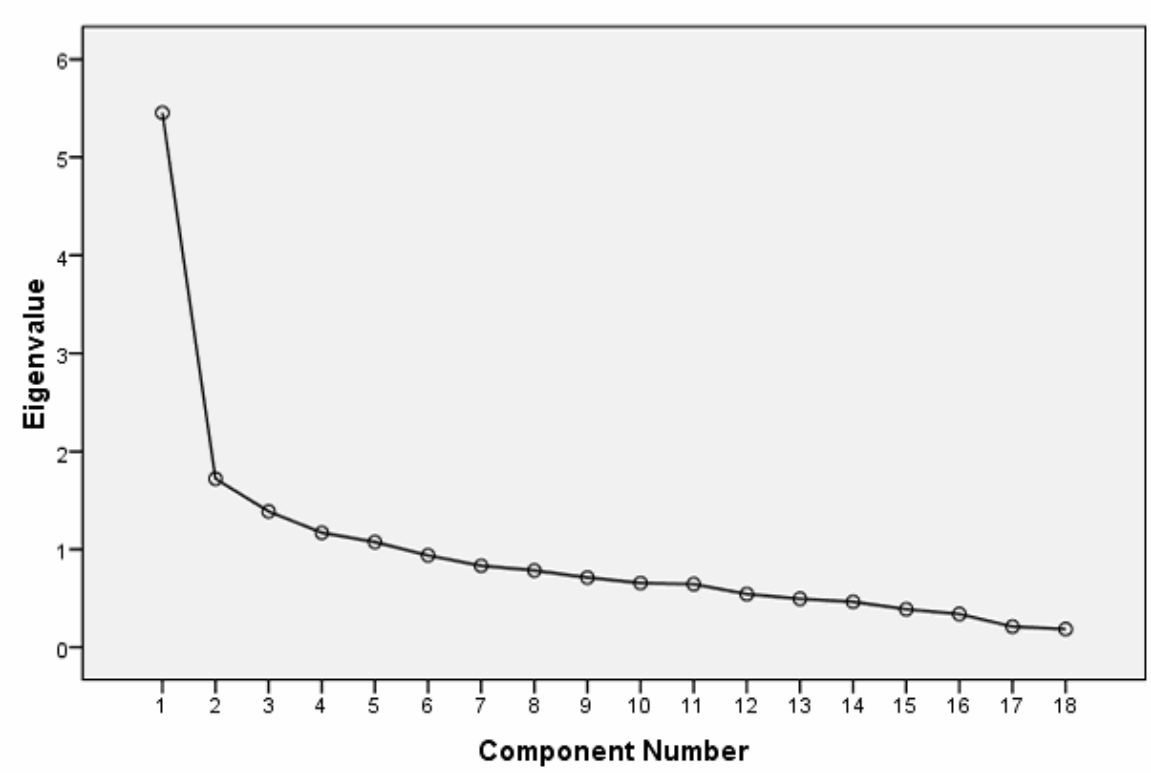

Figure 4.20 Scree Plot

To maximize the reliability of our factor analysis, we decided to use the structure matrix (Table 4.21) for our oblique rotation model. Component 1 is composed of the all four direct presenteeism measures (three 7-day and one 30-day), four questions from PRQ, and 7-day productivity. Component 2 comprises 4 questions from PRQ. Component 3 consists of two questions from achievements and mistakes. Component 4 consists of the 2 self-reported measures of presenteeism (P30 and P7), 5 questions from PRQ, and 7-day productivity. Component 5 consists of 7-day productivity, health limits work, and 1 question from achievements and mistakes. Table 4.23 below summarizes the component information gleaned from this factor analysis.

Table 4.23: Oblique Rotation Summary

Component	Cronbach α	Description
1	0.868	P30, P7, PP7, PS7, 4 PRQ questions, Prod7
2	0.650	4 PRQ questions
3	0.481	2 AM questions
4	0.843	P30, P7, 5 PRQ questions, Prod7
5	0.288	Prod7, HLW, 1 AM question

4.8.2 Orthogonal Rotation

For comparison purposes, we also ran a factor analysis using an orthogonal rotation on the same set of questions, namely those used to determine the productivity covariates of HQWP. Table 4.24 below presents the results of factor analysis using the latent root criterion, with factor loadings above 0.4 highlighted. Again, the five components explained 60.049% of the variance. Note that for orthogonal factor analyses, the structure and pattern matrices are identical, so there was only one factor matrix.

Table 4.24: Orthogonal Rotation Factor Matrix

HQWP Question	Component				
	1	2	3	4	5
PP7.25c	0.876	0.112	0.155	-0.028	-0.099
PS7.25d	0.856	0.146	0.084	-0.025	-0.029
P7.25b	0.819	0.240	0.123	0.041	0.063
P30.25a	0.817	0.277	0.168	-0.025	0.094
PRQ.23a	0.419	0.173	-0.092	0.168	0.288
PRQ.23d	0.005	0.732	0.230	0.032	-0.157
PRQ.23e	0.294	0.713	0.255	-0.025	-0.031
PRQ.23b	0.339	0.646	-0.059	-0.084	0.207
PRQ.23c	0.191	0.618	0.186	0.023	0.093
Prod7.26	0.393	0.550	-0.030	0.043	0.353
PRQ.23g	0.140	0.221	0.737	0.098	0.016
PRQ.23h	0.011	0.081	0.731	-0.016	0.122
PRQ.23i	0.098	0.072	0.667	-0.013	0.113
PRQ.23f	0.400	0.358	0.450	-0.089	-0.114
AM.28	0.008	-0.057	0.009	0.817	-0.031
AM.29	0.001	0.037	0.022	0.802	0.002
AM.27	-0.095	-0.103	0.078	-0.100	0.731
HLW.24	0.137	0.228	0.256	0.067	0.529

The five factors have standardized Cronbach α values of 0.850, 0.764, 0.650, and 0.481, and 0.089, respectively. Although these values are not ideal, they were similar to those we found for the oblique rotation and we felt that they were sufficient for exploratory factor analysis and did not feel like including more factors would be appropriate, so we did not pursue further improvements.

Thus, for the orthogonal rotation model, we have the five-component model described in Table 4.24. Component 1 is composed of the four direct presenteeism measures (P30, P7, PP7, and PS7), and 2 questions from PRQ. Component 2 comprises 4 questions from PRQ and 7-day productivity. Component 3 consists of 4 questions from PRQ. Component 4 consists of 2 questions from achievements and mistakes. Component 5 consists of health limits work and 1 question from achievements and mistakes. Table 4.25 below summarizes the component information gleaned from this factor analysis.

Table 4.25: Orthogonal Rotation Summary

Component	Cronbach α	Description
1	0.850	P30, P7, PP7, PS7, 2 PRQ questions
2	0.764	4 PRQ questions, Prod7
3	0.650	4 PRQ questions
4	0.481	2 AM questions
5	0.089	HLW, 1 AM question

The orthogonal rotation had smaller components and minimal cross-loading compared to the oblique factor analysis, though the two analyses did include two identical components. Overall, the reliability of the factors was higher for the oblique rotation, although both analyses included a component with HLW and the first achievements and mistakes question ("Did you experience any special work success or achievement at any time during the past 30 days?") with extremely low reliability, which implies that those questions are not related to the other presenteeism measures. Our results do give support to the notion that the remaining measures of presenteeism are related. Thus, our main takeaway from our factor analyses is that there is a significant

underlying relationship between direct presenteeism measures (P30, P7, PP7, PS7), the questions that make up the PRQ index, and 7-day productivity (Prod7).

4.9 Economic Analysis

To be sure, great caution needs to be exercised when performing any sort of economic translation of workforce productivity figures, especially given the difficulty in the estimation of these figures to begin with. However, several studies have been done to translate productivity into monetary values, [98-102] further underscoring the need for productivity instruments to have this capability.

A simple economic analysis was run to illustrate the magnitude of indirect cost savings one could achieve with programs that could be used to reduce the absenteeism and presenteeism in a population using health promotion and disease management (HP/DM) programs. Given our results that showed significantly higher levels of presenteeism among faculty and absenteeism among staff, we presumed separate programs for each group focusing on the areas where the most improvement could be realized. Our analysis assumes the following:

- Total Georgia Tech faculty and staff populations of 835 and 4500, respectively
- Rates of absenteeism and presenteeism for the sample population are applicable to the entire GT population
- A yearly discount rate of 6%
- A time horizon of 5 years
- An average yearly salary increase of 4%
- An average faculty salary of $103,900
- An average staff salary of $45,000

- A faculty annual turnover rate of 7% and staff annual turnover rate of 5%
- Overall population growths of 0.05% for both faculty and staff
- Mean staff absenteeism rate of 0.89615 days per month
- Mean faculty presenteeism rate of 0.17035 (proportion of overall work loss)
- Changes in direct (medical, mental health, and pharmaceutical) costs are not considered

Given these parameters, we performed an analysis to see what cost savings could be realized with an HP/DM program that reduces absenteeism 1% per year for staff and reduces presenteeism 1% per year for faculty. Since we assumed that programs have the most impact on current (not new) employees, we also assumed that these reductions only affect employees who have been working at least one year. Given these assumptions, we found that the net present value (NPV) of a staff program that reduces absenteeism by 1% is about $657,000, or approximately $142 per capita, and the NPV of a faculty program that reduces presenteeism by 1% is about $722,700, or about $844 per capita. Assuming that program costs to achieve the desired reductions do not exceed those amounts, such programs would be cost-effective without even taking into account other potential savings such as reductions in direct medical and pharmaceutical costs. Tables summarizing the effect of programs to reduce faculty absenteeism and staff presenteeism are summarized below in Tables 4.26 and 4.27.

Table 4.26: Economic Evaluation of Staff Absenteeism Reduction Program

Variable	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Total employees	4,500	4,523	4,545	4,568	4,591	4,614
New employees	0	338	339	341	343	344
Old employees (1+ years)	0	4,185	4,206	4,227	4,248	4,269
Yearly discount value	6%	6%	6%	6%	6%	6%
Yearly reduction in absenteeism	0.0%	1.0%	1.0%	1.0%	1.0%	1.0%
Absenteeism rate per employee per month	0.896	0.887	0.878	0.870	0.861	0.852
Absenteeism rate per employee per year	10.754	10.646	10.540	10.434	10.330	10.227
Average annual salary	$50,000	$52,000	$54,080	$56,243	$58,493	$60,833
Average annual salary increase	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%
Monetary value of productivity gain	$0	$130,013	$134,531	$139,206	$144,043	$149,048
Discounted value	$0	$122,654	$126,916	$131,326	$135,890	$140,611
Cumulative NPV	$0	$122,654	$249,570	$380,897	$516,786	$657,398
Cumulative NPV per capita	$0	$27	$55	$83	$113	$142

Table 4.27: Economic Evaluation of Faculty Presenteeism Reduction Program

Variable	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Total employees	835	839	843	848	852	856
New employees	0	63	63	63	64	64
Old employees (1+ years)	0	777	780	784	788	792
Yearly discount value	6%	6%	6%	6%	6%	6%
Yearly reduction in presenteeism	0.0%	1.0%	1.0%	1.0%	1.0%	1.0%
Presenteeism rate	0.170	0.169	0.167	0.165	0.164	0.162
Average annual salary	$103,900	$108,056	$112,378	$116,873	$121,548	$126,410
Average annual salary increase	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%
Monetary value of productivity gain	$0	$142,942	$147,909	$153,049	$158,367	$163,870
Discounted value	$0	$134,851	$139,537	$144,386	$149,403	$154,594
Cumulative NPV	$0	$134,851	$274,388	$418,774	$568,176	$722,770
Cumulative NPV per capita	$0	$161	$325	$494	$667	$844

CHAPTER 5

DISCUSSION

5.1 Productivity

As expected, HRQOL and productivity loss, as measured by absenteeism and presenteeism, were negatively correlated. In other words, the healthier one is, the less work loss one suffers. This relationship held for the overall sample, as well as the faculty and staff groups.

Several statistical tests performed support the notion that mental health is more closely tied to productivity than physical health is, as postulated in Hypothesis 2. First, the scatterplots presented in Figures 4.8, 4.9, and 4.10 gave a visual indication that there was a stronger relationship between productivity (measured as absenteeism, presenteeism, and PRQ) and MCS than productivity and PCS. Next, looking at the raw correlations between HRQOL and productivity measures in Table 4.9, we saw that, for the most part, only mental health scales had moderate correlations (> 0.3) with productivity scales, including absenteeism, presenteeism-related questions, and 7-day productivity. The one exception was health limits work (HLW), which had moderate correlations with three of the four physical health scales, as well as all four mental health scales.

All of our regression models also supported the notion that mental health measures are a more significant predictor of productivity than physical health. The MCS scales of role emotional and vitality were found to be a significant predictor for absenteeism, while role emotional was found to be a significant predictor of presenteeism. For PRQ original and revised models, HRQOL predictors varied but each model included

at least two of the four MCS scales of mental health, role emotional, vitality, and social function. Of all of the regression models run, less than half had any predictors from PCS: the overall and staff models for absenteeism; faculty model for presenteeism; and original staff model for PRQ had any physical health scale as a predictor. The physical function scale was a significant predictor in first three models, and general health (which we know to be correlated with mental health measures) was a predictor for the last model.

Another interesting result we found was that absenteeism and presenteeism are not even moderately correlated (Table 4.8), which was not expected since both are measures of productivity. Other results specific to absenteeism and presenteeism measures alone follow in the next two sections.

5.2 Absenteeism

Clearly, absenteeism is easier than presenteeism to measure in most cases. As mentioned earlier, absenteeism usually can be determined by asking those in supervisory roles to report absent days or by asking employees to self-report days absent. The fact that faculty had significantly lower rates of absenteeism than staff could imply that those with flexible work schedules are less likely to report absenteeism since absenteeism is much more difficult to define without a more structured work schedule.

Our analysis found that 30-day absenteeism had a small correlation with physical health, as measured by PCS, and medium correlation with mental health, as measured by MCS, and that these correlations were statistically different. These correlations were also statistically equivalent when looking at the faculty and staff groups separately (Table 4.10). Other SF-36 scales the reported moderate correlations with absenteeism include

vitality and social function (Table 4.9). When comparing the correlation strength of SF-36 scales between the absenteeism of faculty and staff, physical function had statistically weaker correlations than five of the seven other scales (Table 4.12). Lastly, staff reported statistically higher levels of absenteeism than faculty (Table 4.14).

Our binary logistic model for binary absenteeism yielded three models that included the MCS scales of vitality and role emotional as a significant predictor for the overall sample, faculty and staff. As might be expected given our result in Table 4.14 that staff had significantly higher levels of absenteeism than staff, the faculty binary variable was also a significant predictor for the overall model, with faculty having 0.615 times the odds of reporting absenteeism. Other interesting predictors in these absenteeism models include the PCS scale of physical function, job functions, and being of black ethnicity.

Lastly, our economic analysis found that net present value (NPV) of a HP/DM program to decrease staff absenteeism by 1% per year is about $657,000, or $142 per capita, by year 5. Possible programs that could help reduce absenteeism include programs like on-site child care and on-site health clinics, and incentives to adopt or maintain a healthy lifestyle through healthy eating and regular exercise. While such programs could have a cost greater than the $657,000 cost savings estimated, effective programs could result in reductions in absenteeism of more than 1%. Furthermore, economic analyses of programs would also need to include direct medical and pharmaceutical cost reductions, which would further add to the monetary benefits realized.

5.3 Presenteeism

Although presenteeism is normally thought of as a very difficult construct to measure, the HQWP used several covariates related to presenteeism. All six of these measures had at least moderate correlations (> 0.3) with each other, and four of these, P30, P7, PP7, and PS7, had strong correlations (> 0.7) with each other (Table 4.14). These results provide a great amount of reliability in trying to use HQWP to measure the concept of presenteeism.

A significant result we found is that faculty reported statistically higher levels of presenteeism than staff (Table 4.14). This result could imply that those with flexible work schedules (like most faculty) are less likely to be productive while on the job compared to those with a more structured work schedule (like most staff).

Some of our hypothesis testing gave evidence that mental health is more strongly correlated with presenteeism in particular than physical health is. PRQ was found to have a statistically stronger correlation with MCS than PCS, for example (Hypothesis 2). Another interesting result we found is that, when looking at the correlation between MCS and 30-day presenteeism, the staff group does have a significantly stronger correlation than the faculty (Hypothesis 3). Although Hypotheses 4 and 5, which considered whether the MCS scales of role emotional and vitality had stronger correlations to productivity than the other SF-36 scales, had almost no support, these two scales were found to be significant predictors of PRQ, and role emotional was the only scale that was found to be a significant predictor of presenteeism.

When looking at our regression models, RE was the only common predictor (and indeed the only predictor for the overall sample and staff group) for 30-day presenteeism

models. This result was significant because it was the only HRQOL scale that was significant across all models for the three regression models we ran (A30, P30, PRQ). For PRQ, MCS was a significant predictor for all three models (overall, faculty and staff), as it was for A30. These regression results provide further support that mental health measures are better predictors of productivity than physical health measures are.

Our factor analyses, which looked at relationships between presenteeism questions in HQWP, implied that is that there is a significant underlying relationship between direct presenteeism measures (P30, P7, PP7, PS7), the questions that make up the PRQ index, and 7-day productivity (Prod7). As a result, future productivity surveys may be able to make use of constructs like PRQ and 7-day productivity to substitute for or to confirm direct presenteeism constructs like 30-day or 7-day presenteeism. Because presenteeism is such a hard concept to measure as a subjective and self-reported measure, additional constructs that can help guide and confirm its measurement could undoubtedly prove useful in further productivity studies that include presenteeism as a measure. Finally, the factor analysis results give some evidence that presenteeism, as measured by the covariates in HQWP (not including achievements and mistakes, and health limits work), is a reliable construct.

Lastly, our economic analysis found that net present value (NPV) of a HP/DM program to decrease faculty presenteeism by 1% per year exceeds a $722,000, or $844 per capita, by year 5. Possible programs that could help reduce presenteeism include social programs and events to help increase job satisfaction, as well as structural considerations such as natural lighting in buildings. While such programs likely would not affect physical health much, they would likely have a positive impact on mental

health, which could, in turn, reduce direct medical and pharmaceutical costs from reduced mental health expenditures, which would further add to the monetary benefits realized.

5.4 Study Significance

The results of this study and subsequent analysis have several contributions to the knowledge base. First of all, this is one of the very few studies that simultaneously measures productivity and HRQOL. In fact, only three productivity instruments, MWPLQ,[54] SPS-13,[31] and WPAI,[70] have been validated or compared with results from the SF-36 or other HRQOL measures, and only one empirical study was found specifically examining the relationship between productivity and HRQOL, and that study was a secondary data analysis limited to patients with lower back pain.[81]

This thesis is also significant because it includes in its productivity measure presenteeism, an invisible quantity that is difficult to measure with or without the occurrence of absenteeism. Measuring presenteeism accurately is especially important because estimated presenteeism costs usually exceed those of not only absenteeism costs, but also direct (medical and pharmaceutical) costs for employers.[31] In addition, there is not yet an established standard for workforce productivity instruments, let alone presenteeism instruments, making any reasonable theoretical or empirical productivity study a significant contribution to the knowledge base. Furthermore, the instrument used is one of the few productivity surveys that includes questions designed to reduce social desirability bias.

Like many other public health issues, improving workforce health has the potential to produce very real returns but will require upfront investments that all but some forward-thinking companies are willing to make. Berger et al. (2003), who

estimate that the average workforce is effectively reduced 5% to 10% due to health-related problems, even go as far as to say that American companies and organizations that make a commitment to increasing investments in a healthy workforce will be leaders in future gains of US productivity.[103]

Finally, this study could help guide cost-effectiveness analysis for economists and decision-makers in health policy and public health using analyses such as the one presented here. The economic analyses presented here give a simplified but important feel for the magnitude of potential cost savings. Ultimately, more comprehensive and better quality tools to measure presenteeism could help strategists implement disease management and health promotion programs that would reduce health risks before health conditions arise, thereby not only improving overall productivity and reducing absences but also saving companies and employees direct costs, reducing rates of turnover due to increased employee satisfaction, and improving both quantity and quality of life years.[91, 100, 104, 105]

5.5 Study Limitations

Perhaps the most significant limitation of the HQWP instrument is that measurement of productivity varies by company and profession type. For example, knowledge- and production-based jobs can have very different measures of output. Our study assumes that we can use the same instrument to measure the productivity of a sample population, even though they may have different job functions and requirements. Our study has limited generalizability because, although a wide variety of job functions within a university was surveyed, our sample did not include production-based

employees due to the low response rate to our email invitation only, web-based survey. In addition, the and voluntary nature of the study may have induced bias, as certain types of people may have been more inclined to participate in the survey than others.

Another limitation relates to the sample group selected for this study. While a university population may not be representative of all populations, it is an interesting population to use for workforce productivity measure testing because of it consists of a mix of knowledge-based jobs, which typically do not lend themselves naturally to productivity measures, and production-based jobs, which do. Also, since the purpose of this study is not to test an instrument that is slated for widespread use but rather to examine and explore the relationship between variables, the study design is suitable. Other limitations include a potential lag in measuring the effect of HRQOL on productivity. In other words, a lag in productivity may not be evident until a period of time after the onset of an illness or condition. Lastly, our sample included both full- and part-time workers. Although there were very few part-time workers, their responses could have biased the results due to different HRQOL or productivity measures.

Other limitations relate to the actual timing of the survey administration. Since many faculty and some staff working schedules vary during the summer (compared to the rest of the academic year), survey results may not be representative of a normal academic population. For example, one participant remarked that his work schedule did not correspond to his non-summer schedule, as he was not being paid by the university during the summer but was using the summer months to write grant proposals.

Several limitations of this study relate to the nature of social science research. As for the survey itself, another challenge with the faculty in our sample is that they typically

have flexible work schedules, which makes it difficult to measure and compare productivity. In fact, several faculty members mentioned the difficulty of answering the HQWP productivity questions, specifically presenteeism, as their schedules are not based on number of hours worked but rather classes and research projects. Also, since our sample size was not large enough, we combined full- and part-time positions and were not able to study differences between the two groups.

Another limitation to our study is the covariates not included in our survey that also could be significant predictors of productivity. Perhaps the most controversial but potentially significant is that of income level, which has shown mixed correlations with productivity.[106, 107] Other factors that may be appropriate covariates include health factors, such as BMI, alcohol consumption, and smoking status; family and social factors, such as number and age of children, marital status, participation in social organizations or places of worship; and job-related factors such as job satisfaction, benefits, and commute.

Finally, there are some concerns regarding the use of self-reported data to perform economic evaluations. Unfortunately, no gold standard for measuring workforce productivity exists, which is one reason so many instruments exist. Although no instrument to measure productivity is without flaw, self-reported rather than other measures have become the accepted standard because they are more generalizable to a variety of employee populations and industries. In addition, the questions added to reduce social desirability bias are designed to help alleviate the concerns of using self-reported data.[33, 35]

CHAPTER 6

CONCLUSIONS

This study also reveals several interesting results regarding the relationship between HRQOL and productivity. First of all, we found that HRQOL and productivity loss, as measured by absenteeism and presenteeism, were negatively correlated. This result is straightforward and logical, as one would expect healthier people to be more productive at work. Next, we found that absenteeism and presenteeism had statistically small correlations. This result was a bit surprising, as absenteeism and presenteeism are both standard measures of productivity, so one would expect them to be somewhat related.

Next, we found that mental health measures were a stronger predictor of productivity, both absenteeism and presenteeism, than physical health measures. We also found that staff had significantly higher levels of absenteeism, and faculty higher levels of presenteeism. Moreover, staff had significantly higher levels of 7-day productivity, while faculty worked significantly more extra hours to catch up on work.

Among physical health scales alone, we found that HRQOL scales and productivity had very minor statistical differences in correlation strengths. When comparing the correlations of HRQOL scales and productivity, however, there were some statistical differences in correlation strengths. Physical function was statistically weaker than mental scales for absenteeism; bodily pain and role physical were statistically weaker than mental scales for presenteeism; and vitality, role emotional, and mental health were statistically stronger than physical scales for presenteeism-related questions.

Lastly, we found the interesting result that those who work extra hours to make up for lost production have significantly lower social function scores compared to those who do not. While the number of extra hours did not necessarily correlate with lower social function, this result has potential implications for improvements in employee social function, and, in turn, job satisfaction and overall well-being, by discouraging working extra hours.

6.1 Study Implications

The results of this research have several implications. First of all, since we found that mental health components are stronger predictors for productivity, programs to measure, monitor, or improve mental health could yield significant improvements in productivity. While much attention has been paid recently to programs to physical health through diet and exercise, there is comparably much less attention on programs to improve mental health. The importance of mentally healthy employees for workforce productivity is obviously not an entirely new idea, but the results of this study seem to suggest that there could be substantial benefits for employers to implement more programs related to mental health maintenance and improvement.

Secondly, it is clear from our study that absenteeism and presenteeism, though both are measures of productivity loss, are not necessarily directly linked to each other. Specifically, we found that faculty had higher levels of presenteeism, while staff had lower levels of absenteeism. One possibility for the differences could be schedules, as faculty tend to have more flexible work schedules, which implies that absenteeism may be difficult to define but that having a flexible work schedule results in less productive

working time. On the other hand, staff tend to have more structured schedules, meaning their time at work is less flexible, which could lead to increased productivity while at work as well as higher rates of reported absence from work. Thus, programs designed to improve productivity need to consider whether absenteeism, presenteeism, or both need to be addressed, depending on the characteristics of the target population.

Lastly, the result that those who work extra hours to make up for lost production have significantly lower social function scores compared to those who do not could have significant policy implications. Specifically, programs to help balance workload among employees to minimize overtime could help increase social function scores of employees, which could, in turn, improve other aspects of health and general employee morale and satisfaction with work.

6.2 Future Research

Because of the difficulty of performing productivity studies for a general population, many recent studies have started examining the effect of specific health risk factors on productivity. For example, Tsai et al. (2008) examined the role of overweight and obesity on absenteeism for industrial petrochemical workers and estimated the cost of just this one risk factor to be 36% of the total illness absence, a total loss of $1,873,500 per year for an employee population of 4153 employees.[108] In another recent publication, Stewart et al. (2007) estimated the economic impact of lost productive time due to diabetes and neuropathic pain in the US workforce to be about $3.65 billion annually.[109]

However, there are still many unexplored avenues of productivity research for general populations. In fact, perhaps the most important contribution of this work is a

corroboration of the difficulty in creating productivity surveys that are generalizable to multiple job types. For example, full-time versus part-time positions have different demands and requirements, which could clearly affect productivity and its measurement in different ways. As mentioned in the Study Limitations section above, some faculty members encountered difficulty answering questions related to presenteeism due to the flexible work schedules during the summer and among faculty in general. It would be interesting to see, for example, whether the differences found between faculty and staff are generalizable to other flexible time and traditional jobs, respectively. A related issue would be looking at basis of pay differences (i.e., salary versus hourly) as a group comparison study. Surveying academic populations may yield different results during the non-summer months, so another research possibility would be to see whether results differ based on the time of year of survey administration.

Another interesting finding from this study is the fact that an online productivity survey may only be suitable for certain types of positions. Since our online survey was found to have a reasonable response rate for knowledge-based workers but not production-based workers, we were not able to do a group comparison study between the two groups as originally planned. With more time and resources, it would have been interesting to implement the same survey among production-based workers in paper-based form to perform the group comparisons that were one of the original goals of this study.

Although our study found a negative correlation between HRQOL and productivity loss, our sample was too small to explore the shape of the relationship, especially if it is non-linear. Of particular interest would be how productivity changes at

particularly high and low levels of HRQOL compared to average levels of HRQOL. Figure 6.1 below shows a possible depiction of a non-linear relationship between HRQOL and productivity, where at low and high levels of HRQOL, productivity improves only slightly with incremental increases in HRQOL due to floor and ceiling effects, and the greatest gains in productivity are realized for incremental increases in average HRQOL. A larger study would also allow for the exploration of age effects on the relationship between HRQOL and productivity.

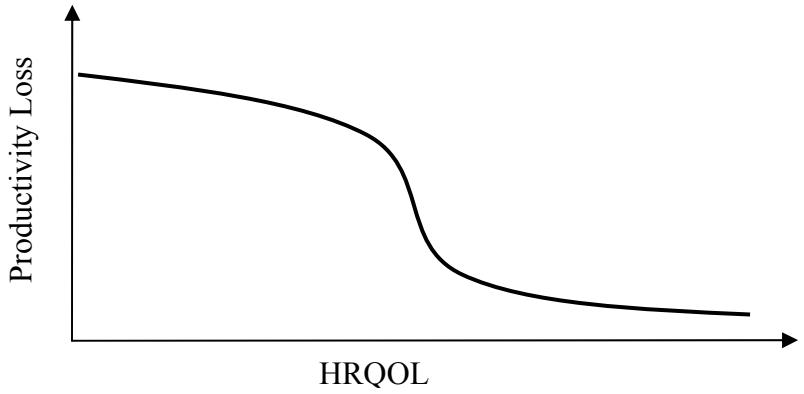

Figure 6.1: Possible Relationship between Productivity Loss and HRQOL

Many recent studies examining the relationship between HRQOL and workforce productivity have focused on those with and without specific health conditions. However, it is clear that more study in exploring the relationship between HRQOL and workforce productivity is needed. Policy implications that result from such study have the potential to affect not only employers and employees, but their families, and indeed, society as a whole.

APPENDIX A: SHORT-FORM 36 (SF-36) QUESTIONNAIRE

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities. Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is: (circle one)

   Excellent 1
   Very good.....2
   Good 3
   Fair.....4
   Poor 5

2. Compared to one year ago, how would you rate your health in general now? (circle one)

   Much better now than one year ago.....1
   Somewhat better than one year ago 2
   About the same as one year ago 3
   Somewhat worse now than one year ago 4
   Much worse now than one year ago 5

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? *(circle one number on each line)*

ACTIVITIES	**Yes, Limited a Lot**	**Yes, Limited a Little**	**No, Not Limited At All**
a. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
b. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
c. Lifting or carrying groceries	1	2	3
d. Climbing **several** flights of stairs	1	2	3
e. Climbing **one** flight of stairs	1	2	3
f. Bending, kneeling, or stooping	1	2	3
g. Walking **more than a mile**	1	2	3
h. Walking **several blocks**	1	2	3
i. Walking **one block**	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health? *(circle one number on each line)*

	YES	**NO**
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Were limited in the **kind** of work or other activities	1	2
d. Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? *(circle one number on each line)*

	YES	**NO**
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Didn't do work or other activities as **carefully** as usual	1	2

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups? (circle one)

Not at all.....1
Slightly.....2
Moderately.....3
Quite a bit.....4
Extremely.....5

7. How much bodily pain have you had during the past 4 weeks? (circle one)

None.....1
Very mild.....2
Mild.....3
Moderate.....4
Severe.....5
Very severe.....6

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? (circle one)

Not at all.....1
A little bit.....2
Moderately.....3
Quite a bit.....4
Extremely.....5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks: *(circle one number on each line)*

	All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
a. Did you feel full of pep?	1	2	3	4	5	6
b. Have you been a very nervous person?	1	2	3	4	5	6
c. Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d. Have you felt calm and peaceful?	1	2	3	4	5	6
e. Did you have a lot of energy?	1	2	3	4	5	6
f. Have you felt downhearted and blue?	1	2	3	4	5	6
g. Did you feel worn out?	1	2	3	4	5	6
h. Have you been a happy person?	1	2	3	4	5	6
i. Did you feel tired?	1	2	3	4	5	6

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)? (circle one)

All of the time.....1
Most of the time.....2
Some of the time.....3
A little of the time.....4
None of the time.....5

11. How TRUE or FALSE is each of the following statements for you? *(circle one number on each line)*

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a. I seem to get sick a little easier than other people	1	2	3	4	5
b. I am as healthy as anybody I know	1	2	3	4	5
c. I expect my health to get worse	1	2	3	4	5
d. My health is excellent	1	2	3	4	5

APPENDIX B

WORKFORCE PRODUCTIVITY QUESTIONNAIRES

Angina-Related Limitations at Work Questionnaire

During the past 4 weeks, how much difficulty did you have with the following because of your angina?

- If you had no difficulty because of angina, circle 0 for NO DIFFICULTY.
- If you were completely unable to perform the task, circle 5, SO MUCH DIFFICULTY COULDN'T DO AT ALL.
- If a statement does not describe your work situation, circle 6, DOES NOT APPLY TO MY WORK.
- Please report on your main job only.

Scale: 0 = NO DIFFICULTY, 1 = A SLIGHT AMOUNT, 2 = SOME, 3 = QUITE A BIT, 4 = A GREAT DEAL, 5 = SO MUCH DIFFICULTY COULDN'T DO AT ALL, 6 = DOES NOT APPLY TO MY WORK

a. Getting home to work (for example, getting to work from parking areas, bus or train stations).
b. Getting started at the beginning of the workday (for example, giving yourself enough time to get ready for work, travel to work and rest before starting)
c. Pacing yourself so you could get through the workday (for example, slowing things down)
d. Following a work routine or schedule without having to rearrange or reassign your tasks
e. Working continuously without needing frequent interruptions, breaks or rests
f. Doing any lifting, carry or moving of objects at work.
g. Exerting yourself physically at work (for example, walking and climbing stairs)
h. Concentrating on your work (for example, not being too distracted by your angina symptoms)
i. Controlling irritability or anger toward the people you work with (including, for example, employers, managers, coworkers, customers or the public)
j. Doing all your work without avoiding certain tasks or rushing through them
k. Handling difficult or stressful work situations
l. Handling the workload
m. Working fast
n. Finishing all your work (for example, without taking unfinished work home)
o. Accomplishing as much work as you would like
p. Feeling secure in your job
q. Controlling worry or anxiety about what others at work might think of you

Figure B.1 Angina-Related Limitations at Work Questionnaire (ALWQ)

1. During the past 4 weeks, did you miss any time from your job or business because of your angina?
(Circle one)
Yes .... 1 Go to question 2
No .... 2 Skip to end of questionnaire

2. During the past 4 weeks, were there days when you were completely unable to do any paid work at all or do any of your job tasks at home because of your angina?
(Circle one)
Yes 1
No .... 2 Skip to question 5

3. How many full work days did you miss during the past 4 weeks because of your angina?

_____
Fill in the total number of full days missed

4. Did you receive all, part or none of your regular pay for the days that you missed? (Include payment for work you do at home.)
(Circle one)
All 1
Part 2
None 3

5. During the past 4 weeks, were there days that you did some work at a job or business but put in fewer hours than usual because of your angina?
(Circle one)
Yes 1
No .... 2 Skip to end of questionnaire

6. Think about the days you worked but put in fewer hours than usual during the past 4 week period. For example, you came in late, left early or took time out in the middle of the workday due to angina symptoms, doctor visits or angina treatment. What was the total number of hours you missed over the 4 week period?

_____
Fill in the total number of hours missed

7. Did you receive all, part or none of your regular pay for the hours you missed? (Include payment for work you do at home.)
(Circle one)
All 1
Part 2
None 3

Figure B.1 continued

Health and Performance Questionnaire

1. How many full days of work did you miss in the past 30 days not including vacation or maternity leave?

1. None
2. One
3. More than one (specify number) _ _ _ _ _ _ _ _ _
4. Don't know
5. Refused

1a. How many of these days were in the past week?

1. None *(go to question 4)*
2. One *(go to question 2)*
3. More than one (specify number) _ _ _ _ _ _ _ _ _ *(go to question 3)*
4. Don't know
5. Refused

2. Was that because of problems with your own health, the health of someone else, or for some other reason?

1. Own health
2. Other's health
3. Other reason

Go to question 4

3. How many of these days did you miss because of problems with your own health?

4. How many days in the past 30 days did you either come in late for work or leave early?

1. None *(go to question 9)*
2. One *(go to question 5)*
3. More than one (specify number) _ _ _ _ _ _ _ _ _ *(go to question 7)*
4. Don't know
5. Refused

5. How many hours did you miss on that day?

6. Did you miss this time because of problems with your own health, the health of someone else, or for some other reason?

1. Own health
2. Other's health
3. Other reason

Go to question 9

Figure B.2 MacArthur Health and Work Performance Questionnaire (MHPQ)

7. On average, about how many hours of work did you miss on each of those days?

8. How many of those days was your reduced time at work because of problems with your own health?

9. How many days in the past 30 did you either come in early, work late, or work on your day off in order to catch up on your work?

1. None *(go to question 12)*
2. One *(go to question 10)*
3. More than one (specify number) _ _ _ _ _ _ _ _ _ *(go to question 11)*
4. Don't know
5. Refused

10. How many extra hours of work did you put in that day?
Go to question 12

11. Altogether, about how many extra hours of work did you put in on those days combined?

12. The next questions are about the time you spent at work over the past 30 days. How often during that time did you have each of the following experiences:

12a.How much of the time was your speed of work or productivity higher than expected?

1. All of the time *(go to question 12c)*
2. Most of the time
3. About half of the time
4. Some of the time
5. A little of the time
6. None of the time
7. Don't know
8. Refused

12b.How much of the time was your speed of work or productivity lower than expected?

Scored as above

12c.How much of the time did you do no work at times when you were supposed to be working?

Scored as above

12d.How much of the time did you find yourself not working as carefully as you should?

Scored as above

12e.How much of the time was the quality of your work lower than expected?

Scored as above

12f. How much of the time did you find yourself daydreaming and not concentrating on your work?

Scored as above

12g.How much of the time did you have trouble getting along with others at work?

Scored as above

Figure B.2 continued

12h.How much of the time did you have trouble controlling your emotions when you were around people at work?

Scored as above

12i. How much of the time did you get along well with others at work?

Scored as above

13. During the time you were at work in the past 30 days, how often did health problems limit you in the kind or amount of work you could do compared to usual?

1. All of the time
2. Most of the time
3. About half of the time
4. Some of the time
5. A little of the time
6. None of the time
7. Don't know
8. Refused

14. On a scale from 0 to 10, where 0 is the worst job performance anyone could have at your job and 10 is the performance of a top worker, what number describes your overall job performance on the days you worked during the past 30 days?

14a.Using the same 0 to 10 scale, how would you rate your job performance during the past 7 days?

15. Using the same 0 to 10 scale, how would you rate your usual job performance?

16. How would you rate the usual job performance of most workers on your job?

17. How many days in the past 7 was your speed of work or productivity lower than expected?

1. All of the time
2. Most of the time
3. About half of the time
4. Some of the time
5. A little of the time
6. None of the time
7. Don't know
8. Refused

18. Did you experience any special work success or achievement at any time during the past 30 days?

1. Yes
2. No
3. Don't know
4. Refused

Figure B.2 continued

19. Did you have any special work failure, make any big mistakes, or miss a major deadline at any time during the past 30 days?

1. Yes
2. No
3. Don't know
4. Refused

20. Did you make any big mistake at work during the past 30 days that either caused an accident or that created a safety risk for yourself or for others?

1. Yes
2. No
3. Don't know
4. Refused

Figure B.2 continued

Health and Work Questionnaire

Life and Work Satisfaction

This next group of questions asks about how satisfied you are with your life, relationships with friends and family, and your work.

1. Overall, how **stressed** have you felt this week?

Not stressed at all									Very stressed
1	2	3	4	5	6	7	8	9	10

2. How **satisfied** were you this week with the *physical environment* in which you work (e.g., amount of noise, temperature where you work)?

Very dissatisfied									Very satisfied
1	2	3	4	5	6	7	8	9	10

3. How **personally rewarding** did you find your **work** this week?

Not rewarding at all									Very rewarding
1	2	3	4	5	6	7	8	9	10

4. How **personally rewarding** did you find your *personal life outside of work* this week?

Not rewarding at all									Very rewarding
1	2	3	4	5	6	7	8	9	10

5. How **satisfied** were you this week with your **relationships with friends and family**?

Very dissatisfied									Very satisfied
1	2	3	4	5	6	7	8	9	10

6. How **satisfied** were you overall with **your job** this week?

Very dissatisfied									Very satisfied
1	2	3	4	5	6	7	8	9	10

7. How **satisfied** were you this week with your relationships with your **coworkers**?

Very dissatisfied									Very satisfied
1	2	3	4	5	6	7	8	9	10

8. How **satisfied** were you this week with your relationships with your **supervisors**?

Very dissatisfied									Very satisfied
1	2	3	4	5	6	7	8	9	10

Figure B.3 Health and Work Questionnaire (HWQ)

9. How much **control** did you feel you had over how **you did your job** this week?

No control at all									Total control
1	2	3	4	5	6	7	8	9	10

10. How easy was it to **communicate** with your **supervisor** this week?

Not easy at all									Very easy
1	2	3	4	5	6	7	8	9	10

11. How easy was it to **communicate** with your **family and friends** this week?

Not easy at all									Very easy
1	2	3	4	5	6	7	8	9	10

The next set of questions asks you about how you felt about your work this week. Rate each question on a scale from 1 to 10, where 10 reflects the highest level you think you could possibly achieve and 1 reflects the lowest level you have ever experienced at work.

12. How would you and the following people describe your **EFFICIENCY** this week?

	My worst ever									My best possible
a. Self	1	2	3	4	5	6	7	8	9	10
b. Supervisor	1	2	3	4	5	6	7	8	9	10
c. Co-workers	1	2	3	4	5	6	7	8	9	10

13. How would you and the following people describe the **OVERALL QUALITY** of your work this week?

	My worst ever									My best possible
a. Self	1	2	3	4	5	6	7	8	9	10
b. Supervisor	1	2	3	4	5	6	7	8	9	10
c. Co-workers	1	2	3	4	5	6	7	8	9	10

14. How would you and the following people describe the **OVERALL AMOUNT** of work you did this week?

	My worst ever									My best possible
a. Self	1	2	3	4	5	6	7	8	9	10
b. Supervisor	1	2	3	4	5	6	7	8	9	10
c. Co-workers	1	2	3	4	5	6	7	8	9	10

Think of your worst level of efficiency ever and your best possible efficiency, then rate how efficient you felt you were this week as compared to your worst ever and best possible.

Figure B.3 continued

Example: Let's say that you feel that you were so efficient this week that it is close to being your best possible performance. In this case your answer would fall somewhere between 8 and 9.

	My worst ever									My best possible
15. Rate your **highest level** of efficiency this week:	1	2	3	4	5	6	7	8	9	10
16. Rate your **lowest level** of efficiency this week:	1	2	3	4	5	6	7	8	9	10

THIS WEEK, *How frequently did you:*

	Never									Almost always
17. Become **annoyed** with or **irritated** by co-workers, boss/supervisor, clients/ customers/vendors or others?	1	2	3	4	5	6	7	8	9	10
18. Become **impatient** with others at work?	1	2	3	4	5	6	7	8	9	10
19. How often did you get into **conflicts** with others at work?	1	2	3	4	5	6	7	8	9	10
20. Become **restless** while at work?	1	2	3	4	5	6	7	8	9	10
21. Lose **interest** or become **bored** with your work?	1	2	3	4	5	6	7	8	9	10
22. Had difficulty **concentrating** at work?	1	2	3	4	5	6	7	8	9	10
23. **Fail to finish** assigned tasks?	1	2	3	4	5	6	7	8	9	10
24. Feel **too exhausted** to do your work?	1	2	3	4	5	6	7	8	9	10

THANK YOU for your time.

Figure B.3 continued

Migraine Work and Productivity Loss Questionnaire

Instructions

These questions ask about the impact of your most recent migraine headache and its treatment on your daily life. Please read every question. Some questions may look like others, but each one is different. Please take the time to read and answer each question carefully by circling the appropriate number or by filling in the answer as requested. If you are unsure about how to answer a question, please give the best answer you can. You may be asked to skip certain questions or even entire sections. Please be sure you are following the instructions carefully.

Thank you for your assistance.

These questions ask about how much impact your most recent migraine headache has had on your ability to perform your regular activities. When choosing an answer, please think about all aspects of the migraine headache including its symptoms and treatment.

1. Think about the entire period of time your most recent migraine headache lasted. How much of the time did you have difficulty performing your work or your regular daily activity such as housework or going to school be- cause of your migraine headache (for example, it took extra effort)?

(Circle one)
All of the time…..... 1
Most of the time..... 2
A good bit of the time..... 3
Some of the time..... 4
A little of the time..... 5
None of the time 6

Work

- Please continue with question 2 if your most recent migraine headache occurred when you usually work at a job or business for pay.
- If you were not working for pay or planning to work for pay during this headache, check the following box and skip to Work Difficulty. ☐

2. How many total hours of work did you miss because of your most recent migraine headache or migraine headache treatment? (If you did not miss any work, fill in `0' and skip to the next page.

_____
Fill in total number of hours missed

Figure B.4 Migraine Work and Productivity Loss Questionnaire (MWPLQ)

3. How many of these hours were missed before you took your initial dose of medication? (If you took medication as soon as your migraine headache started, or if no hours were missed, fill in `0'.)

_____
Fill in hours missed before initial dose

4. How many of these hours were missed after you took your initial dose of medication? (If none, fill in `0'.)

_____
Fill in hours missed after initial dose

5. How many hours altogether did you work while you had symptoms due to the headache or its treatment? (If none, fill in `0'.)

_____
Fill in total hours worked with symptoms during the entire migraine episode

Work difficulty
Please complete this section if you usually worked at a job or business during all or any part of the time that you had your most recent migraine headache. If you are not employed or usually did not work on the days your most recent migraine occurred, check the following box and skip to the end. ☐

6. (a) What is the name of your job? What are the main things you do?
If you have more than one job, please report on your main job only.

_____
_____

(b) How many hours each week do you usually work? If you have more than one job, please report on your main job only.

_____
Fill in total hours per week

(c) If you have a second job, how many hours each week do you usually work at your second job?

_____
Fill in total hours per week

Figure B.4 continued

These questions ask about difficulties you had working because of your most recent migraine headache or migraine headache treatment. After each statement, circle the number that best describes how much difficulty you had during the entire period of time from when your migraine headache started until you stopped working.

- If you have had no difficulty because of your migraine headache, circle 0 for NO DIFFICULTY.
- If you are completely unable to perform the task because of your migraine headache, circle 5 for SO MUCH DIFFICULTY, COULDN'T DO AT ALL.
- If a statement does not describe your work situation, circle 6 for DOES NOT APPLY TO MY WORK.
- Please report on your main job only.

7. Think about the entire period of time your migraine headache lasted. How much difficulty did you have with the following because of your most migraine headache or migraine headache treatment?
(Circle one number on each line)

NO DIFFICULTY (0)
A SLIGHT AMOUNT (1)
SOME (2)
QUITE A BIT (3)
A GREAT DEAL (4)
SO MUCH DIFFICULTY COULDN'T DO AT ALL (5)
DOES NOT APPLY TO MY WORK (6)

(a) Getting started at the beginning of the workday (for example, giving yourself enough time to get ready for work, travel to work and rest before starting)

(b) Pacing yourself so you could get through the workday

(c) Following a routine or schedule without having to rearrange your workday

(d) Working when there is little fresh air, poor ventilation, fumes, odors or smells

(e) Working near bright or flashing lights

(f) Working in noisy areas

(g) Reading or using your eyes when working

(h) Doing things that require you to use your body (for example, walking, bending, reaching and lifting)

Figure B.4 continued

(i) Concentrating on your work (for example, not being distracted by your migraine headache)

(j) Thinking clearly

(k) Controlling irritability or anger toward people you work with (for example, employers, managers, coworkers, customers or the public)

(l) Talking with people in-person, at meetings or on the phone

(m) Doing your work carefully without making mistakes

(n) Doing of all your work without putting tasks aside or rushing through them

(o) Working fast and not slowing down

(p) Finishing all of your work (for example, without taking unfinished work home)

(q) Accomplishing as much as you would like

(r) Not missing too much work

8. Think about the entire period of time your migraine headache lasted. How much difficulty altogether did you have working because of your most recent migraine headache or migraine headache treatment?
(Circle one number)

NO DIFFICULTY (0)
A SLIGHT AMOUNT (1)
SOME (2)
QUITE A BIT (3)
A GREAT DEAL (4)
SO MUCH DIFFICULTY COULDN'T DO AT ALL (5)
DOES NOT APPLY TO MY WORK (6)

9. How would you rate your effectiveness on the job during the entire period of time you had this migraine headache? (100% means you are at your best and 0% means you are at your worst.)
_____ %
Fill in percent

Figure B.4 continued

Osterhaus Technique

In the **last 12 months,** about often have you had migraine attacks?(Please select only one).

_____less than once a month
_____once a month
_____2 times a month
_____3 times a month
_____once a week
_____2-3 times a week
_____4-6 times per week
_____every day

About how many days in the **past 4 weeks** did you work while you had migraine symptoms? _____ days

Compared with your normal work performance, please estimate how effective you are at your job when you have migraine symptoms at work. (100% = fully effective) _____% effective

During a typical migraine, about how long is your work affected by migraine symptoms?_____ hours

During the **past 30 days,** how many days of work or school have you missed because of migraine headaches? (Please state to the nearest half day. If none, please write "0")

_____days of work missed
_____days of school missed
_____not employed

During the **past 12 months,** how many times did you visit an emergency room or emergency clinic to receive treatment for a migraine? (Please select only one).

_____0 (no times)
_____1 time
_____2 times
_____3 times
_____4 times
_____5 times
_____6 times
_____More than 6 times
(please estimate how many times) _____

During the **past 12 months,** how many times did you visit a clinic, HMO or doctor's office for treatment or evaluation of migraines? (Please select only one).

_____0 (no visits)
_____1-2 visits
_____3-4 visits
_____5-6 visits
_____ 7-8 visits
_____9-10 visits
_____11-12 visits
_____More than 12 visits
(please estimate how many times) _____

During the **past 12 months,** how many times were you admitted to a hospital or clinic (overnight) for a migraine? (Please select only one).

_____0 (no times)
_____1 time
_____2 times
_____3 times
_____4 times
_____5 times
_____6 times
_____More than 6 times

What is your date of birth? _____
month / day / year

What is your sex? _____Male _____Female

What is your current employment status? (Please select only one).

_____Employed full time
_____Employed part time
_____Self-employed
_____Homemaker
_____Unable to work because of migraines
_____Unable to work for health reasons other than migraines
_____Other (please specify) _____
_____Looking for work
_____Retired
_____Student

If you are employed, which one of the following best describes your occupation? (Please select only one).

_____Managerial, professional
_____Technical, sales and administrative support
_____Service occupations
_____Precision production, craft and repair
_____Operators, fabricators, labourers
_____Farming, forestry and fishing
_____Student
_____Other (please specify) _____

Figure B.5 Sample Items from the Osterhaus Technique (OT)

QQ instrument

The questionnaire 'Consequences of Illness' covers production losses without absence through the QQ method on a daily basis. On all days, the exact Dutch phrasing was:

Op de schaal hieronder kunt u aangeven hoeveel werk u vandaag hebt gedaan in uw normale werktijd ten opzichte van een normale dag. Een 1 omcirkelen betekent dat u niets kon doen en een 10 betekent dat u evenveel als normaal kon doen.

1 2 3 4 5 6 7 8 9 10

Zo goed als niets Evenveel als normaal

Wilt u inschatten hoe de kwaliteit was van het werk dat u vandaag hebt gedaan ten opzichte van normaal. Een 1 omcirkelen betekent dat uw werk van zeer slechte kwaliteit was en een 10 betekent een even hoge kwaliteit als normaal.

1 2 3 4 5 6 7 8 9 10

Erg slechte kwaliteit Zelfde kwaliteit als normaal

Translated into English:
Could you indicate how much work you actually performed today during regular hours as compared to normal on the scale below?

1 2 3 4 5 6 7 8 9 10

Practically nothing Normal quantity

Could you indicate the quality of the work you performed today as compared to normal on the scale below?

1 2 3 4 5 6 7 8 9 10

Very poor quality Normal quality

Figure B.6 Quantity and Quality Instrument (QQ)

At the end of the week the respondents had to complete the O method and the VR method. These questions covered the past week. The exact phrasing in Dutch in the questionnaire was:

Hoeveel dagen in de afgelopen week bent u wèl naar uw werk geweest, terwijl u last had van gezondheidsproblemen?
LET OP: dagen dat u zich ziek gemeld heeft niet meerekenen
…… dagen

Wij willen u vragen op de meetschaal hieronder te omcirkelen hoe efficiënt u volgens u gewerkt heeft op de dagen dat u wèl op uw werk was terwijl u last had van gezondheidsproblemen. Op de meetschaal betekent 10 dat uw werk niet werd beïnvloed, 1 betekent dat u zeer slecht in staat was uw werk uit te voeren.

1 2 3 4 5 6 7 8 9 10
Maximaal inefficiënt Even efficiënt als normaal

Hoeveel uur zou u in de AFGELOPEN WEEK langer hebben moeten werken om het werk dat u niet hebt kunnen doen als gevolg van gezondheidsproblemen in te halen?
LET OP: dagen dat u zich ziek gemeld heeft niet meerekenen
…… uur

Translated into English:

How many days during this past week did you go to work while suffering from health problems?
NOTE: Do not count the days on which you reported sick……days

Please circle on the scale below the degree of efficiency you consider yourself to have worked with on the days you did go to work while suffering from health problems. On this scale 10 means your work was not affected and 1 means that you were hardly capable of performing your work.

1 2 3 4 5 6 7 8 9 10
Very inefficiently As efficient as normal

How many hours extra would you have had to work to catch up on tasks you were unable to complete in normal working hours due to health problems IN THE PAST WEEK?
NOTE: Do not count the days on which you reported sick.
……hours

Figure B.6 Quantity and Quality Instrument (QQ)

Stanford Presenteeism Scale – 6

Directions: Please describe your work experiences **in the past month**. These experiences may be affected by many environmental, as well as personal factors, and may change from time to time. For each of the following statements, please check one of the following responses to show your agreement or disagreement with this statement in describing *your* work experiences **in the past month.**

Please use the following scale:

. . . I strongly disagree with the statement
. . . I somewhat disagree with the statement
. . . I am uncertain about my agreement with the statement
. . . I somewhat agree with the statement
. . . I strongly agree with the statement

	Strongly disagree	Somewhat disagree	Uncertain	Somewhat agree	Strongly agree
1. Because of my concern/s, the stresses of my job were much harder to handle.	[ ]	[ ]	[ ]	[ ]	[ ]
2. Despite having my concern/s, I was able to finish hard tasks in my work.	[ ]	[ ]	[ ]	[ ]	[ ]
3. My concern/s distracted me from taking pleasure in my work.	[ ]	[ ]	[ ]	[ ]	[ ]
4. I felt hopeless about finishing certain work tasks, due to my concern/s.	[ ]	[ ]	[ ]	[ ]	[ ]
5. At work, I was able to focus on achieving my goals despite my concern/s.	[ ]	[ ]	[ ]	[ ]	[ ]
6. Despite having my concern/s, I felt energetic enough to complete all my work.	[ ]	[ ]	[ ]	[ ]	[ ]

Figure B.7.A Stanford Presenteeism Scale – 6 Question Version (SPS-6)

Stanford Presenteeism Scale – 13

1. **Do you have any of the following health conditions? Please check all that apply, and also check which one you consider your <u>primary condition</u> (*the condition that has affected you most in the past 4 weeks*). If you have none of these conditions, please mark this box and do not complete the survey.** (Choices: Allergies; Arthritis or joint pain/stiffness; Asthma; Back or neck disorder; Breathing disorder-bronchitis or emphysema; Depression anxiety or emotional disorder; Diabetes; Heart or circulatory problems-artery disease, high blood pressure, angina; Migraines/chronic headaches; Stomach or bowel disorder; and Other ____)

For 2 – 11: In thinking about how your <u>primary</u> condition affected your ability to do your job, how often <u>in the past 4 weeks</u>: (Always / Frequently / About half the time / Occasionally / Never / No answer)

2. **Were you able to finish hard tasks?**
3. **Did you find your attention wandering?**
4. **Were you able to focus on achieving work goals?**
5. **Did you feel energetic enough to work?**
6. **Were the stresses of your job hard to handle?**
7. **Did you feel hopeless about finishing your work?**
8. **Were you able to focus on finding a solution when unexpected problems arose in your work?**
9. **Did you need to take breaks from your work?**
10. **Were you able to work with other people on shared tasks?**
11. **Were you tired because you lost sleep?**

12. **Given your primary health condition, what percentage of your usual productivity level were you able to achieve while working over the <u>last 4 weeks</u>?** (Place X on continuous scale 1-100)

13. **Because of your <u>primary</u> condition as you identified in question 1, how many work hours did you miss in the <u>past 4 weeks</u>?** (Place X on continuous scale 0-40+)

Figure B.7.B Stanford Presenteeism Scale – 13 Question Version (SPS-13)

Unnamed Hepatitis Instrument

1 During the past 4 weeks, how many days have you not been able to work due to your hepatitis or its treatment?'

2 During the past 4 weeks, have you had to work shorter hours because of your hepatitis or its treatment?

3 During the past 4 weeks, were you less productive in your work activity because of your hepatitis or its treatment?

Figure B.8 Unnamed Hepatitis Instrument (UHI)

Work Limitations Questionnaire

In the past 2 weeks, how much of the time did your physical health or emotional problems make it difficult for you to do the following?

	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
a. do your work without stopping to take breaks or rests	❑1	❑2	❑3	❑4	❑5	❑0
b. stick to a routine or schedule	❑1	❑2	❑3	❑4	❑5	❑0
c. keep your mind on your work	❑1	❑2	❑3	❑4	❑5	❑0
d. speak with people in person, in meetings or on the phone	❑1	❑2	❑3	❑4	❑5	❑0
e. handle the workload	❑1	❑2	❑3	❑4	❑5	❑0

Note: Items a. and b. are from the Time Demands scale. Items c.and d. are from the Mental-Interpersonal Demands scale. Item e. is from the Output Demands scale.

In the past 2 weeks, how much of the time were you **ABLE TO DO** the following without difficulty caused by physical health or emotional problems?

(Mark one box on each line a. through f.)

	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
a. walk or move around different work locations (for example, go to meetings)	❑1	❑2	❑3	❑4	❑5	❑0
b. use hand-held tools or equipment (for example, a phone, pen, keyboard, computer mouse, drill, hairdryer, or sander)	❑1	❑2	❑3	❑4	❑5	❑0

Note: Items a. and b. are from the Physical Demands scale.

Figure B.9 Sample Items from the Work Limitations Questionnaire (WLQ)

Work Productivity and Activity Impairment General Health Questionnaire

The following questions ask about the effect of your health problems on your ability to work and perform regular activities. By health problems we mean any physical or emotional problem or symptom. *Please fill in the blanks or circle a number, as indicated.*

1. Are you currently employed (working for pay)? ____ NO ____ YES
   If NO, check "NO" and skip to question 6.

The next questions are about the **past seven days**, not including today.

2. During the past seven days, how many hours did you miss from work because of your health problems? *Include hours you missed on sick days, times you went in late, left early, etc., because of your health problems. Do not include time you missed to participate in this study.*

   _____HOURS

3. During the past seven days, how many hours did you miss from work because of any other reason, such as vacation, holidays, time off to participate in this study?

   _____HOURS

4. During the past seven days, how many hours did you actually work?

   _____HOURS *(If "0", skip to question 6.)*

Figure B.10.A Work Productivity and Activity Impairment - General Health Questionnaire (WPAI-GH)

5. During the past seven days, how much did your health problems affect your productivity while you were working?

Think about days you were limited in the amount or kind of work you could do, days you accomplished less than you would like, or days you could not do your work as carefully as usual. If health problems affected your work only a little, choose a low number. Choose a high number if health problems affected your work a great deal.

Consider only how much health problems affected productivity while you were working.

Health problems had no effect on my work	0 1 2 3 4 5 6 7 8 9 10	Health problems completely prevented me from working

CIRCLE A NUMBER

6. During the past seven days, how much did your health problems affect your ability to do your regular daily activities, other than work at a job?

By regular activities, we mean the usual activities you do, such as work around the house, shopping, childcare, exercising, studying, etc. Think about times you were limited in the amount or kind of activities you could do and times you accomplished less than you would like. If health problems affected your activities only a little, choose a low number. Choose a high number if health problems affected your activities a great deal.

Consider only how much health problems affected your ability to do your regular daily activities, other than work at a job.

Health problems had no effect on my daily activities	0 1 2 3 4 5 6 7 8 9 10	Health problems completely prevented me from doing my daily activities

CIRCLE A NUMBER

Figure B.10.A continued

The following questions ask about the effect of your PROBLEM on your ability to work and perform regular activities. Please fill in the blanks or circle a number, as indicated.

1. Are you currently employed (working for pay)? _____ NO ___ YES

If NO, check "NO" and skip to question 6.

The next questions are about the past seven days, not including today.

2. During the past seven days, how many hours did you miss from work because of problems associated with your PROBLEM? Include hours you missed on sick days, times you went in late, left early, etc., because of your PROBLEM. Do not include time you missed to participate in this study.

_____ HOURS

3. During the past seven days, how many hours did you miss from work because of any other reason, such as vacation, holidays, time off to participate in this study?

_____HOURS

4. During the past seven days, how many hours did you actually work?

_____HOURS (If "0", skip to question 6.)

Figure B.10.B Work Productivity and Activity Impairment - Specific Health Problem Questionnaire (WPAI-SHP)

5. During the past seven days, how much did your PROBLEM affect your productivity while you were working?

Think about days you were limited in the amount or kind of work you could do, days you accomplished less than you would like, or days you could not do your work as carefully as usual. If PROBLEM affected your work only a little, choose a low number. Choose a high number if PROBLEM affected your work a great deal.

Consider only how much PROBLEM affected productivity while you were working.

PROBLEM had no effect on my work	0 1 2 3 4 5 6 7 8 9 10	PROBLEM completely prevented me from working

CIRCLE A NUMBER

6. During the past seven days, how much did your PROBLEM affect your ability to do your regular daily activities, other than work at a job?

By regular activities, we mean the usual activities you do, such as work around the house, shopping, childcare, exercising, studying, etc. Think about times you were limited in the amount or kind of activities you could do and times you accomplished less than you would like. If PROBLEM affected your activities only a little, choose a low number. Choose a high number if PROBLEM affected your activities a great deal.

Consider only how much PROBLEM affected your ability to do your regular daily activities, other than work at a job.

PROBLEM had no effect on my daily activities	0 1 2 3 4 5 6 7 8 9 10	PROBLEM completely prevented me from doing my daily activities

CIRCLE A NUMBER

Figure B.10.B continued

Worker Productivity Index

Lost Hours per Week Due to Absenteeism =
(Total Illness Hours + Total STD* Hours)/Weeks Employed

Lost Hours per week Due to Failure to Meet Productivity Standard =
[100% − (Overall Score†)/(0.5)] × Average Weekly Staffing Hours

Total Lost Hours per Week =
Lost Hours per Week Due to Absenteeism
+
Lost Hours per Week Due to Failure to Meet Productivity Standard

---

* STD, short-term disability.
† If < 0.5.

Figure B.11 Worker Productivity Index (WPI)

APPENDIX C: HQWP RECRUITMENT LETTER

June 2007

Dear <Name>:

You are being asked to be a volunteer in a research study. Health and its effect on productivity is one of the most pressing questions facing employers today. We are performing a study to examine this relationship and its implications. Results may help employers become better at managing all aspects of the health of their employees. You have been randomly selected from the Georgia Tech employee population to help us with this study, and **we need your help to make this study successful**.

This survey should take approximately **10-15 minutes** to complete, and no elements of this survey have foreseeable risks. Any personally identifiable information will be kept confidential, and the survey itself is a web-based survey secured using 256-bit SSL encryption. We are offering the opportunity to participate in a **raffle for one of four 30GB (video) iPods** by submitting your name and contact email or phone number in the raffle section at the end of the survey (one entry per person). Note that completion of the survey is not necessary in order to enter the raffle. **Minors are not eligible to participate in this study** but are eligible to enter the raffle.

Please be frank. This instrument can only provide useful information if your answers are realistic. There are no costs to you except your time. Your participation in this study is voluntary. You have the right to change your mind and leave the study at any time without giving any reason and without penalty. Staff from the Health Systems Institute will collect, compile, and report the survey results in aggregate only. **Individual survey responses will be kept confidential, and no one at your workplace will ever see your answers.**

Should you have any questions about the study, you may contact David Huang, Georgia Tech Project Manager, at (404) 385-0140 or david.huang@hsi.gatech.edu. If you have any questions about your rights as a research subject, you may contact Melanie Clark, Georgia Tech Office of Research Compliance at (404) 894-6942 or melanie.clark@gtrc.gatech.edu.

The secure link to the survey is:
https://www.surveykey.com/s.cfm?SID=45730170-FF9E-87C0-4BBFA3599E07C90D

Should you prefer to receive and return a paper copy of the survey instead (with the same opportunity to enter the iPod drawing), please send an email to David Huang at david.huang@hsi.gatech.edu with your request. The deadline for completing either survey is July 16, 2007. **Thank you for your time and participation!**

If you complete the survey in the link above, it means that you have read -- or have had read to you -- the information contained in this letter, you are not a minor, and you would like to be a volunteer in this study.

Sincerely,

François Sainfort, Ph.D.
William H. George Professor of Health Systems
Georgia Institute of Technology

APPENDIX D: HQWP INSTRUMENT

HRQOL/Work Productivity Survey (HQWP)

Part 1: Health-Related Quality of Life (HRQOL)

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities. Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is: (circle one)
   - Excellent 1
   - Very good.....2
   - Good 3
   - Fair.....4
   - Poor 5

2. Compared to one year ago, how would you rate your health in general now? (circle one)
   - Much better now than one year ago.....1
   - Somewhat better than one year ago 2
   - About the same as one year ago 3
   - Somewhat worse now than one year ago 4
   - Much worse now than one year ago 5

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? *(circle one number on each line)*

ACTIVITIES	Yes, Limited a Lot	Yes, Limited a Little	No, Not Limited At All
a. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
b. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
c. Lifting or carrying groceries	1	2	3
d. Climbing **several** flights of stairs	1	2	3
e. Climbing **one** flight of stairs	1	2	3
f. Bending, kneeling, or stooping	1	2	3
g. Walking **more than a mile**	1	2	3
h. Walking **several blocks**	1	2	3
i. Walking **one block**	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health? *(circle one number on each line)*

	YES	NO
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Were limited in the **kind** of work or other activities	1	2
d. Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? *(circle one number on each line)*

	YES	NO
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Didn't do work or other activities as **carefully** as usual	1	2

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups? (circle one)

Not at all.....1
Slightly.....2
Moderately.....3
Quite a bit.....4
Extremely.....5

7. How much bodily pain have you had during the past 4 weeks? (circle one)

None.....1
Very mild.....2
Mild.....3
Moderate.....4
Severe.....5
Very severe.....6

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? (circle one)

Not at all.....1
A little bit.....2
Moderately.....3
Quite a bit.....4
Extremely.....5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks: *(circle one number on each line)*

	All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
a. Did you feel full of pep?	1	2	3	4	5	6
b. Have you been a very nervous person?	1	2	3	4	5	6
c. Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d. Have you felt calm and peaceful?	1	2	3	4	5	6
e. Did you have a lot of energy?	1	2	3	4	5	6
f. Have you felt downhearted and blue?	1	2	3	4	5	6
g. Did you feel worn out?	1	2	3	4	5	6
h. Have you been a happy person?	1	2	3	4	5	6
i. Did you feel tired?	1	2	3	4	5	6

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)? (circle one)

All of the time.....1
Most of the time.....2
Some of the time.....3
A little of the time.....4
None of the time.....5

11. How TRUE or FALSE is each of the following statements for you? *(circle one number on each line)*

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a. I seem to get sick a little easier than other people	1	2	3	4	5
b. I am as healthy as anybody I know	1	2	3	4	5
c. I expect my health to get worse	1	2	3	4	5
d. My health is excellent	1	2	3	4	5

Part 2: Work Productivity

INSTRUCTIONS: This survey asks questions regarding your productivity at work during the past 30 days. For questions with multiple choice answers, **circle the one response** that best fits. If you have not worked for pay in the past 30 days, please draw a diagonal line through the next three pages and proceed to **Part 3: Classification Questions**, beginning with **Question 30**.

12. How many full days of work did you miss in the past 30 days not including vacation or maternity leave?
    a. None *(go to question 16)*
    b. One
    c. More than one (specify number) _____

13. How many of these days were in the past week?
    a. None *(go to question 16)*
    b. One
    c. More than one (specify number) _____

14. Of the days missed in the past 30 days, was that because of problems with your own health, the health of someone else, or for some other reason? (circle all that apply)
    a. Own health
    b. Other's health
    c. Other reason

15. How many of the past 30 days did you miss because of problems with your own health? _____

16. How many days in the past 30 days did you either come in late for work or leave early?
    a. None *(go to question 20)*
    b. One
    c. More than one (specify number) _____

17. How many hours did you miss on that day or on average for each of those days? _____

18. Did you miss this time because of problems with your own health, the health of someone else, or for some other reason? (circle all that apply)
    a. Own health
    b. Other's health
    c. Other reason

19. For how many of the past 30 days was your reduced time at work because of problems with your own health?

_____

20. How many days in the past 30 did you either come in early, work late, or work on your day off in order to catch up on your work?

a. None *(go to question 23)*
b. One
c. More than one (specify number) _____ *(go to question 22)*
d. Don't know

21. How many extra hours of work did you put in that day? _____ (g*o to question 23)*

22. Altogether, about how many extra hours of work did you put in on those days combined? _____

23. The next questions are about the time you spent at work over the past 30 days. How often during that time did you have each of the following experiences (check one box in each row):

		All of the time	**Most of the time**	**About half of the time**	**Some of the time**	**A little of the time**	**None of the time**	**Don't know**
a	How much of the time was your speed of work or productivity higher than expected?							
b	How much of the time was your speed of work or productivity lower than expected?							
c	How much of the time did you do no work at times when you were supposed to be working?							
d	How much of the time did you find yourself not working as carefully as you should?							
e	How much of the time was the quality of your work lower than expected?							
f	How much of the time did you find yourself daydreaming and not concentrating on your work?							
g	How much of the time did you have trouble getting along with others at work?							
h	How much of the time did you have trouble controlling your emotions when you were around people at work?							
i	How much of the time did you get along well with others at work?							

24. During the time you were at work in the past 30 days, how often did health problems limit you in the kind or amount of work you could do compared to usual?

a. All of the time
b. Most of the time
c. About half of the time
d. Some of the time
e. A little of the time
f. None of the time
g. Don't know

25. The next set of questions ask you to rate job performance on a **scale from 0 to 10**, where 0 is the worst job performance anyone could have at your job and 10 is the performance of a top worker (check one box in each row):

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
a. How would you rate your overall job performance on the days you worked during the past **30** days?											
b. How would you rate your job performance during the past **7** days?											
c. How would **your peers** rate your job performance during the past **7** days?											
d. How would **your supervisor** rate your job performance during the past **7** days?											

26. How many days in the past 7 was your speed of work or productivity lower than expected?

a. All of the time
b. Most of the time
c. About half of the time
d. Some of the time
e. A little of the time
f. None of the time
g. Don't know

27. Did you experience any special work success or achievement at any time during the past 30 days?

a. Yes
b. No
c. Don't know

28. Did you have any special work failure, make any big mistakes, or miss a major deadline at any time during the past 30 days?

a. Yes
b. No
c. Don't know

29. Did you make any big mistake at work during the past 30 days that either caused an accident or that created a safety risk for yourself or for others?

a. Yes
b. No
c. Don't know

Part 3: Classification Questions

The following questions are classification purposes only.

30. What is your gender?
    - ☐ Male
    - ☐ Female

31. Which of the following includes your age?
    - ☐ Under 18
    - ☐ 18 - 24
    - ☐ 25 - 34
    - ☐ 35 - 44
    - ☐ 45 - 54
    - ☐ 55 - 64
    - ☐ 65+

32. Are you of Hispanic or Latino origin?
    - ☐ Yes
    - ☐ No

33. What is your race? (check all that apply)
    - ☐ American Indian or Alaska Native
    - ☐ Asian
    - ☐ Black or African American
    - ☐ Native Hawaiian or Other Pacific Islander
    - ☐ White
    - ☐ Other Race: _____

34. Which of the following best describes your main occupational classification?
    - ☐ Managerial
    - ☐ Professional and technical
    - ☐ Administrative and clerical
    - ☐ Production, construction, operations, maintenance, material handling
    - ☐ Service
    - ☐ Sales and related
    - ☐ Agricultural, forestry, fishing

35. How would you describe your job?
    - Full-time faculty
    - Full-time staff
    - Part-time faculty
    - Fart-time staff

Thank you for your participation!

Don't forget to fill in the following page to enter the raffle for one of four 30GB Video iPods!

Raffle

If you would like to enter the raffle for one of four 30GB video iPods, please check the appropriate box and include your name and a contact email and/or phone number. All information on this page will be collected separately from survey results and destroyed when the raffle has been held.

Check one box:

___ I would like to enter the raffle for one of four 30GB Video iPods.

Contact name: _____

Contact email and/or phone number: _____

(Note that if any winner cannot be contacted within 2 weeks of the survey conclusion, an alternate winner will be chosen.)

___ I do not desire to enter the raffle.

<The following text will only appear for respondents who choose to receive a paper copy of the survey.>

Please return all pages via campus mail to:

David Huang
HQWP Study
Health Systems Institute
Georgia Institute of Technology
Campus Mail Code 0477

Alternatively, you may return via U.S. mail to:

HQWP Study
Attn: David Huang
Health Systems Institute
Georgia Institute of Technology
901 Atlantic Dr., Suite 4100
Atlanta, GA 30332-0477

REFERENCES

1. Guyatt GH, Feeny DH, Patrick DL. Measuring Health-Related Quality-of-Life. *Annals of Internal Medicine.* Apr 1993;118(8):622-629.

2. Evans CJ, Mertzanis P, Abetz L. Measurement strategies for indirect costs in economic evaluations. *Expert Review of Pharmacoeconomics & Outcomes Research.* December 2003;3(6):703-716.

3. Chu C, Dwyer S. Employer role in integrative workplace health management - A new model in progress. *Disease Management & Health Outcomes.* 2002;10(3):175-186.

4. Cronbach LJ. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika.* September 1951;16(3):297-334.

5. Stephens MA. Edf Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association.* 1974;69(347):730-737.

6. Shrout PE, Fleiss JL. Intraclass Correlations - Uses in Assessing Rater Reliability. *Psychological Bulletin.* Mar 1979;86(2):420-428.

7. Field A. *Discovering Statistics Using SPSS*. Second ed: Sage Publications Ltd; 2005.

8. Cohen RJ, Swerdlik ME. *Psychological Testing and Assessment: An Introduction To Tests and Measurement*: McGraw-Hill; 2002.

9. Aiken LR, Groth-Marnat G. *Psychological Testing and Assessment*. 12th ed: Allyn & Bacon; 2005.

10. Kline P. *Handbook of Psychological Testing*. 2nd ed: Routledge; 2000.

11. Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study - an Application of Methods for Monitoring the Results of Medical-Care. *Jama-Journal of the American Medical Association.* Aug 1989;262(7):925-930.

12. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (Sf-36) .1. Conceptual-Framework and Item Selection. *Medical Care.* Jun 1992;30(6):473-483.

13. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (Sf-36) .2. Psychometric and Clinical-Tests of Validity in Measuring Physical and Mental-Health Constructs. *Medical Care.* Mar 1993;31(3):247-263.

14. Garratt AN, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal.* Jun 2002;324(7351):1417-1419.

15. Ware JE. SF-36.org: A community for measuring health outcomes using SF tools. http://www.sf-36.org. Accessed February 12, 2008.

16. McHorney CA, Ware JE, Lu JFR, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (Sf-36) .3. Tests of Data Quality, Scaling Assumptions, and Reliability across Diverse Patient Groups. *Medical Care.* Jan 1994;32(1):40-66.

17. von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press; 1947.

18. Essink-Bot M, Bonsel GJ. How to Derive Disability Weights. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, eds. *Summary Measures of Population Health*. Geneva: World Health Organization; 2002.

19. Fryback DG. Methodological Issues in Measuring Health Status and Health-related Quality of Life for Population Health Measures: A Brief Overview of the 'HALY' Family of Measures. In: Field MJ, Gold MR, eds. *Summarizing Population Health – Directions for the Development and Application of Population Metrics*. Washington, DC: National Academy Press; 1998.

20. Hazen G. Multiattribute Structure for QALYs: IEMS Department, Northwestern University; 2003.

21. Dolan P. Modelling the Relationship Between the Description and Valuation of Health States. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, eds. *Summary Measures of Population Health*. Geneva: World Health Organization; 2002.

22. Kongnakorn T, Sainfort F. Modeling Health Outcomes for Economic Analysis. In: Brandeau ML, Sainfort F, Pierskalla WP, eds. *Operations Research and Health Care : A Handbook of Methods and Applications*: Springer; 2004.

23. Marinus J, Ramaker C, van Hilten JJ, Stiggelbout AM. Health related quality of life in Parkinson's disease: a systematic review of disease specific instruments. *Journal of Neurology Neurosurgery and Psychiatry.* Feb 2002;72(2):241-248.

24. Revicki DA, Irwin D, Reblando J, Simon GE. The Accuracy of Self-Reported Disability Days. *Medical Care.* Apr 1994;32(4):401-404.

25. Hackett RD. Work Attitudes and Employee Absenteeism - a Synthesis of the Literature. *Journal of Occupational Psychology.* Sep 1989;62(3):235-248.

26. Johns G. Absenteeism Estimates by Employees and Managers - Divergent Perspectives and Self-Serving Perceptions. *Journal of Applied Psychology.* Apr 1994;79(2):229-239.

27. Johns G. How Often Were You Absent - a Review of the Use of Self-Reported Absence Data. *Journal of Applied Psychology.* Aug 1994;79(4):574-591.

28. Pauly MV, Nicholson S, Xu J, et al. A general model of the impact of absenteeism on employers and employees. *Health Economics.* Apr 2002;11(3):221-231.

29. Brouwer WBF, van Exel NJA, Koopmanschap MA, Rutten FFH. Productivity costs before and after absence from work: as important as common? *Health Policy.* Aug 2002;61(2):173-187.

30. Koopman C, Pelletier KR, Murray JF, et al. Stanford presenteeism scale: Health status and employee productivity. *Journal of Occupational and Environmental Medicine.* Jan 2002;44(1):14-20.

31. Collins JJ, Baase CM, Sharda CE, et al. The assessment of chronic health conditions on work performance, absence, and total economic impact for employers. *Journal of Occupational and Environmental Medicine.* Jun 2005;47(6):547-557.

32. Goetzel RZ, Long SR, Ozminkowski RJ, Hawkins K, Wang SH, Lynch W. Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting US employers. *Journal of Occupational and Environmental Medicine.* Apr 2004;46(4):398-412.

33. Evans CJ, Crawford B. Data Collection Methods in Prospective Economic Evaluations: How Accurate Are the Results? *Value in Health.* July/August 2000;3(4):277-286.

34. Lynch W, Riedel JE. *Measuring Employee Productivity: A guide to self-assessment tools.* Scottsdale, Arizona: Institute for Health and Productivity Management; 2001.

35. Loeppke R, Hymel PA, Lofland JH, et al. Health-related Workplace productivity measurement: General and migraine-specific recommendations from the ACOEM expert panel. *Journal of Occupational and Environmental Medicine.* Apr 2003;45(4):349-359.

36. Ozminkowski RJ, Goetzel RZ, Chang S, Long S. The application of two health and productivity instruments at a large employer. *Journal of Occupational and Environmental Medicine.* Jul 2004;46(7):635-648.

37. Prasad M, Wahlqvist P, Shikiar R, Shih YCT. A review of self-report instruments measuring health-related work productivity - A patient-reported outcomes perspective. *Pharmacoeconomics.* 2004;22(4):225-244.

38. Chapman LS. Presenteeism and its role in worksite health promotion. *American Journal of Health Promotion.* Mar-Apr 2005;19(4):1-8.

39. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics.* 2004;22(3):165-184.

40. Lerner DJ, Amick BC, Malspeis S, Rogers WH, Gomes DRJ, Salem DN. The angina-related limitations at work questionnaire. *Quality of Life Research.* Jan 1998;7(1):23-32.

41. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): A new measure to assess treatment effects. *Psychopharmacology Bulletin.* 1997;33(1):13-16.

42. vanRoijen L, EssinkBot ML, Koopmanschap MA, Bonsel G, Rutten FFH. Labor and health status in economic evaluation of health care - The health and labor questionnaire. *International Journal of Technology Assessment in Health Care.* Sum 1996;12(3):405-415.

43. vanRoijen L, Essinkbot ML, Koopmanschap MA, Michel BC, Rutten FFH. Societal Perspective on the Burden of Migraine in the Netherlands. *Pharmacoeconomics.* Feb 1995;7(2):170-179.

44. Meerding WJ, Ijzelenberg W, Koopmanschap MA, Severens JL, Burdorf A. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. *Journal of Clinical Epidemiology.* May 2005;58(5):517-523.

45. Kessler RC, Barber C, Beck A, et al. The world health organization health and work performance questionnaire (HPQ). *Journal of Occupational and Environmental Medicine.* Feb 2003;45(2):156-174.

46. Kessler RC, Ames M, Hymel PA, et al. Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to evaluate the indirect workplace costs of illness. *Journal of Occupational and Environmental Medicine.* Jun 2004;46(6):S23-S37.

47. Wang PS, Beck A, Berglund P, et al. Chronic medical conditions and work performance in the health and work performance questionnaire calibration surveys. *Journal of Occupational and Environmental Medicine.* Dec 2003;45(12):1303-1311.

48. Kessler RC. HPQ Data Consortium. http://www.hpq.org. Accessed February 12, 2008.

49. Halpern MT, Shikiar R, Rentz AM, Khan ZM. Impact of smoking status on workplace absenteeism and productivity. *Tobacco Control.* Sep 2001;10(3):233-238.

50. Shikiar R, Halpern MT, Rentz AM, Khan ZM. Development of the Health and Work Questionnaire (HWQ): an instrument for assessing workplace productivity in relation to worker health. *Work.* 2004;22(3):219-229.

51. Lerner DJ, Amick BC, Malspeis S, et al. The migraine work and productivity loss questionnaire: Concepts and design. *Quality of Life Research.* Dec 1999;8(8):699-710.

52. Hartmaier SL, Santanello NC, Epstein RS, Silberstein SD. Development of a Brief 24-Hour Migraine-Specific Quality-of-Life Questionnaire. *Headache.* Jun 1995;35(6):320-329.

53. Santanello NC, Hartmaier SL, Epstein RS, Silberstein SD. Validation of a New Quality-of-Life Questionnaire for Acute Migraine Headache. *Headache.* Jun 1995;35(6):330-337.

54. Davies GM, Santanello N, Gerth W, Lerner D, Block GA. Validation of a migraine work and productivity loss questionnaire for use in migraine studies. *Cephalalgia.* Jun 1999;19(5):497-502.

55. Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare Resource and Lost Labour Costs of Migraine Headache in the US. *Pharmacoeconomics.* Jul 1992;2(1):67-76.

56. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity losses without absence: measurement validation and empirical evidence. *Health Policy.* Jul 1999;48(1):13-27.

57. Turpin RS, Ozminkowski RJ, Sharda CE, et al. Reliability and validity of the Stanford Presenteeism Scale. *Journal of Occupational and Environmental Medicine.* Nov 2004;46(11):1123-1133.

58. McHutchison JG, Ware JE, Bayliss MS, et al. The effects of interferon alpha-2b in combination with ribavirin on health related quality of life and work productivity. *Journal of Hepatology.* Jan 2001;34(1):140-147.

59. Stewart WF, Ricci JA, Leotta C. Health-related lost productive time (LPT): Recall interval and bias in LPT estimates. *Journal of Occupational and Environmental Medicine.* Jun 2004;46(6):S12-S22.

60. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *Jama-Journal of the American Medical Association.* Jun 2003;289(23):3135-3144.

61. Stewart WF, Ricci JA, Chee E, Morganstein D. Lost productive work time costs from health conditions in the United States: Results from the American productivity audit. *Journal of Occupational and Environmental Medicine.* Dec 2003;45(12):1234-1246.

62. Stewart WF, Ricci JA, Leotta C, Chee E. Validation of the work and health interview. *Pharmacoeconomics.* 2004;22(17):1127-1140.

63. Burton WN, Chen CY, Conti DJ, Schultz AB, Edington DW. The association between health risk change and presenteeism change. *Journal of Occupational and Environmental Medicine.* Mar 2006;48(3):252-263.

64. Burton WN, Chen CY, Conti DJ, Schultz AB, Pransky G, Edington DW. The association of health risks with on-the-job productivity. *Journal of Occupational and Environmental Medicine.* Aug 2005;47(8):769-777.

65. Burton WN, Pransky G, Conti DJ, Chen CY, Edington DW. The association of medical conditions and presenteeism. *Journal of Occupational and Environmental Medicine.* Jun 2004;46(6):S38-S45.

66. Lerner D, Reed JI, Massarotti E, Wester LM, Burke TA. The Work Limitations Questionnaire's validity and reliability among patients with osteoarthritis. *Journal of Clinical Epidemiology.* Feb 2002;55(2):197-208.

67. Lerner D, Amick BC, Lee JC, et al. Relationship of employee-reported work limitations to work productivity. *Medical Care.* May 2003;41(5):649-659.

68. Lerner D, Amick BC, Rogers WH, Malspeis S, Bungay K, Cynn D. The work limitations questionnaire. *Medical Care.* Jan 2001;39(1):72-85.

69. Reilly MC. Reilly Associates. http://www.reillyassociates.net. Accessed February 12, 2008.

70. Reilly MC, Zbrozek AS, Dukes EM. The Validity and Reproducibility of a Work Productivity and Activity Impairment Instrument. *Pharmacoeconomics.* Nov 1993;4(5):353-365.

71. Reilly MC, Tanner A, Meltzer EO. Work, classroom and activity impairment instruments - Validation studies in allergic rhinitis. *Clinical Drug Investigation.* May 1996;11(5):278-288.

72. Chirban JT, Jacobs RJ, Warren J, et al. The 36-Item Short Form Health Survey (SF-36) and Work Productivity and the Impairment (WPAI) Questionnaire in Panic Disorder. *Disease Management and Health Outcomes.* March 1997;1(3):154-164.

73. Wahlqvist P, Carlsson J, Stalhammar NO, Wiklund I. Validity of a Work Productivity and Activity Impairment questionnaire for patients with symptoms of

gastro-esophageal reflux disease (WPAI-GERD) - Results from a cross-sectional study. *Value in Health.* Mar-Apr 2002;5(2):106-113.

74. Wahlqvist P, Guyatt GH, Armstrong D, et al. The work productivity and activity impairment questionnaire for patients with gastroesophageal reflux disease (WPAI-GERD) - Responsiveness to change and English language validation. *Pharmacoeconomics.* 2007;25(5):385-396.

75. Burton WN, Conti DJ, Chen CY, Schultz AB, Edington DW. The role of health risk factors and disease on worker productivity. *Journal of Occupational and Environmental Medicine.* Oct 1999;41(10):863-877.

76. Goetzel RZ, Ozminkowski RJ, Long SR. Development and reliability analysis of the Work Productivity Short Inventory (WPSI) instrument measuring employee health and productivity. *Journal of Occupational and Environmental Medicine.* Jul 2003;45(7):743-762.

77. Ozminkowski RJ, Goetzel RZ, Long SR. A validity analysis of the work productivity short inventory (WPSI) instrument measuring employee health and productivity. *Journal of Occupational and Environmental Medicine.* Nov 2003;45(11):1183-1195.

78. Brouwer WBF, Meerding WJ, Lamers LM, Severens JL. The relationship between productivity and health-related QOL - An exploration. *Pharmacoeconomics.* 2005;23(3):209-218.

79. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press; 1996.

80. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs measurement through quality of life? A response to the recommendation of the Washington Panel. *Health Economics.* May-Jun 1997;6(3):253-259.

81. Lamers LM, Meerding WJ, Severens JL, Brouwer WBF. The relationship between productivity and health-related quality of life: An empirical exploration in persons with low back pain. *Quality of Life Research.* Apr 2005;14(3):805-813.

82. Allgulander C, Jorgensen T, Wade A, et al. Health-related quality of life (HRQOL) among patients with Generalised Anxiety Disorder: evaluation conducted alongside an escitalopram relapse prevention trial. *Current Medical Research And Opinion.* Oct 2007;23(10):2543-2549.

83. Kantz ME, Harris WJ, Levitsky K, Ware JE, Davies AR. Methods for Assessing Condition-Specific and Generic Functional Status Outcomes after Total Knee Replacement. *Medical Care.* May 1992;30(5):MS240-MS252.

84. Nerenz DR, Repasky DP, Whitehouse FW, Kahkonen DM. Ongoing Assessment of Health-Status in Patients with Diabetes-Mellitus. *Medical Care.* May 1992;30(5):MS112-MS124.

85. Wagner AK, Keller SD, Kosinski M, et al. Advances in Methods for Assessing the Impact of Epilepsy and Antiepileptic Drug-Therapy on Patients Health-Related Quality-of-Life. *Quality of Life Research.* Apr 1995;4(2):115-134.

86. Ilieva J, Baron S, Healey NM. Online surveys in marketing research: pros and cons. *International Journal of Market Research.* 2002;44(3):361-376.

87. Couper MP. Web surveys - A review of issues and approaches. *Public Opinion Quarterly.* Win 2000;64(4):464-494.

88. Saarni SI, Suvisaari J, Sintonen H, Koskinen S, Harkanen T, Lonnqvist J. The health-related quality-of-life impact of chronic conditions varied with age in general population. *Journal Of Clinical Epidemiology.* Dec 2007;60(12):1288-1297.

89. Feyrer J. Aggregate evidence on the link between age structure and productivity. *Population And Development Review.* 2008;34:78-99.

90. Hodgson TA, Meiners MR. Cost-of-Illness Methodology - a Guide to Current Practices and Procedures. *Milbank Memorial Fund Quarterly-Health and Society.* 1982;60(3):429-462.

91. Hellerstein JK, Neumark D, Troske KR. Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations. *Journal Of Labor Economics.* Jul 1999;17(3):409-446.

92. Papoulis A, Pillai SU. *Probability, Random Variables and Stochastic Processes*. 4th ed. Boston: McGraw-Hill; 2002.

93. Hair JF, Black B, Babin B, Anderson RE, Tatham RL. *Multivariate Data Analysis*: Prentice Hall; 2005.

94. Norušis MJ. *SPSS 16.0 Advanced Statistical Procedures Companion*: Prentice Hall; 2008.

95. Durbin J, Watson GS. Testing For Serial Correlation In Least Squares Regression .2. *Biometrika.* 1951;38(1-2):159-178.

96. Field MJ, Gold MR, eds. *Summarizing Population Health – Directions for the Development and Application of Population Metrics*. Washington, DC: National Academy Press; 1998.

97. Brown TA. *Confirmatory Factor Analysis for Applied Research*: The Guilford Press; 2006.

98. Cyr A, Hagen S. Measurement and quantification of presenteeism. *Journal Of Occupational And Environmental Medicine.* Dec 7 2007;49(12):1299-1300.

99. Liljas B. How to calculate indirect costs in economic evaluations. *Pharmacoeconomics.* Jan 1998;13(1):1-7.

100. Nicholson S, Pauly MV, Polsky D, et al. How to present the business case for healthcare quality to employers. *Appl Health Econ Health Policy.* 2005 2005;4(4):209-218.

101. Nicholson S, Pauly MV, Polsky D, Sharda C, Szrek H, Berger ML. Measuring the effects of work loss on productivity with team production. *Health Economics.* Feb 2006;15(2):111-123.

102. Berger ML, Murray JF, Xu J, Pauly M. Alternative valuations of work loss and productivity. *Journal of Occupational and Environmental Medicine.* Jan 2001;43(1):18-24.

103. Berger ML, Howell R, Nicholson S, Sharda C. Investing in healthy human capital. *Journal of Occupational and Environmental Medicine.* Dec 2003;45(12):1213-1225.

104. Holzer HJ. The Determinants Of Employee Productivity And Earnings. *Industrial Relations.* Fal 1990;29(3):403-422.

105. Ichniowski C, Shaw K, Prennushi G. The effects of human resource management practices on productivity: A study of steel finishing lines. *American Economic Review.* Jun 1997;87(3):291-313.

106. Lazear EP. Performance pay and productivity. *American Economic Review.* Dec 2000;90(5):1346-1361.

107. Bloom DE, Canning D. Public health - The health and wealth of nations. *Science.* Feb 2000;287(5456):1207-+.

108. Tsai SP, Ahmed FS, Wendt JK, Bhojani F, Donnelly RP. The impact of obesity on illness absence and productivity in an industrial population of petrochemical workers. *Annals Of Epidemiology.* Jan 2008;18(1):8-14.

109. Stewart WF, Ricci JA, Chee E, Hirsch AG, Brandenburg NA. Lost productive time and costs due to diabetes and diabetic neuropathic pain in the US workforce. *Journal of Occupational and Environmental Medicine.* Jun 2007;49(6):672-679.

VITA

David T. Huang was born in Denver, Colorado on August 15, 1978 and grew up in Englewood, Colorado and Irvine, California. He received a B.S.E. in electrical and computer engineering, computer science, and mathematics from Duke University, in Durham, North Carolina in 1998 and M.S. in operations research from Georgia Tech in 2000. He also worked in the telecommunications industry for two years before returning to Georgia Tech to pursue his doctorate in industrial and systems engineering. During his time at Georgia Tech, he developed a significant interest in public health, which led him to complete an M.P.H. in quantitative methods from the Harvard School of Public Health in Boston, Massachusetts in 2008. Mr. Huang is an avid runner and has completed four marathons, including the 2008 Boston Marathon. He also enjoys travel, reading, guitar, and violin.