Illinois Workplace Wellness Study

Pre-Analysis Plan

Damon Jones, University of Chicago and NBER
David Molitor, University of Illinois at Urbana-Champaign and NBER
Laura Payne, University of Illinois at Urbana-Champaign
Julian Reif, University of Illinois at Urbana-Champaign

July 11, 2016

1. Introduction

This plan outlines the hypotheses to be tested in the analysis of the impact and behavioral response to the introduction of a comprehensive workplace wellness program at the University of Illinois at Urbana-Champaign (UIUC). This pre-analysis plan has been created prior to the collection of data.

2. Overview of the Study

a. Motivation

Workplace wellness programs have become a $6 billion industry and are widely touted as a way to improve employee well-being, reduce health care costs by promoting prevention, and increase workplace productivity. Yet, there is little rigorous evidence available to support these claims, partly because the voluntary nature of these programs means that participants may differ from nonparticipants for reasons unrelated to the causal effects of the wellness program. We will implement a randomized control trial to identify the effects of incentives on wellness program participation, produce causal estimates of the effect of wellness programs on health outcomes, determine what kinds of employees benefit from wellness programs the most, and test for the presence of peer effects in wellness participation.

b. Experimental Design

Our experiment consists of a baseline survey, followed by random assignment to a control or one of four treatment groups, A, B, C, or D. Individuals assigned to a treatment group are offered incentives to complete a biometric screening and health risk assessment. Thereafter, members of the treatment group are further given incentives to participate in up to two wellness programs, one in each semester of the school year. One year later, we will follow up with a subset of study participants, again offering a follow-up survey and biometric screening. Survey, biometric screening, and health risk assessment data will be combined with administrative data from health insurers and the university human

resources department.

At baseline, employees will be invited via postcard and email to participate in the study by completing an online survey. In return for completing the survey, participants will be given a $30 Amazon.com gift card. Consent will be obtained before the survey is taken.

Following the baseline survey, employees will be assigned to either the control group or a treatment group. Members of the control group have no further intervention in the first year of the program. Members of the treatment group will be invited to participate in the **iThrive** wellness program consisting of the following:

i. First, they are given the opportunity to participate in a biometric screening and health risk assessment. The biometric screening is scheduled on campus and takes approximately 15 minutes. The biometric test will measure: (1) anthropometrics such as height, weight, and waist circumference (to assess obesity and overweight status); (2) resting blood pressure (to assess hypertension); (3) blood glucose (to assess diabetes risk); and (4) total, LDL, and HDL cholesterol levels, total cholesterol ratio, and triglycerides (to assess risk of cardiovascular disease). Biometric screening is carried out by a third-party vendor, Presence Health.

ii. After completing the biometric screening, the participants are invited to complete an online health risk assessment (HRA). The HRA is questionnaire designed to identify areas of health improvement, by asking a series of questions related to wellness, health status, nutrition, healthy activities, desire to improve health, preventative health measures. The HRA is also prepopulated with biometric information from the above screening. Upon completion, participants are given customized feedback on areas of improvement. The HRA is administered by a third-party vendor, Wellsource.

iii. Upon completion of the HRA, participants are given the option to enroll in up to two wellness courses, one in the fall semester and one in the spring semester. Courses are designed by the UI Wellness Center and include an Active Living class; self-paced online health challenges in physical activity, weight management, and healthy eating; a weight management class; a tobacco cessation hotline; a stress management class; a Tai Chi class; and a chronic disease management class.

Members of the treatment group will be offered financial incentives for completion of the different stages of the iThrive program. Treatment groups A, B and C will receive $0, $100, and $200, respectively, upon successfully completing

both the biometric screening and HRA. Within each treatment group, half of the participants will be offered $25 for each wellness course completed, for up to two courses. The other half will be offered $75 for completion of each wellness course. All incentives will be made known to treatment group members at the onset of treatment group assignment. Throughout the first year of the program, members of the treatment group will have access to an online portal that provides information on treatment group assignment, current progress, accrued incentives, scheduling for biometric screening, HRA access, and wellness program enrollment.

A final treatment group, D, will be offered the same incentives as groups A, B, and C, but will feature a clustered design. Treatment will be administered at the peer network level, based on responses to the baseline survey regarding workplace peers. Relative to the treatment groups A, B, and C, members of treatment group D will only have one member (of approximately 3 members) of a given peer group assigned to treatment. The remaining members will be assigned to the control group. Within the clustered design, assignment to the incentive structure of groups A, B, or C will be equally likely.

One year after the launch of the study, we will administer a follow-up survey and biometric screening among a subset of control group members and members of treatment group C. Members of the follow-up study will be randomly offered either $25 or $35 to complete a follow-up survey. We will also randomly offer $150 or $200 for completion of a biometric screening at follow-up, with equal probability. Follow-up study participants will be made aware of these incentives after the first year of the study is complete, and prior to the launch of the follow-up.

The timeline of the study is as follows:

i. **July 11 – Aug 1, 2016:** Baseline Survey is administered online
ii. **August 1 – August 7, 2016:** Treatment assignment
iii. **August 8 – September 9, 2016:** Biometric screenings
iv. **August 22 – September 23, 2016:** HRA is administered online
v. **September 26 – December 2, 2016:** Fall wellness courses
vi. **January – May, 2017:** Spring wellness courses
vii. **July – August, 2017**: Follow-up survey
viii. **August – September, 2017**: Follow-up biometric screening

c. Sample Selection and Treatment Assignment

The initial pool for our study includes 12,459 benefits eligible employees at the UIUC. Specifically, the set of eligible employees include all University of Illinois employees satisfying three criteria as of June 10, 2016: a) physically located on

the UIUC campus, b) not terminated, and c) eligible for benefits through the Illinois Department of Central Management Services. From this set of employees, 15 were excluded due to their direct involvement with the approval, implementation or design of the study—members of the research team, members of the IRB review panel involved in the study design, staff directly involved in collecting program data, and family members of the research team.

Of the 12,459 employees in the initial pool, the "Core Sample" will consist of all employees who respond to a baseline survey. We estimate a response rate of approximately 50%, resulting in a Core Sample of 6,000 employees. All subsample sizes below will be based on this estimate of a 50% baseline survey response rate. Employees will then be assigned to one of 5,600 clusters, based on treatment assignment. There are two types of treatment assignment:

i. **Individual Cluster:** 5,400 of the employees will be assigned to either a control group, or one of three treatment groups, at the individual level. In particular, 2,700 employees will be assigned to the control group, 900 will be assigned to treatment group A, 900 will be assigned to treatment group B, and 900 will be assigned to treatment group C.

ii. **Peer Effect Cluster:** 600 employees will be assigned to a fourth group, treatment group D, clustered at the peer group level. Peer groups will be identified based on self-reported peer networks, measured using the baseline survey. We expect an average peer-groups size of 3, resulting in 200 additional clusters.

In the second year of the study, members of the control group and treatment group C will be invited to participate in a follow-up survey and biometric screening. We will invite 600 members of the control group and 600 members of treatment group C.

At the time of taking the baseline survey, participants will be informed that they may be contacted for follow up treatments, but will not know control or treatment group status. After the baseline survey, employees will be assigned to a treatment or control group using pseudo-random numbers generated by a computer program according to the following steps:

i. Peer networks will be identified using baseline survey responses.
ii. Among the set of peer networks, we will randomly select 200 networks to be assigned to treatment group D. Randomization will be stratified by peer network size, age, gender, and employment class.
iii. The remaining employees will be randomly assigned to the control group or treatment group A, B or C. Randomization will be stratified by age, gender, annual salary, race, and employment class (Faculty,

Academic Professional, and Civil Service). We will require there to be at least eight people per strata (two for each control and/or treatment group) in each treatment cell. If there are fewer than eight people, then we will aggregate strata as necessary.

iv. One year following the original study, we will randomly choose 600 members of the control group and 600 members of treatment group C for our follow up sample.

3. Data Sources

a. Baseline Survey (A): Self-reported health, workplace, and demographic information will be collected via an online survey with up to 66 questions. Only a subset of questions are answered by participants, based on skip logic.

b. Health Insurance Claims Data (B): As a part of consenting to our study, participants will grant us access to health insurance claims data. These include total costs of services, bill amounts, and diagnosis codes. Health insurance data will be collected from 2015 – 2020. Claims data are currently only available for employees in two insurance companies, which together comprise 70% of employees in the initial pool with health benefits through the University.

c. Human Resources Data (C): Participants will also grant us access to UIUC human resources data, including absenteeism, turnover, employment unit, department, tenure, salary, age, race, sex, and benefit elections.

d. Biometric Screening (D): A third-party vendor will measure height, weight, waist circumference, resting blood pressure, total cholesterol, total cholesterol ratio, HDL cholesterol, LDL cholesterol, triglycerides, glucose levels. In addition, a set of questions will be asked prior to the revelation of the results of the biometric screening, to measure expectations of weight, height, cholesterol level, blood pressure, glucose level, and body mass index. A second biometric screening will be administered among a subset of participants during the second year of the study.

e. Health Risk Assessment (E): The health risk assessment survey will measure self-reported wellness, health status, nutrition, healthy activities, desire to improve health, preventative health measures. Answers to the adaptive survey are combined to create customized indices of health risk.

f. Wellness Course Data (F): We will collect data on enrollment, participation and completion of wellness courses. Enrollment is recorded using our online registration system. Participation and completion are monitored by wellness instructors. Completion is defined as participation in at least 80% of the weeks

that a course is provided.

g. **Follow-Up Survey (G):** Identical to baseline survey, administered to a subset of the members of the control group and treatment group C.

h. **Follow-Up Biometric Screening (H):** Identical to the first biometric screening, similarly administered to a subset of the members of the control group and treatment group C.

4. Hypotheses

Using our combined data, we will identify the effects of incentives on wellness program participation, produce causal estimates of the effect of wellness programs on health outcomes, determine what kinds of employees benefit the most from wellness programs, and test for the presence of peer effects in wellness participation. See Appendix 1 for a glossary of data source references (e.g. A1 is question 1 on our baseline survey). Our hypotheses can be summarized in the following 5 groups:

a. **Participation Outcomes:** As we increase the incentives for completion of the biometric screening, HRA, and wellness programs, we expect the level of participation to increase.

b. **Selection Outcomes:** Participants may differ from the average employee, in terms of health, health care utilization, and productivity. The average level of health among wellness participants compared to all baseline survey participants is theoretically ambiguous. Under **average** advantageous selection, participants are healthier, lower cost, and more productive, while under **average** adverse selection, the opposite is true. In addition, the marginal participant may differ from the average participant, causing the composition of participants to differ as incentives for participation are varied. Under **marginal** advantageous selection, higher incentives draw in healthier, lower cost, and more productive participants, while under **marginal** adverse selection, the opposite is true.

c. **Short-Run Health and Productivity Outcomes:** Take-up of biometric screening, HRA, and wellness programs may potentially increase health, wellbeing, satisfaction, and productivity measures measured one year following the intervention. Even though health measures may improve, health care utilization may also increase, causing an ambiguous effect on health care costs in the short-run.

d. **Peer Effects:** Having a close workplace friend or colleague who also participates in a biometric screening, HRA, or wellness activity may increase one's own participation in wellness programs.

e. **Heterogeneous Treatment Effects:** The baseline characteristics of participants may cause differential impacts of the incentives on participation and health outcomes. In addition, the extent to which biometric screening information deviates from elicited expectations may cause differential impacts on wellness participation and health outcomes.

Hypothesis Group A: *Incentives for participation in the biometric screening, HRA, and wellness programs will have a positive effect on participation, increasing in the size of incentive.*

The following indicators will comprise the family of outcomes in this domain:

Participation Outcomes:

i. Scheduled a biometric screening (D10)
ii. Completed a biometric screening (D11)
iii. Completed HRA (E1)
iv. Enrolled in a fall wellness course (F1)
v. Completed fall wellness course (F2)
vi. Enrolled in a spring wellness course (F3)
vii. Complete spring wellness course (F4)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis Group B: *In the case of **average** advantageous (adverse) selection, employees who select into wellness programs have higher (lower) baseline health, lower (higher) health costs, and higher (lower) productivity. Consequently, differences in baseline health and productivity between participants and non-participants will be positive (negative), while the analogous difference in baseline health costs will be (negative) positive. In addition, under **marginal** advantageous (adverse) selection, the differences in baseline health and productivity between participants and non-participants will be negatively (positively) correlated with incentive size, while the analogous difference in baseline health costs will be positively (negatively) correlated with incentive size.*

Hypothesis B.1: Under **average** advantageous (adverse) selection, the difference in baseline health measures between participants and non-participants will be positive (negative). Under **marginal** advantageous (adverse) selection, the difference in baseline health measures between participants and non-participants will be negatively (positively) correlated with incentive size.

The following variables will be used to measure baseline health:

Baseline Survey:

i. Had at least one previous health screening (Any of A1-A5, A8-A9=”yes”)
ii. Physically active (A11=”More active”)
iii. Trying to be more active (A12=”Yes” or A13=”Yes”)
iv. Smoking status:
    1. Current smoker: A16=”Yes” and A17=”Every day” or “Some days”
    2. Former smoker: A16=”Yes” and A17”=”Not at all”
v. Other tobacco use (A22 and A23 != “Not at all”)
vi. Drinking:
    1. Drinker: A24!=0
    2. Heavy drinker: A25>=4 if female, A25>=5 if male
vii. Has at least one chronic health condition (A27)
viii. Self-reported health (A28)
    1. Health is excellent or very good
    2. Health is not poor
ix. Problems with physical activities or pain (A29-A31)
    1. A29=”Somewhat”, “Quite a lot”, “Could not do physical activities” or A30 = “Some”, “Quite a lot”, “Could not do daily work” or A31=”Mild”, ”Moderate”, “Severe”, “Very severe”
x. Energy (A32=”An extraordinary amount”, “Quite a lot”)
xi. Emotional health (A33=”Moderately”, ”Quite a lot”, “Extremely”)
xii. Overweight status (A39=”Overweight” or A40=”Very overweight”)
xiii. Bad health status (A40=”High”, =”Very high” or A41=”High” or “Very high” or A42=”High” or “Very high”)
xiv. Sedentary job (A53=”None at all” or =”Some, but less than 1 hour”)

These baseline health variables can be divided up into two different domains: primary outcomes of interest, and secondary outcomes of interest. The first domain (“primary outcomes”) will include the following variables: current smoker, has at least one chronic health condition, both self-reported health questions, overweight status, and problems with physical activities or pain. Any baseline health variables not in the first domain will be in the second domain (“secondary outcomes”). A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis B.2: Under **average** advantageous (adverse) selection, the difference in baseline health costs and utilization between participants and non-participants will be negative (positive). Under **marginal** advantageous (adverse) selection, the difference in baseline health costs and utilization between participants and non-participants will be negatively (positively) correlated with incentive size.

The following variables will be used to measure baseline health costs and utilization:

Baseline Survey:

i. Drug utilization (A34>0 or A35>0)
ii. Physician or ER utilization (A36!=”None”)
iii. Hospital utilization (A37!=”None”)

Insurance Claims:

iv. Number of claims (B1)
v. Number of bed days (B2)
vi. Allowed amount for claims (B4)
vii. Amount paid by plan for claim (B5)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis B.3: Under **average** advantageous (adverse) selection, the difference in baseline productivity between participants and non-participants will be positive (negative). Under **marginal** advantageous (adverse) selection, the difference in baseline productivity between participants and non-participants will be positively (negatively) correlated with incentive size.

The following variables will be used to measure baseline productivity:

Baseline Survey:

i. Days missed (A45)

Human Resources Data:

i. Salary (C1)
ii. Absenteeism rate (C2)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis B.4: Under **marginal** advantageous (adverse) selection, average health measures, as measured by biometric screening results, will be negatively (positively) correlated with incentive size, conditional on participating in the biometric screening.

The following variables will be used to measure health conditional on biometric screening:

Biometric Screening Data:

i. BMI (D1)
ii. Waist circumference (D2)
iii. Resting blood pressure (D3)
iv. Total cholesterol (D4)
v. Total cholesterol ratio (D5)
vi. HDL cholesterol (D6)
vii. LDL cholesterol (D7)
viii. Triglycerides (D8)
ix. Glucose levels (D9)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis B.5: Under **marginal** advantageous (adverse) selection, average health measures, as measured by HRA screening results, will be negatively (positively) correlated with incentive size, conditional on participating in the HRA.

The following variables will be used to measure health conditional on HRA participation:

HRA Data:

i. Health risk score (E1)
ii. Nutrition risk score (E2)
iii. Healthy activities risk score (E4)
iv. Readiness to change score (E5)
v. Biometrics score (E6)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis Group C: *Participation in biometric screening, HRA, and wellness activities will result in increases in health status and productivity one year later. Participation will have an ambiguous effect on health care costs and utilization.*

Hypothesis C.1: Follow-up health status will be positively affected by wellness treatment, as measured by the reduced form relationship between follow-up health and treatment incentive sizes for biometric screening/HRA and/or wellness activity participation. The same will hold for the IV estimate of the effect of participation in biometric screening/HRA and/or wellness activity participation on follow-up health, instrumented for by incentive size.

The following variables will be used to measure follow-up health:

Follow-Up Survey:

i. Had at least one previous health screening (Any of A1-A5, A8-A9=”yes”)
ii. Physically active (A11=”More active”)
iii. Trying to be more active (A12=”Yes” or A13=”Yes”)
iv. Smoking status:
    1. Current smoker: A16=”Yes” and A17=”Every day” or “Some days”
    2. Former smoker: A16=”Yes” and A17”=”Not at all”
v. Other tobacco use (A22 and A23 != “Not at all”)
vi. Drinking:
    1. Drinker: A24!=0
    2. Heavy drinker: A25>=4 if female, A25>=5 if male
vii. Has at least one chronic health condition (A27)
viii. Self-reported health (A28)
    1. Health is excellent or very good
    2. Health is not poor
ix. Problems with physical activities or pain (A29-A31)
    1. A29=”Somewhat”, “Quite a lot”, “Could not do physical activities” or A30 = “Some”, “Quite a lot”, “Could not do daily work” or A31=”Mild”, ”Moderate”, “Severe”, “Very severe”
x. Energy (A32=”An extraordinary amount”, “Quite a lot”)
xi. Emotional health (A33=”Moderately”, ”Quite a lot”, “Extremely”)
xii. Overweight status (A39=”Overweight” or A40=”Very overweight”)
xiii. Bad health status (A40=”High”, ”Very high” or A41=”High” or “Very high” or A42=”High” or “Very high”)
xiv. Sedentary job (A53=”None at all”, ”Some, but less than 1 hour”)

Follow-Up Biometric Screening:

xv. BMI (H1)
xvi. Waist circumference (H2)
xvii. Resting blood pressure (H3)
xviii. Total cholesterol (H4)
xix. Total cholesterol ratio (H5)
xx. HDL cholesterol (H6)
xxi. LDL cholesterol (H7)
xxii. Triglycerides (H8)
xxiii. Glucose levels (H9)
xxiv. Deviation in expected biometrics from actual biometrics (D10)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis C.2: Follow-up productivity will be positively affected by wellness treatment, as measured by the reduced form relationship between follow-up productivity and treatment incentive sizes for biometric screening/HRA and/or wellness activity participation. The same will hold for the IV estimate of the effect of participation in biometric screening/HRA and/or wellness activity participation on follow-up productivity, instrumented for by incentive size.

The following variables will be used to measure follow-up productivity and job satisfaction:

Follow-Up Survey:

ii. Days missed (A45)
iii. Job satisfaction
    1. A46=”Very satisfied”
    2. A46=”Very satisfied”, ”Somewhat satisfied”

Human Resources Data:

iv. Salary (H1)
v. Absenteeism rate (H2)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

Hypothesis C.3: Follow-up health costs will be decreased by wellness treatment, as measured by the reduced form relationship follow-up health costs and between treatment incentive sizes for biometric screening/HRA and wellness activity participation. The same will hold for the IV estimate of the effect of participation in biometric screening/HRA and wellness activity participation, instrumented by incentive size.

The following variables will be used to measure follow-up health costs and utilization:

Follow-Up Survey:

i. Drug utilization (A34>0 or A35>0)
ii. Physician or ER utilization (A36!=”None”)
iii. Hospital utilization (A37!=”None”)

Insurance Claims:

iv. Number of claims (B1)

v. Number of bed days (B2)
vi. Allowed amount for claims (B4)
vii. Amount paid by plan for claim (B5)

Hypothesis Group D: *The share of one's peers who are induced to participate in the biometric screening/HRA and/or wellness programs increases one's own likelihood to participate in these activities.*

Hypothesis D.1: Participants in the suppressed peer treatment clusters (treatment group D) will have lower participation rates than analogous members of the standard treatment groups (treatment groups A, B, and C).

The following indicators will comprise the family of outcomes in this domain:

Participation Outcomes:

i. Scheduled a biometric screening (D10)
ii. Completed a biometric screening (D11)
iii. Completed HRA (E1)
iv. Enrolled in a fall wellness course (F1)
v. Completed fall wellness course (F2)
vi. Enrolled in a spring wellness course (F3)
vii. Complete spring wellness course (F4)

A standardized participation outcome impact will be obtained, as described in our methodology section below. In addition, we will report each outcome separately.

In addition, peer networks will be identified based on self-reported close friends at work, via the baseline survey:

Baseline Survey:

i. Close co-workers (A43)

Hypothesis Group E: The effect of treatment incentives on participation and short-run outcomes will vary based on baseline characteristics.

We will look for heterogeneity in participation with respect to the following:

i. Age (C3 ≥ 50)
ii. Sex (C4)
iii. Race (C5, white v. Non-white)
iv. Ethnicity (C6, Hispanic v. Non-Hispanic)
v. Baseline health status

1. All categories from baseline survey measuring health (these are outlined above in Hypothesis B.1)

vi. Annual Salary (C1, above median)
vii. Employment Class (C7)
viii. Deviation in expected biometrics from actual biometrics (D10)

5. Estimation Methodology

a. Treatment Effects

For our main participation outcomes, we will restrict analysis to members of treatment groups A, B, and C. First, we will pool groups B and C, and compare them to group A, as follows:

$$P_i = \alpha + \beta_{B,C} T_{i,B,C} + \gamma_{75} T_{i,75} + \Gamma X_i + \varepsilon_i$$

where P_i is the participation outcome, $T_{i,B,C}$ is an indicator for membership in treatment group B or C, $T_{i,75}$ is an indicator for receiving the $75 incentive for wellness program completion, and X_i is a vector of baseline control variables We will only include control variables that are good predictors of the outcome variable. We will identify those control variables by estimating a LASSO regression with five-fold cross validation. We will include in that LASSO regression all available variables from our baseline survey and HR data. Next, we will separately estimate the effect of treatment B and C, relative to treatment A, as follows:

$$P_i = \alpha + \beta_B T_{i,B} + \beta_C T_{i,C} + \gamma_{75} T_{i,75} + \Gamma X_i + \varepsilon_i$$

where now $T_{i,B}$ and $T_{i,C}$ are indicators for membership in treatment groups B and C, respectively.

Our next set of estimates will test for adverse or advantageous selection, again among members of treatment groups A, B, and C. We will estimate the following regressions:

$$X_i = \alpha_A T_{i,A} + \alpha_B T_{i,B} + \alpha_C T_{i,C} + \delta_A T_{i,A} \times P_i + \delta_B T_{i,B} \times P_i + \delta_C T_{i,C} \times P_i + \varepsilon_i$$

where X_i is baseline variable measured from either the baseline survey, administrative data, or health insurance data, and $T_{i,A}$ is an indicator for membership in treatment group A. To test for **average** advantageous or adverse selection, we will test the sign of the δ coefficients. To test for **marginal** advantageous or adverse selection, we will compare the magnitudes of the δ coefficients across treatment groups. In the case that a variable is top-coded, we will use a censored regression model to account for the structure of the limited dependent variable.

Our second set of selection estimates are estimated conditional on either completing the biometric screening or the HRA. In particular, we will estimate the following:

$$X_i = \alpha + \pi_B T_{i,B} + \pi_C T_{i,C} + \varepsilon_i$$

In this case, we can only test for **marginal** advantageous or adverse selection by comparing the magnitudes of the π coefficients. In the case that a variable is top-coded, we will use a censored regression model to account for the structure of the limited dependent variable.

In the above specifications, we will extend the analysis to test for selection with respect to the wellness program incentives, by adding an interaction term $T_{75} \times P_i$.

Our next set of treatment effects measure the short-run effect of wellness participation on health, health utilization, and productivity. We will compare members of the control group and members of treatment group C as follows:

$$Y_i = \alpha + \theta_C T_{i,C} + \theta_{75} T_{i,75} + \varepsilon_i$$

In addition, we will estimate local average treatment effects of HRA/biometric completion and wellness program completion:

$$Y_i = \alpha + \theta_{HRA/Screen} HRA/Screen_i + \theta_{Wellness} Wellness_i + \varepsilon_i$$

screening and HRA, and $Wellness_i$ is a count of the number of wellness programs completed. These two regressors will be instrumented for by treatment group membership (B or C versus A) and wellness program incentive.

In order to test for peer effects, we will compare participation outcomes among members of the clustered treatment group D to those of members of treatment groups A, B, and C. Analysis will follow the recommendations of Angrist (2014). Specific, we will estimate the following specification among members of treatment groups A, B, C, and D with comparable peer network sizes:

$$P_i = \alpha + \rho_D T_{i,D} + \varepsilon_i$$

where $T_{i,D}$ is an indicator for membership in treatment group D.

b. Heterogeneous Effects

In order to identify heterogeneous effects in participation, we will interact

baseline characteristics with our treatment effect estimates. We will estimate the following equation:

$$P_i = \alpha + \beta_{B,C} T_{i,B,C} + \gamma_{75} T_{i,75} + \lambda_{B,C} T_{i,B,C} \times X_i + \lambda_{75} T_{i,75} \times X_i + \Gamma X_i + \varepsilon_i$$

We will also test for heterogeneity under the more flexible specification:

$$\begin{aligned} P_i \quad = \quad & \alpha + \beta_B T_{i,B} + \beta_C T_{i,C} + \gamma_{75} T_{i,75} + \Gamma X_i \\ & + \lambda_B T_{i,B} \times X_i + \lambda_C T_{i,C} \times X_i + \lambda_{75} T_{i,75} \times X_i + \varepsilon_i \end{aligned}$$

Our estimates of heterogeneity in the case of short-run health and productivity effects will be obtained as follows in reduced form:

$$Y_i = \alpha + \theta_C T_{i,C} + \theta_{75} T_{i,75} + \phi_C T_{i,C} \times X_i + \phi_{75} T_{i,75} \times X_i + \varepsilon_i$$

and, additionally, as follows for IV estimates:

$$\begin{aligned} Y_i \quad = \quad & \alpha + \theta_{HRA/Screen} HRA/Screen_i + \theta_{Wellness} Wellness_i \\ & + \lambda_{HRA/Screen} HRA/Screen_i \times X_i + \lambda_{Wellness} Wellness_i \times X_i + \varepsilon_i \end{aligned}$$

One of our interaction terms, the deviation between expected biometrics screening results and actual screening results, is measured conditional on taking the biometric screening, and, furthermore, could be influenced by treatment assignment. We will first test for significant differences in this measure across treatment groups, and will only include it if the measure does not systematically vary with treatment group.

c. Multiple Outcomes, Standardized Treatment Effects, Multiple Inference

Many of our domains of interest, such as “baseline health”, can be measured by several different variables. We will address this by summarizing estimates within a domain using standardized treatment effects. We will also report estimates for each of the individual outcomes within a domain, in which case we will report both regular p-values and p-values adjusted to account for multiple outcomes. We will follow the methodology employed by Finkelstein et al. (2012) in all cases.

d. Missing Data and Questions with limited Variation

If the outcome variable is missing for a substantial fraction of the sample, we will not conduct regression analysis using that outcome. If that happens, we will note it in our write-up. Also, we will not report results if the outcome variable is identical for 90% or more of the sample. If a control variable is missing for a significant fraction of the population, we will omit it from our regressions. For

control variables that are missing for a minor share of respondents, we will use a dummy variable to indicate a missing value, and retain the control variable. Decisions to include or exclude control variables on this basis will be done prior to treatment assignment and observation of outcome variables.

e. Badly reported responses

The baseline survey asks respondents to report their age and their gender. These data are also available from our administrative HR dataset, so we will be able to assess whether or not respondents are responding truthfully to these two questions on the baseline survey. We will define a respondent to be "untruthful" if they report their age or their gender incorrectly. We will report as a robustness test results omitting those participants who answer these questions incorrectly. If omitting these individuals significantly affects our results, then we will remove them from our preferred specification.

f. Follow-Up Survey Attrition

Our follow-up analysis can only be conducted among employees that remain employed at the university and furthermore, among employees that agree to complete the follow-up survey and biometric screening. We will first examine If treatment assignment is related to attrition at these various levels at a statistically significant level. If not, then we will conduct analysis ignoring attrition. However, if we do find a significant difference in attrition, we will implement bounds on our treatment effect estimates, using Lee bounds in the case of differential employment turnover, and using sharper bounds to address differential survey and/or biometric screening response, which rely on variation in follow-up incentives to calculate the bounds.

6. Contingency Plans

a. Low Baseline Survey Response

If we receive less than 2,000 responses in the 7 days following the launch of the baseline survey, then we will increase the Amazon.com gift card amount from $30 to $50 to encourage a larger response rate. If that happens, we will also award an extra $20 to participants who already responded in the first 7 days.

b. Peer Effects Analysis

If the response rate to question A43 on the baseline survey is insufficiently high, then we will not implement the peer effects portion of the field experiment (treatment group D). Similarly, if we determine that we do not have enough power to rule out meaningful peer effects, then we will not implement the peer effects portion. If the peer effects treatment group is dropped, then we will use

the savings from that to increase the sample sizes of treatment groups A, B, and C. Determination of whether to carry out the peer effect analysis will be carried out prior to treatment assignment and observation of any outcome variables.

c. Low Baseline Screening Rates

If rates of participation in the biometric screening are much lower than projected, we may add another treatment arm, Group E, comprised of individuals from the control group. Group E will receive the same treatment as Groups A, B, and C, except that the screening incentive will be $400. The number of individuals invited to Group E will depend on budget and a projected response rate based on early participation in Groups A, B, and C. Determination of whether to add another treatment arm will be done prior to observation of any outcome variables from the screening rate portion of the experiment.