Wellness Program Outreach and Effectiveness: A Randomized Controlled Trial

Analysis Plan

Date: October 1, 2024

Abstract

We will conduct a randomized controlled trial to evaluate the impact of an outreach campaign designed to increase engagement with Pack Health, a Quest Diagnostics wellness program providing individual health coaching for weight management and diabetes prevention. While employee wellness and disease-management programs have the potential to improve wellbeing and reduce healthcare costs, their effectiveness is often undermined by low engagement and selection bias in participant comparisons. This study will test whether an outreach approach that auto-enrolls eligible individuals—employees and their spouses/domestic partners—into the program, with the option to opt out, can increase engagement and improve health outcomes compared to the standard invitation-based approach.

Participants will be randomly assigned to one of two groups: the intervention group, which will receive auto-enrollment outreach, and the control group, which will receive traditional invitation-based outreach. The primary outcome will measure program engagement as the share who begin the program. Secondary outcomes will include additional measures of engagement, such as the number of modules completed, health outcomes observed in subsequent risk assessments and healthcare claims, and employee retention. This trial will provide evidence on whether an opt-out framing in outreach campaigns can enhance the effectiveness of wellness programs, ultimately informing best practices for population health management.

PRINCIPAL INVESTIGATOR: Joseph Doyle[1], PhD

CO-INVESTIGATORS: David Molitor[2], PhD; Guilherme Marques de Amorim[2], PhD; Nicholas Torsiello[3], PhD MPH

AFFILIATIONS:

1. Massachusetts Institute of Technology
2. University of Illinois Urbana-Champaign
3. Quest Diagnostics

1. Introduction

This document describes our analysis plan for: Wellness Program Outreach and Effectiveness: A Randomized Controlled Trial. This will serve as a record of the planned analyses to avoid concerns over searching for results *ex post*. We anticipate that further analyses will be conducted as more outcome data become available and as new ideas are inspired by the initial findings. It is also possible that our planned analyses may be modified based on what we learn during the initial launch phase. Any such additions will be noted in a later version of this plan with an indication of the date on which they were added.

The outline of this plan is as follows: Section 2 briefly describes the intervention and trial details. Section 3 describes the data and the outcomes of interest; Section 4 presents the main empirical models, including planned analysis for heterogeneous effects; and Section 5 concludes with caveats and interpretation issues.

2. Intervention

The program we are studying is Pack Health, a Quest Diagnostics company that offers individual health coaching to improve health (e.g., weight loss). The intervention consists of an outreach campaign that notifies eligible participants they have been auto-enrolled in the program. The campaign then encourages the scheduling of an initial appointment with the program. The control group receives similar campaign messages, but the communications invite eligible participants to join the program.

Enrollment

The trial will launch in October 2024. Enrollment in the trial will be conducted by Quest Diagnostics in a process like the one currently used to enroll eligible Quest Diagnostics members (employees and spouses/domestic partners) into Pack Health. Members will be enrolled if they satisfy the inclusion and exclusion criteria described below. Projecting from Pack Health eligibility rates among Quest members in 2023 and an estimated 40% informed consent rate, we

anticipate trial enrollment of about 2,000 members. Actual enrollment will depend on how many meet the following criteria:

Inclusion criteria:

1. Provided informed consent.
2. Those who are eligible for the Pack Health program and its outreach will be included in the study with the exceptions of those who fall under the exclusion criteria below.

   Pack Health eligibility criteria include:
   - Employees/spouses/domestic partners over the age of 18
   - Metabolic syndrome defined as results indicating 3 or more of the following risk factors:
     - High waist circumference (>35 inches for women and >40 inches for men)
     - High triglycerides (≥150 mg/dl)
     - Low HDL cholesterol (<50 mg/dl for women and <40 mg/dl for men)
     - High Blood Pressure (≥130/85)
     - High Fasting Glucose (≥100 mg/dl)

Exclusion criteria:

1. Anyone under the age of 21. They will not be part of the study and will receive messages as if they were in the control group.
2. Anyone observed in claims data with a historical diagnosis of:
   - Anorexia nervosa
   - Bulimia nervosa
   - Binge-eating disorder
   - Body dysmorphic disorder
   - Major depression
   - Post-traumatic stress disorder (PTSD)
   - Severe anxiety disorder

Randomization

The randomization will take place through the data informatics teams at Quest Diagnostics and Pack Health in collaboration with MIT. MIT will send a database to Pack Health that includes the randomized flags for inclusion in the treatment (auto-enrollment) and control groups within randomization strata: HbA1c > 2023 Median, Weight > 2023 Median, and Female. Pack Health will receive a list of eligible participants from Quest Diagnostics and slot them into the open rows in the database. Those in the treatment group will then receive the auto-enrollment outreach campaign, while those in the control group will receive the standard outreach campaign. The outreach campaign is implemented by Pack Health.

3. Data sources and outcomes

3.1. Data sources

Data are provided by Quest Diagnostics and Pack Health. These include annual risk assessments (lab and survey results), healthcare claims, wellness program participation, and employment and job-related information to measure relational and environmental factors (e.g. salary class, job class, leave of absence information, site code to match co-workers, and tenure to measure effects of experience), and demographics (age, race, ethnicity, sex). We will receive data on members enrolled in the trial.

3.2. Primary outcome

Pack Health program utilization is the primary endpoint measured by the share who begin the program.

3.3. Secondary outcomes

In addition to the primary engagement outcome, we will measure complementary measures of engagement, including the number of modules completed and reaching the threshold for program completion (9 modules).

Secondary outcomes will also include annual risk assessment measures, including weight and HbA1c. We will also examine healthcare claims and categories of claims including inpatient, emergency, and outpatient utilization. Retention in the sample will be examined as well, both because it will be important to understand the sample selection due to differential retention to interpret the other outcomes, and, among employees, retention is of independent interest as it could reflect employee satisfaction and lower costs due to employee turnover.

4. Empirical Model

Our initial analysis will estimate intent-to-treat models using linear regression. Consider an outcome, Y_i, such as an indicator for Pack Health enrollment, for each subject i. The primary specification is:

$$Y_i = \beta_0 + \beta_1 Treat_i + \beta_2 X_i + \epsilon_i \quad (1)$$

For our analysis, $Treat_i$ is an indicator equal to one if the subject was randomized to the treatment group and zero if the subject was randomized to the control group. X_i is a vector of control variables. These control variables should be uncorrelated with the treatment indicator, but they can aid in the precision of the estimate because they are correlated with the outcome Y_i. First, we will include strata indicators. We will also include the lag of the dependent variable where appropriate, which substantially improves power for clinical outcomes. For test results that would be the most recent test result prior to enrollment in the study. For utilization outcomes such as inpatient visits, we will include the number of such episodes observed in the 12 months prior to study enrollment. For individuals with incomplete data in the prior 12 months, we will include an indicator for that status.

We also plan to include standard demographic controls as available, including age, race, ethnicity, and employee status (recall that female is included as a strata control). 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.

We will also estimate models where the outcome is the difference between the baseline and the follow-up exam instead of including the baseline exam result as a control:

$$\Delta Y_i = \beta_0 + \beta_1 Treat_i + \beta_2 X_i + \epsilon_i. \quad (2)$$

This is a more restrictive model but may improve statistical precision and requires fewer controls. For those with missing lagged values in this difference model, we will estimate models for those with non-missing values. We will also test whether missing lagged value is balanced across treatment and control for each measure. For statistical inference, we will calculate heteroscedasticity-robust standard errors.

To consider robustness, we will report models with only strata indicators as controls, strata indicators plus the lag of the dependent variable because this can substantially improve statistical power, and models with full controls. For models of binary outcomes, we will also compare marginal impacts using a logit specification.

The coefficient β_1 from equation (1) measures the effect of intention-to-treat: the causal effect of being randomized into the treatment group. This can differ from the effect of the program if there is non-compliance: some subjects randomized to the treatment group may decide not to join the Pack Health program or may quit it after a short time. As an alternative specification, we will estimate the relationship between Pack Health engagement and health and retention outcomes using Two Stage Least Squares (2SLS), with the initial randomized assignment to treatment or control used as an instrumental variable for Pack Health engagement. Specifically, we will estimate the following equation via 2SLS:

$$Y_i = \gamma_0 + \gamma_1 InitialEngagement_i + \gamma_2 X_i + \epsilon_i, \quad (3)$$

where $InitialEngagement_i$ is an indicator for completing an initial appointment. The 2SLS estimate of γ_1 measures the average causal effect of the program among those who engage due to the randomly assigned treatment status as a complementary parameter to the intent-to-treat analysis described above. In secondary analyses, we will explore versions of equation (3) in

which initial engagement is replaced by other measures of Pack Health engagement, such as the number of sessions attended and reaching the threshold for program completion.

In addition, if there is differential attrition from the study across treatment and control groups, we will explore bounding estimates under a range of assumptions about effect sizes for those not observed.

Planned Heterogeneity Analyses

In addition to outcome comparisons across all study participants, we will consider heterogeneous treatment effects across participants. We will test for different effects of the program by randomization strata, age, race, ethnicity, employee (vs. spouse/domestic partner), and terciles of baseline weight and HbA1c.

We plan to consider machine-learning techniques to choose the optimal controls in the regression analyses and to shed light on the types of members where the effects are particularly large. This will likely entail normalizing the outcomes so that we measure effects relative to each participant's baseline rather than absolute differences.

Statistical Power

For our primary engagement measure of the increase in take-up, for a control group mean of 5%, we have 80% power to detect a 3.7 percentage point increase. This is well within the range of the literature for opt-out interventions (Chapman, 2010).

Using the mean of 10 modules completed and standard deviation of 5.7 (conditional on starting), we can also calculate the MDE for number of modules to be 0.4. We expect the control group to complete 0.5 modules on average, so we could detect an increase to 0.9 for the treatment group. Again, this is well powered given that a 10ppt increase in take up should result in an average number of modules completed to be 1.5 for the treatment group.

For health outcomes, Table 1 shows that with the approximate anticipated sample size of 2000, we would have 80% power to detect a 0.035 point reduction in HbA1c and a loss of 1.88 pounds

comparing the treatment and control groups. If the outreach increases engagement with the program by 10 percentage points (e.g. from 5% take-up to 15% take-up),[1] then the implied reduction in HbA1c due to participation in the program would be a 0.35-point reduction; similarly, the change in weight due to participation would be 18.8 pounds.

If the increase in take-up is double (an increase of 20 percentage points), the resulting implied effects of the program would fall in half.

In the end, we are powered to detect small changes in take up, but we are only powered to detect relatively large, but not impossible, improvements based on the program.

For future research, a sample size of 5000 would provide power to detect an implied reduction in HbA1c due to participation in the program of 0.2 points, and a decline in weight approximately 12 pounds.

[1] In Chapman et al. (2010), a similar opt-out randomized trial for flu vaccination resulted in 12 percentage-point increase in take up.

Table 1: Statistical Power for Clinical and Claims-Based Outcomes

				Minimum Detectable Effect		Program Effect	
Measure	Mean	SD	R-squared	Unadjusted	Adjusted	Unadjusted	Adjusted
BMI	32.5	6.51	0.846	0.816	0.320	8.160	3.200
Total cholesterol	190	38.1	0.541	4.776	3.236	47.760	32.360
Diastolic blood pressure	79.7	10.1	0.236	1.266	1.107	12.660	11.070
Fasting blood glucose	99.3	13.3	0.275	1.667	1.420	16.670	14.200
HDL	48.8	13.9	0.706	1.742	0.945	17.420	9.450
Health quotient score	72.8	8.44	0.499	1.058	0.749	10.580	7.490
HbA1c	5.52	0.45	0.614	0.056	0.035	0.560	0.350
LDL	115	32.3	0.536	4.049	2.758	40.490	27.580
Systolic blood pressure	126	14.7	0.274	1.843	1.570	18.430	15.700
Triglycerides	147	89.3	0.335	11.194	9.128	111.940	91.280
Weight (lb)	205	46.3	0.895	5.804	1.881	58.040	18.810
Waist circumference (in)	39.4	5.85	0.690	0.733	0.408	7.330	4.080
Take-up	0.05	0.22	—	0.037	—	—	—
Modules completed	0.5	2.5	—	0.433	—	—	—

Notes: Means and standard deviations come from measure values from the 2018 annual risk assessment, Blueprint for Wellness (BFW). R-squared refers to the R-squared between the indicated outcome in 2018 vs. 2019. Minimum detectable effects are derived using a sample size of 2000 and the means and standard deviations from the 2018 BFW. Adjusted MDEs adjust SDs for reductions in variance due to controlling for the lag/previous year's value of the dependent variable. Actual program effects represent the effect the program would have to have on those who actually participate to attain the indicated minimum detectable effect, assuming that the default results in a 10 percentage point increase in uptake of the Pack Health program. Take-up and modules completed assume a take-up rate for the treatment group of 15%.

References

Chapman GB, Li M, Colby H, Yoon H. Opting In vs Opting Out of Influenza Vaccination. *JAMA*. 2010;304(1):43–44. doi:10.1001/jama.2010.892