ILLINOIS
Gies College of Business

DOES WORKPLACE WELLNESS WORK?
THE ILLINOIS WORKPLACE WELLNESS STUDY

Authors: David Molitor and Julian Reif
Released: Fall 2022
Prepared for: FIN 550 Big Data Analytics

1 DOES WORKPLACE WELLNESS WORK?

Some of the main drivers of high health care spending in the US are related to smoking and chronic diseases such as cardiovascular disease, diabetes, high blood pressure, and obesity. Increasing the use of preventive care could potentially help people avoid these diseases, thereby lowering costs. How can this best be accomplished? One widely touted solution is increased use of workplace wellness programs, which have become popular among US employers that offer health benefits to their employees.

The workplace is an attractive venue to promote preventive care, since workers spend a lot of time there. Moreover, firms have a vested interest in finding ways to improve worker health. Healthier workers are less absent from work, are more productive, and have lower health care costs—something especially appealing to firms that provide health insurance benefits. Many employers seemingly agree with this logic: the workplace wellness industry has more than quadrupled since 2011, drawing in $8 billion in annual revenue and covering over 50 million American workers. Among large US firms offering health benefits in 2019, 84% also offered a wellness program.

While the specific characteristics of a workplace wellness program vary across employers, most programs include at least one of three main components. The first is an annual health screening, where a clinician visits the workplace and measures employee health outcomes such as blood pressure, cholesterol levels, and body weight. The second component is a health risk assessment, a questionnaire designed to assess lifestyle health habits. The third component is a set of wellness activities, such as self-paced walking programs or stress management classes, that employees can participate in throughout the year. Best practice guides advise employers to let employees take paid time to participate in wellness programs and to provide financial incentives for completing the various components, in order to maximize participation.

The widespread adoption of workplace wellness programs has been fueled in part by studies suggesting that these programs generate substantial medical and absenteeism cost savings. One widely cited summary of these studies estimates that medical spending falls by $3.27, and absenteeism costs by $2.73, for every dollar spent on wellness programs (Baicker, Cutler, and Song 2010). However, the accuracy of these estimates has been contested, with some critics arguing that these programs may not work as intended. Most prior studies of wellness programs estimated the savings generated by those programs by comparing people who chose to participate in the program to those who did not, or by comparing firms that implemented a wellness program to firms that did not.

These comparisons raise concerns about selection bias, which arises when individuals in a group differ systematically from the larger population. In the context of workplace wellness programs, one might suspect that the employees who are most eager to join a wellness program lead healthier lifestyles than the typical employee. Selection bias makes it difficult to draw conclusions from a comparison of wellness program participants to non-participants. Are the participants healthier because the wellness program improved their health, or were the participants simply healthier from the start? There are ways to adjust for some of these differences using regression analysis, but how can one adjust for factors such as intrinsic motivation or dislike of exercise?

2 THE VALUE OF RANDOMIZATION

A randomized controlled trial (RCT) is a research design that, if properly implemented, can overcome concerns about selection bias. In its simplest form, an RCT randomly assigns one group of individuals to a “treatment group,” which receives an intervention, and assigns the remaining individuals to a “control

group," which does not receive the intervention. After the intervention has concluded, the average effect of the intervention on an outcome, Y, can be calculated as the difference in means of that outcome between the treatment group and control group. More formally, let $\overline{Y_T}$ be the average of the outcome for the treatment group, and let $\overline{Y_C}$ be the average for the control group. Then the average treatment effect (ATE) of the intervention on the outcome is equal to:

$$ATE = \overline{Y_T} - \overline{Y_C}$$

Because assignment to the treatment group is random—individuals are prohibited from choosing whether to be in the treatment or control group—the only systematic difference between treatment and control is that members of the treatment group received the intervention and members of the control group did not. Thus, any significant differences that arise between the groups, as measured by the ATE, must have been caused by the intervention. In this way, an RCT allows researchers to estimate causal effects without suffering from selection bias.

3 THE ILLINOIS WORKPLACE WELLNESS STUDY

The Illinois Workplace Wellness Study implemented a randomized evaluation of a workplace wellness program at the University of Illinois at Urbana-Champaign (Jones, Molitor, and Reif 2019; Reif et al. 2020). The study included 4,834 employees, 3,300 of whom were randomly assigned to the treatment group. The remaining 1,534 employees were assigned to the control group.

All members of the treatment group were invited to take paid time off to participate in a comprehensive workplace wellness program called iThrive, which ran for two years and included an annual health screening, an annual health risk assessment, and weekly wellness activities. Treatment group members who successfully completed the program earned financial rewards of up to $650. To evaluate the effects of the program, employee data were collected from online surveys, university employment records, health insurance claims, campus gym visit records, and running event records.

Here are a few resources that offer additional background on the study:

- Study website, with a summary results and links to articles: https://nber.org/workplacewellness
- New York Times coverage: https://www.nytimes.com/2018/08/06/upshot/employer-wellness-programs-randomized-trials.html
- Public use data: https://github.com/reifjulian/illinois-wellness-data
    - Examples of how to load the data in R are included on the homepage README
    - Data codebooks describing all variables are found at /data/codebooks
    - More data documentation at /documentation

4 REFERENCES

Baicker, Katherine, David Cutler, and Zirui Song, "Workplace Wellness Programs Can Generate Savings," *Health Affairs*, 29 (2010), 304–311.

Jones, Damon, David Molitor, and Julian Reif. "What do workplace wellness programs do? Evidence from the Illinois workplace wellness study." *Quarterly Journal of Economics* 134, no. 4 (2019): 1747-1791.

Reif, Julian, David Chan, Damon Jones, Laura Payne, and David Molitor. "Effects of a workplace wellness program on employee health, health beliefs, and medical use: A randomized clinical trial." *JAMA Internal Medicine* 180, no. 7 (2020): 952-960.

5 CASE DESCRIPTION

1. Suppose we want to estimate the effect of workplace wellness programs on employee health care spending. One approach would be to compare health care spending at firms with wellness programs to firms without such programs. What condition would need to be true for this comparison to describe the causal effect of workplace wellness on health care spending? Is this condition likely to be satisfied? Why or why not?

2. Suppose we want to estimate the effect of workplace wellness among employees in the Illinois study. One approach is to compare outcomes for employees eligible to participate in the wellness program to ineligible employees. What condition would need to be true for this comparison to describe the causal effect of workplace wellness on employee outcomes? Is this condition likely to be satisfied? Why or why not?

3. What do "treatment" and "control" mean in the context of this study? How does this differ from the definition of a "participant" and "non-participant"? How many employees were in the treatment group, and how many in the control group? Of those in the treatment group, how many participated in the initial (screening) segment of the wellness program in the first year? The answer to this question can be found in the publications of the Illinois Workplace Wellness Study, but use the public-use data (described in Section 3) to form your answers.

4. Find the "claims" dataset in the public use data repository. For each outcome in the claims dataset, as measured pre-randomization (i.e., prior to August 2016), report the following in a four-column table (one row per outcome): **column (1)** Variable description; **column (2)** Control group mean; **column (3)** Treatment group mean; and **column (4)** p-value on the difference. Use linear regression to calculate all these values, and report and describe this equation in your answer below. Do the results in this table support or undermine the claim that treatment group assignment was random?

5. For each outcome in the "claims" dataset, as measured in the first year following randomization, report the following in a three-column table (one row per outcome): **column (1)** Variable description, **column (2)** estimated difference between *treatment and control groups* (no demographic controls) along with standard error in parentheses, **column (3)** estimated difference between treatment and control groups (with demographic controls) along with standard error in parentheses. Your demographic controls should include indicator variables for sex (male/female), race (white/nonwhite), middle age group (37-49/not 37-49), and oldest age group (50+/not 50+). Use linear regression to calculate all these values, and report and describe this equation in your answer below. What do these estimates indicate about causal treatment effects or selection bias? Should we expect the estimates to differ much between columns (2) and (3)?

6. For each outcome in the "claims" dataset, as measured in the first year following randomization, report the following in a three-column table (one row per outcome): **column (1)** Variable description, **column (2)** estimated difference between *participants and non-participants* (no demographic controls) along with standard error in parentheses, **column (3)** estimated difference between participants and non-participants (with demographic controls) along with standard error in parentheses. Your demographic controls should include indicator variables for sex (male/female), race (white/nonwhite), middle age group (37-49/not 37-49), and oldest age group (50+/not 50+). Use linear regression to calculate all these values, and report and describe this equation in your answer below. What do these estimates indicate about causal treatment effects or selection bias? Should we expect the estimates to differ much between columns (2) and (3)?

6 CASE DELIVERABLES

You should **deliver the following two files** as part of your final project:

1. **Executive Summary**. Produce an executive summary report in PDF format of your approach and findings. The document should have three sections: (I) a concise overview of the case and objectives, (II) answers to each questions asked in the Case Description above, and (III) a conclusion that draws lessons for employers who may be considering adopting a wellness program in their workplace. Your executive summary needs to be thoughtful, but it should not be more than 3 double-spaced pages, font 12, in length.

2. **Script**. Along with your executive summary, turn in a script that contains programming commands perform your analysis. It should be created such that if you were to give your script to someone else, they could run the script and generate exactly the same results. To help us to understand your code, annotate your code with brief comments and follow recommended programming style practices.

7 CASE GRADING

1. **Executive Summary (60%)**. Your grade on this component will be based primarily on your ability to clearly communicate your objectives, methodologies, and results. Avoiding jargon, correct grammar, and proper sentence and paragraph structure will all be considered.

2. **Script (40%)**. Your grade on this component will be based on two factors. First, how easy is it to follow your script? Clearly commenting your code and using informative naming conventions will help. Second, are your results replicable? If the raw data files are in the same folder as your script, we should be able to run your code without any modification and replicate the results in your final project.