Fresh Food Farmacy: A Randomized Controlled Trial

Analysis Plan

Date: April 10, 2019

Abstract

We will conduct a randomized controlled trial to estimate the impact of Geisinger Health's Fresh Food Farmacy program (FFF). The program brings a "food-as-medicine" approach to address a key social determinant of health (food insecurity) for patients with diabetes. A Fresh Food Farmacy dietitian prescribes enough fresh food for two meals per day, five days per week, for everyone in the patient's household. Prescriptions for the food are filled once a week at no charge at a local clinic. Complementary services include dietitian consultation, recipes, cooking classes, biometric tracking, diabetes self-management training and checks that participants are following best practices such as eye and foot exams. By randomly assigning patients to either start the FFF program immediately or in six months, we will be able to estimate the impact of the FFF program on a wide range of clinical outcomes, healthy behaviors, measures of wellbeing, and health care utilization.

PRINCIPAL INVESTIGATORS: John Bulger[1], DO MBA and Joseph Doyle[2], PhD

CO-INVESTIGATOR(S): Marcella Alsan[3], MD, PhD, John Cawley[4], PhD, Allison Hess[1], Michele Passaretti[5], RN

AFFILIATIONS:

1. Geisinger Health Plan
2. Massachusetts Institute of Technology
3. Stanford University
4. Cornell University
5. Geisinger Health

1. Introduction

This document details our analysis plan for: Fresh Food Farmacy: 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. The intervention

In July 2016, Geisinger began piloting the Fresh Food Farmacy (FFF) program for food-insecure patients with diabetes in Shamokin, PA. Fresh Food Farmacy brings a "food-as-medicine" approach to treat these patients. Participants receive a referral to the Fresh Food Farmacy dietitian who provides a prescription for fresh food for the patient and everyone else in their household, enough for 2 meals per day, 5 days per week. (They do not provide 3 meals per day, 7 days per week because it is assumed that the households have other sources of meals, and the program wishes to minimize waste.) The program lasts indefinitely; there is no pre-determined end date.

FFF participants receive education on how to use the food. This includes dietitian consultations, new recipes each week, and cooking classes. Next, the program expects, but does not compel, participants to attend diabetes self-management classes from a prominent program developed at Stanford University. Clinic staff include nurses who check whether

patients have received preventive care, such as eye and foot exams and influenza vaccination. Participants' biometric measures (HbA1c, fasting glucose, weight, BMI, blood pressure, total cholesterol, LDL cholesterol, and triglycerides) and medication usage are tracked on a regular basis. On a case-by-case basis, participants may receive cooking equipment to help in food preparation.

FFF has received considerable attention (e.g. Aubrey, 2017; Davidson 2017; Feinberg et al., 2017 and 2018), as it uses a holistic approach to address a key Social Determinant of Health (food insecurity), and because providers, public health officials, and payers are increasingly willing to consider such innovative approaches to improve the health of patients. Indeed, over 100 healthcare systems have consulted with Geisinger about their operation and are considering adopting it. Geisinger recognizes that a key next step is to allow for a randomized-controlled trial in order to accurately measure the effectiveness of the FFF program.

Geisinger Health piloted the FFF in Shamokin, PA, and plans to expand the program in the coming months, first in Scranton, PA and then in Lewistown, PA. The program faces capacity constraints such that the program cannot be offered to all eligible participants on the first day. We will carry out a pragmatic randomized controlled trial; eligible patients who are randomized to the treatment group will begin the FFF program immediately, while those randomized to the control group will begin the FFF program 6 months later. This will allow a rigorous examination of the effects of the program for outcomes 6 months after randomization (when the treatment group has been in the FFF program for 6 months and the control group has not yet started the FFF program).

Enrollment & Randomization

Enrollment in the trial will be conducted by Geisinger in a process similar to the one currently used to enroll patients into FFF.

Patients must satisfy the criteria based on the records at the time of recruitment noted in the IRB protocol. In particular, they must be adults with Type II Diabetes and an HbA1c $\geq$ 8.0 (as determined in Geisinger Health EMR data), food insecure (as determined from a two-question

survey instrument), and whose residence is within 25 miles or 45 minute drive from the FFF clinic (a distance that could change in the future), as measured by the patient's ZIP code centroid.

Geisinger Health patients who are *potentially* eligible will be randomized to determine the order in which they are recruited. (To clarify, this is not the randomization for assignment into the treatment and control groups, it is a randomization to determine the order in which potentially eligible patients are contacted and then later randomized to the treatment and control groups.) Depending on the speed of recruitment, we will consider comparing patients enrolled earlier vs. later to consider the evolution of A1c over a longer time horizon than 6 months (e.g. someone recruited 3 months into the trial and assigned to control will have HbA1c measured at 6+3=9 months after "potential enrollment").

Recruitment will be conducted over the phone; the program will be described, consent will be sought, and then patients will be randomized to either treatment ("start FFF now") or control ("start FFF later"). Note that all patients eligible for FFF, whether or not they agree to participate in the study, will be randomized to determine the timing with which they will enroll in FFF. This is the way the program is being introduced, akin to a lottery, and ensures that there is no incentive to refuse to participate in the study in order to start FFF sooner. Recruitment staff will discover whether the patient is randomized to start now or start later in a tamper proof REDCap application, which consults a randomized allocation table that the recruiter cannot access and a treatment status that is locked along with the patient's identifying information and study ID. The recruiter will then set up the patient's first appointment for those in the treatment group and will reach out to patients 5 months later to set up appointments for the group randomized to start later. We anticipate finding the patients 5 months later because all subjects are Geisinger Health patients, whose current address is on file, and we will ask for two other contacts at the time of recruitment.

The randomization is stratified using levels of HbA1c: <8.0 (in the event that the threshold is lowered), ≥ 8.0 and <9.5, and ≥ 9.5 to attempt to ensure balance on a key pre-study control and because we are interested in estimating effects separately by these categories to inform

threshold setting in future applications. The randomization happens separately by location (Scranton and Lewistown) as well.

The trial will launch in April 2019 in Scranton, and likely June 2019 in Lewistown. We plan to continue enrollment until we attain at least 500 subjects (half in treatment) or when the recruitment period has lasted 2 years. If recruitment slows, we will consider lowering the eligibility threshold to an HbA1c of at least 7.5 rather than 8.0, and we may consider expanding the geographic limit.

3. Data sources and outcomes

3.1. Data sources

There are three main sources of data:

A. Geisinger Health Electronic Health records. These include clinical outcomes that are collected as part of the study. Subjects go to a Geisinger lab for a blood test at no cost (more on potential attrition is discussed below); the blood is tested for HbA1c, fasting glucose, total cholesterol, LDL cholesterol, and triglycerides. Weight, BMI and blood pressure are also recorded. The records also include inpatient, outpatient, and ED encounters, and scheduling information. All subjects are Geisinger Health patients. Utilization is largely at Geisinger facilities, although inpatient stays at other hospitals may be visible in the EHR record.

B. Fresh Food Farmacy operational data. This includes encounter data at FFF, including number of weeks in which the patient picked up food, their choices of food, dietitian appointments, and classes. Program participants are surveyed upon entry into FFF, and again at 6 months and at 12 months, and the surveys have been modified to measure outcomes of interest. This will allow us to ask questions about diet, healthy behaviors, self-efficacy, and self-assessed health.

C. Geisinger Health Plan data. For subjects who receive their insurance from the Geisinger Health Plan, expected to be more than half of the subjects, paid claims are available. This

allows us to measure utilization at providers other than Geisinger Health, albeit on a smaller and selected sample. Highmark (another local insurer) may share utilization data for their patients that are participating in the program as well.

3.2. Primary outcome

The primary outcome we will analyze is the subject's HbA1c at 6 months into the trial. This is a standard biomarker for diabetes that is influenced by diet and medication adherence. It is a measure of whether the program is improving health; and it is strongly associated with healthcare utilization (e.g. Oglesby et al., 2006, Juarez et al., 2013).

3.3. Secondary outcomes

6-month outcomes

Secondary outcomes of interest include other biomarkers at 6 months: fasting glucose, weight, BMI, blood pressure, total cholesterol, LDL cholesterol, and triglycerides.

We will also examine outcomes based on self-reported survey data, including diet, exercise, and other healthy behaviors, self-efficacy, self-assessed physical and mental health, and patient satisfaction with Geisinger Health as a whole. We will also measure whether the patients have improved their understanding of diabetes and healthy eating.

Other outcomes help us understand the nature of the services offered. The control group has access to dietitian consultations and diabetes self-management classes as part of the standard of care, but we believe the take-up of such services will be higher among the treatment group. We ask about these consultations and classes in the patient surveys to verify that this is the case. We also survey participants about adherence to diabetes best preventive care and whether the subjects are signed up for SNAP benefits, as the program encourages take-up.

Using encounter data, we will also document the level of engagement with the program at the FFF clinic, including the number of weeks that the participants picked up food, the number of dietitian encounters, and the number of classes taken.

Health care utilization is another important secondary outcome. We will examine, at six months, the difference between the treatment and control group in: probability of any inpatient stay, any ED visit, any hospital visit (inpatient or ED). We will also examine the number of such visits, as well as the number of outpatient visits outside the FFF program. We will also examine medication usage, including insulin. Related, we will test for differences in paid claims both in those categories and in the amount paid among the Geisinger Health Plan beneficiaries and check that treatment and control are balanced on inclusion in this sample. We expect there may be outliers in payments, so we will estimate models in levels, logs, and check for differences in the distribution including the median. Any effect of the program on such outcomes may take time to materialize and thus may not be visible at 6 months.

12-month outcomes

We will examine the same outcomes at 12 months as we did at 6 months. In doing so, we will be comparing the treatment group (which will have had 12 months of exposure to the FFF program) to the control group (which will have had 6 months exposure to the FFF program). This will be informative about whether the rate of expected improvement in outcomes slows over time.

Health care utilization will also be examined at 12 months. As mentioned earlier, we expect that any effect of FFF on utilization may be visible only with a lag. We have statistical power for the more common 12-month outcomes: any inpatient stay, any ED visit, and any hospital visit (inpatient or ED). We will also examine the number of such visits, as well as the number of outpatient visits outside the Fresh Food Farmacy.

To allow time for the program to affect outcomes, we will also consider outcomes within other timeframes: 90-365 days and 90-180 days.

The program re-assesses food insecurity after one year, so food insecurity will be asked at 12 months; given the nature of the program we hypothesize that the FFF program should considerably reduce food insecurity.

Diet and Healthy behaviors

The 6- and 12-month surveys also ask about diet, exercise, and smoking behavior. In addition, whether routine examinations such as eye exams have been conducted are asked in the survey—measures that we can verify in the electronic health record for exams conducted at Geisinger. This will provide some information about utilization outside of Geisinger and serve as a check on the accuracy of the survey responses.

To investigate whether the Fresh Food Farmacy is a complement to other healthy behaviors, we will investigate changes in behaviors that we can observe in addition to the survey questions. These include the rate at which the subjects make other appointments at Geisinger. We also plan to investigate medication adherence. As a separate, related outcome, we will look at drug utilization for drugs used to treat diabetes to test whether the Fresh Food Farmacy can substitute for this treatment.

In addition, the EHR data will allow us to determine whether a subject completes wellness visits in terms of PCP and endocrinologist visits along with diabetes Healthcare Effectiveness Data and Information Set (HEDIS) measures, as another set of healthy behaviors that may be improved as a result of program participation.

Other Household members

Recall that FFF provides fresh food for all members of the participant's household. We will test whether the FFF program has spillover benefits to these other household members. Using address information, we will examine such household members that appear in the Geisinger EHR data and test for impacts on lab results and healthcare utilization.

4. Empirical Model

Our initial analysis will estimate intent-to-treat models using linear regression. Consider an outcome, Y_i, such as HbA1c, for each subject i. The primary specification is:

$$Y_i = \beta_0 + \beta_1 1(\text{Treatment})_i + \beta_2 X_i + \varepsilon_i$$

For our analysis, $1(\text{Treatment})_i$ 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. First, we will include strata indicators for location and HbA1c category. We will also include the lag of the dependent variable, which substantially improves power: for clinical 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.

We also plan to include standard demographic controls: age (5-year age bins), sex (indicator that the patient is male), race/ethnicity (indicators for African American and Hispanic). For the clinical test results, we will also include a control for the number of days from the baseline measure to the follow-up test in order to improve precision. If this number of days is not balanced across treatment and control, we will control for the number of days between the baseline test and the date of enrollment into the study. We will also include the month-year of entry into the study unless this measure is highly correlated with the number of days between blood tests and reduces statistical precision.

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:

$$\Delta Y_i = \beta_0 + \beta_1 1(\text{Treatment})_i + \beta_2 X_i + \varepsilon_i$$

This is a more restrictive model but may improve statistical precision and requires fewer controls. For models that control for lagged values or 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. Note that all patients have a baseline HbA1c that was used in recruitment.

To consider robustness, we will report models with only strata indicators as controls, strata indicators plus the lag of the dependent variable because this substantially improves statistical power, and models with full controls.

The coefficient β_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 FFF program or may quit it after a short time. (Given the value of the free food, we expect that refusal and dropout rates will be low.) At the initial pilot site of Shamokin PA, 17% of those who began the program were no longer in the program at 6 months.

As an alternative specification, we will estimate the equation using Two Stage Least Squares (2SLS), with the initial randomized assignment to treatment or control used as an instrumental variable for the fraction of months engaged in FFF. The 2SLS estimate measures the average causal effect of the program for those who choose to remain in it as a complementary parameter to the intent-to-treat analysis described above.

That is, for subject i the structural equation is:

$$\text{HbA1c at 6 months}_i = \beta_0 + \beta_1 \left\{ \left(\frac{1}{6}\right) * \text{Months of Engagement} \right\}_i + \beta_2 X_i + \varepsilon_i$$

And the first-stage is:

$$\text{Months of Engagement}_i = \alpha_0 + \alpha_1 Treatment_i + \alpha_2 X_i + \vartheta_i$$

We re-scale the months of engagement in the structural equation by 6 for 6-month outcomes, and by 12 for 12 month outcomes so that the interpretation of our estimate of β_1 is for full engagement, which eases comparisons to the intent-to-treat estimates.

In addition, if there is differential attrition from the study across treatment and control groups (despite our efforts to obtain outcomes from each subject including obtaining multiple forms of contact information for subjects and their family members and compensation for subjects' time)

we will explore bounding estimates under a range of assumptions about effect sizes for those not observed.

For survey responses that include categorical answers, outcomes will be binary measures of whether the answers are in the top and bottom quintiles as separate outcomes and will test robustness to using nonlinear estimators such as the multinomial logit. To avoid concerns about multiple hypothesis testing for multiple measures within each category of questions (diet, other healthy behaviors, diet understanding, diabetes understanding, self-assessed health, self-efficacy, and patient satisfaction), we will compile answers to these questions into indexes and test effects of the program on a change in these indexes. We will also report p-values corrected for the multiple tests when presenting results for each question individually.

Planned Heterogeneity Analysis

In addition to outcome comparisons across all of the participants, we will consider heterogeneous treatment effects across patients. First, *a priori* we are interested in whether the program is especially effective for different levels of baseline HbA1c in order to help target the program in the future. For example, if effect sizes decrease with the baseline A1c, we may want to expand the program to those with A1c $\geq$ 7.5.

We are also interested in whether the program has different effects across the two locations (Scranton and Lewistown) to investigate the external validity of the estimates.

We will also test for different effects of the program by sex, as the cooking instruction may differ in effectiveness for men versus women. Household structure may be important given that the food goes to the household: differences between single and multi-person households will be estimated.

We hypothesize that the FFF program may have greater impacts among those of lower income (for whom the food may be especially helpful) and may vary across education levels (those with more education may be better able to utilize the information they are given about food preparation and diabetes management or they may already possess that information); these hypotheses will be tested using ZIP code level education and income measures, although there

may be little variation. We will also test differences by type of payer, with Medicaid providing another proxy for income as well as independent interest in effects for this group given interest among Medicaid systems in this type of intervention.

To further investigate the importance of information, we can compare results for individuals who score relatively well on the baseline survey questions about understanding of diabetes and diet. We will create a score for those questions and split the sample based on the median value of the initial knowledge. Another source of heterogeneity that could be considered in other contexts is the cost of using the program. We plan to compare results by distance between home and the clinic based on the median distance to the clinic from the residential ZIP code.

HbA1c can be affected by diet, lifestyle, and drug adherence. To consider the latter, we will compare patients who are on insulin versus those who are not on insulin at the start of the trial.

We plan to consider whether machine-learning techniques to choose the optimal controls in X and to shed light on the types of patients where the effects are particularly large are feasible with the given sample size and statistical power. This will likely entail normalizing the outcomes so that we measure effects relative to each participant's baseline rather than absolute differences.

Interim results from the trial will be monitored on an ongoing basis by the MIT/J-PAL researchers (Drs. Alsan, Cawley and Doyle), while the clinician team will be blinded from these interim results so that such results do not affect program management in a way that would not be feasible outside of the study setting. Such blinding adds credibility to the results so that they are more likely to apply to other programs in the future that are not part of a study.

Statistical Power and Sample Size Considerations

Using data for current (i.e. self-selected, non-randomized) FFF participants, we conducted parametric power calculations. These take into account the increase in power from controlling for baseline characteristics, but qualitatively similar results were found when we did not include

controls and when we used nonparametric (simulation) methods to estimate power. Similar results were also found when we based the calculations on similar patients in the counties where the two new sites (Lewistown and Scranton) will be located. All power calculations were conducted using STATA Version 15.

The table below shows that with the expected sample size of 500 we will have substantial power to detect small minimal detectable effects (MDE), as well as the ability to conduct heterogeneity analyses for the clinical outcomes. For utilization outcomes, we will have power to detect the (albeit) large reductions in utilization found in the observational literature.

Table 2: Statistical Power to Detect Targeted Minimal Detectable Effects (MDE)

Variable	Mean	Std.Dev.	$\sqrt{1-R^2}$	Target MDE	Required N for MDE
HbA1c after 180 days	8.58	1.74	0.66	10%	55
HbA1c after 365 days	8.56	1.67	0.66	10%	56
Weight after 180 days	228	64	0.96	10%	228
Weight after 365 days	228	64	0.96	10%	228
Hospital Visit (Inpatient or ED) 0-180 days	0.39	0.49	0.85	30%	397
Hospital Visit (Inpatient or ED) 0-365 days	0.48	0.50	0.87	30%	278
Hospital Visit (Inpatient or ED) 90-365 days	0.40	0.49	0.92	30%	447

We subsequently simulated statistical power for a model that estimates the effects of full engagement. For our primary outcome of HbA1c at 6 months, with power=80% and alpha=0.05 and planned sample size of 500, we can detect:

- Intent to treat effect of assignment to treatment: reduction of 3.7%
- IV treatment effect of full engagement: reduction of 5.2%

Section 5: Conclusion

If the Fresh Food Farmacy is found to improve outcomes, a next phase of research could unpack which components of the program are most effective. For example, particular parameters of such programs could be explored such as the amount of food, how food is obtained (delivery or at a clinic), the importance of diabetes education with the food, and how long the patient is offered the program. The length of the program could shed light on the nature of habit formation and preference formation among patients with diabetes.

The proposed RCT will provide important information about the effects of the novel FFF program that has received considerable attention to inform future take up. It is also an example of a way to introduce interventions aimed at addressing Social Determinants of Health in a rigorous way to provide evidence for the design and scaling of future interventions.

References

Aubrey, Allison. "Fresh Food By Prescription: This Health Care Firm Is Trimming Costs — And Waistlines" NPR Morning Edition. May 8, 2017.

Davidson, Adam. "A Bipartisan Way to Improve Medical Care" The New Yorker. May 29, 2017

Feinberg, Andrea T., Jonathan R. Slotkin, Allison Hess, and Alistair R. Erskine. "Geisinger Treats Diabetes by Giving Away Free, Healthy Food." Harvard Business Review. December 19, 2017.

Feinberg, Andrea T., Allison Hess, Michelle Passaretti, Stacy Coolbaugh, Thomas H. Lee. "Prescribing Food as a Specialty Drug," NEJM Catalyst. April 10, 2018.

Juarez et al. "Association Between Sustained Glycated Hemoglobin Control and Healthcare Costs." Am J Pharm Benefits; 2013 ; 5(2): 59–64.

Oglesby AK, Secnik K, Barron J, Al-Zakwani I, Lage MJ. The association between diabetes related medical costs and glycemic control: a retrospective analysis. Cost Eff Resour Alloc. 2006; 4:1.