STUDY ANALYSIS PLAN

Comparative Effectiveness of Family vs. Individually Focused Diabetes Education and Support

Family Support for Health Action (FAM-ACT)

Rosland AM, Heisler M, Kieffer EC, Piatt GA, Valbuena F, Yabes, JG

University of Pittsburgh sIRB: STUDY20110344

June 22, 2022

Principal Investigator	**Ann-Marie Rosland, MD MS** Associate Professor of Medicine Research Scientist, VA Center for Health Equity Research and Promotion Director, Caring for Complex Chronic Conditions Research Center University of Pittsburgh Pittsburgh, PA roslandam@pitt.edu
Protocol identification number	v2.2.1 062222
ClinicalTrials.gov identifier	NCT03812614
Author(s)	**Jonathan G. Yabes, PhD** Assistant Professor of Medicine, Biostatistics, and Clinical and Translational Science University of Pittsburgh 200 Meyran Avenue, Suite 300 Pittsburgh, PA 15213 (412) 605-0158 jgy2@pitt.edu

CONTENTS

ABBREVIATIONS

CACE – Complier Average Causal Effect
DSME – Diabetes Self-Management Education
FAM-ACT – Family Support for Health Action
HbA1c – Hemoglobin A1c
DSMES – Individual Patient-Focused Diabetes Self-Management Education and Support
GLMM – Generalized Linear Mixed Model
ICE – Intercurrent Events
ICC – Intra-Class Correlation
ITT – Intent-To-Treat
MCAR – Missing Completely at Random
MAR – Missing at Random
MNAR – Missing Not at Random
RCT – Randomized Controlled Trial
RMPW – Ratio-of-Mediator-Probability
SAP – Statistical Analysis Plan
SBP – Systolic Blood Pressure
SP – Support Person
UKPDS – UK Prospective Diabetes Study

1.0 INTRODUCTION

This statistical analysis plan (SAP) describes the statistical methods to be used for analyzing and reporting data collected for the Family Support for Health Action (FAM-ACT), a randomized, parallel-arm, comparative effectiveness trial.

2.0 STUDY OBJECTIVES

FAM-ACT aims to compare the effectiveness of a family supporter-focused (FAM-ACT) intervention relative to an individual patient-focused diabetes self-management education and support (I-DSMES). Specifically:

1. Determine the effectiveness of FAM-ACT, compared to I-DSMES, in improving patients' diabetes-related health outcomes (primary: Hemoglobin A1c [HbA1c]; secondary: blood pressure)
2. Determine the effectiveness of FAM-ACT, compared to I-DSMES, in improving patients' health behaviors and perceived support
3. Examine theoretical-model-driven moderators and mediators of FAM-ACT participation on patient outcomes

3.0 STUDY DESIGN

This is a randomized, parallel-arm, comparative effectiveness trial designed to test effectiveness of FAM-ACT in improving diabetes health outcomes and self-management behaviors and perceptions among patients with type 2 diabetes. The study arms consist of:

- Treatment arm: FAM-ACT
    - Patient and Support Person/SP (dyad) will receive a Diabetes Complications Risk Assessment profile and introduction session, SP-focused information/skills training through 4-6 extended DSME sessions, case management contacts with a Community Health Worker (CHW) throughout

the duration of the 6-month intervention, and guidance on how to prepare for and participate in healthcare appointments.

- Control arm: I-DSMES
    - Patient only will receive a Diabetes Complications Risk Assessment profile and introduction session, 4-6 group DSME sessions, case management contacts with CHW throughout the duration of the 6-month intervention, and guidance on how to prepare for and participate in healthcare appointments.

3.1 SAMPLE SIZE AND POWER CONSIDERATIONS

We powered our study to detect a clinically meaningful difference in our primary outcome HbA1c at 6 months (Aim 1) based on a linear mixed model with intervention arm, linear time (0, 6, and 12 months), and interaction between intervention arm and time. We assumed that, due to randomization, the mean difference at baseline is zero and that the intervention effect is represented by the between group difference in time slope (captured by the intervention arm by time interaction), or equivalently, differences at follow-up time points. All calculations assumed at least 80% power and two-sided tests at level $\alpha = 0.05$ using the R package `longpower` (1). The standard deviation of HbA1c at any given time point was estimated to be between 1.5% to 2.2% based on previous studies (2–5). The correlation between timepoints, as measured by the intra-class correlation coefficient (ICC) ranged from 0.5 to 0.7 based on the RCT data at the Federally Qualified Health Center (FQHC) where the study is being conducted. Table 1 shows the total sample size required for various intervention effects.

Table 1: Total sample size (N) required for different intervention effects assuming a range of ICCs and SDs for *HbA1c*

SLOPE DIFFERENCE	MEAN DIFFERENCE AT 6 MONTHS	ICC	SD			
			1.50%	1.75%	2.00%	2.20%
0.067	0.40%	0.5	442	601	785	950
		0.6	354	481	628	760
		0.7	265	361	471	570
0.083	0.50%	0.5	283	385	503	608
		0.6	227	308	402	487
		0.7	170	231	302	365
0.010	0.60%	0.5	197	268	349	423
		0.6	157	214	280	338
		0.7	118	161	210	254
0.116	0.70%	0.5	145	197	257	311
		0.6	116	157	206	249
		0.7	87	118	154	187

To detect a 6-month mean difference of 0.6% in HbA1c, the total sample size required is 214 (highlighted in table above). We chose this sample size based detecting a clinically meaningful effect while considering the study size recruitment feasibility. The detectable effect is similar to what was observed in prior studies. We assumed an HbA1c standard deviation (SD=1.75%) and ICC (ICC=0.6) within the range of potential values reported in the

literature. If these estimates were conservative (i.e., if true SD=1.5% and true ICC=0.7), our sample size achieves at least 80% power to detect mean difference as small as 0.4%.

Anticipating 20% attrition based on the rate observed in prior 6-month and 12-month RCTs at the FQHC, we propose to recruit a total of 268 dyads (134 per group).

3.2 BLINDING

All assessments will be conducted by blinded research associates. Due to the nature of the intervention, participants, their SPs, and care providers cannot be blinded.

4.0 STUDY ENDPOINTS AND COVARIATES

4.1 PRIMARY ENDPOINT

The primary outcome is the change in patient glycemic control from baseline to 6 months. Glycemic control is measured by HbA1c in %.

4.2 SECONDARY ENDPOINTS

Secondary outcome measures include the following:

- Change in HbA1c from baseline to 12 months
- Change in blood pressure from baseline to 6 months and to 12 months
    - Blood pressure will be measured by an automatic blood pressure monitor, taking the average of two readings taken 2 minutes apart. We will average the SBP at 6 months and 12 months and compare to SBP at baseline.
- Change from baseline to 6 months and to 12 months in diabetes distress in patient and in SP
    - The Problem Areas in Diabetes (PAID-5) Scale is comprised of 5 closed-ended items with response options ranging from 0 ('not a problem') to 4 ('serious problem'). The scale's 5 items will be summed to create a total score with a range of 0 to 20. A total score of ≥8 indicates possible diabetes-related emotional distress that warrants further assessment, with higher scores suggesting greater diabetes-related emotional distress.
- Change from baseline to 6 months in diabetes self-care behaviors in patient - including domains of diet, exercise, blood sugar testing, blood pressure checking, foot care and medication taking.
    - The Summary of Diabetes Self-Care Activities (SDSCA) is a brief self-report instrument for measuring levels of self-management across different domains of the diabetes regimen. Our study will score results within each domain by averaging the item scores (each ranging from 0-7) within each domain.
- Change from baseline to 6 months in self-efficacy of patient and of SP
    - The Self-Efficacy for Managing Chronic Diseases Scale is comprised of 5 items asking respondents to indicate how confident they are that they regularly can perform tasks related to their diabetes management. Responses range from 0 to 10, with higher numbers indicating greater self-efficacy.
- Change from baseline to 6 months in patient activation
    - The Patient Activation Measure (PAM)-10 uses a 4-point scale (1=strongly disagree to 4=strongly agree), where respondents indicate the extent to which statements related to being ready, willing and able to manage their health and health care accurately describe them. Responses are summed to create a total score with higher numbers indicating greater activation.
- Patient perceived overall satisfaction with SP support for diabetes from baseline to 6 months

  - Items will assess patient's satisfaction with the support they receive from their SP and whether they feel like they would be worse off without their SP's help with their diabetes care. Responses will be rated on a 7-point scale (1=strongly disagree to 7=strongly agree). Responses are summed to create a total score with higher numbers indicating greater satisfaction.
- Patient perception of SP support from baseline to 6 months
  - Assessment will use the 8-item Important Other Climate Questionnaire (IOCQ) and non-supportive behaviors using 3 similarly-structured items addressing SP irritation, criticism and argumentativeness. All items are rated on a 7-point scale (1=strongly disagree to 7=strongly agree), with non-supportive behavior items being reversed scored.

5.0 STATISTICAL METHODS

5.1 GENERAL APPROACH

For all analyses, the overall level of significance will be set to α=0.05. To achieve maximum power for the primary endpoint, and because analyses involving the secondary outcomes, mediation, and moderation are considered hypothesis generating, no adjustment for multiple testing will be used. Data analyses will be performed using SAS v9.4 (SAS Institute, Cary, NC) or the latest version of R (currently 4.2.2; https://cran.r-project.org/).

We will examine baseline data for prognostically important differences across the two study groups. All main outcome analyses will be conducted at the individual patient level, with treatment group as the primary grouping factor. Prior to hypothesis testing, data will be examined using standard univariate descriptive measures and bivariate statistical measures of association, as well as graphical displays. Statistical assumptions required for the methods of analysis outlined in this section will be rigorously tested. The failure of statistical assumptions necessary for the techniques planned will be addressed in one of the following ways: 1) a different method of analysis will be used (e.g., non-parametric tests versus parametric tests); or 2) the data will be appropriately transformed to a distribution that meets the assumptions of the proposed statistical test.

5.2 ANALYSIS OF PRIMARY EFFICACY ENDPOINT (AIM 1)

The primary analysis will be intent-to-treat (ITT). Section 6.0 discuss details of how missing data and intercurrent events (ICEs) will be analyzed. We will test the hypothesis that the change in HbA1c from baseline to 6 months differed between FAM-ACT and individual focused arm.

We will first graphically examine how HbA1c change over time. We will then conduct our main analyses using linear mixed-effects models that include repeated measures at baseline, 6 months, and 12 months. To evaluate the effect of the intervention over time, we will include time in the model plus an interaction term for time x intervention arm. We will also explore models with time as a categorical instead of continuous variable and use the model that fits best, based on Akaike's Information and Schwarz's Bayesian Criteria. The primary hypotheses (change from baseline to 6 months) will be tested using appropriate linear contrasts. Treatment effect estimates will be presented along with 95% confidence intervals. The model will be adjusted for our design variables (baseline HbA1c and whether the patient and SP live together), baseline insulin use, and age. If baseline HbA1c and insulin use are highly correlated, we will prioritize HbA1c in the primary model and adjust for insulin use in sensitivity analyses.

5.3 ANALYSIS OF SECONDARY ENDPOINTS (AIM 2A)

Secondary health behavior and social support outcomes will be analyzed in a manner similar to the primary outcome. For categorical and ordinal outcome measures, we will use generalized linear mixed models (GLMM)

with logit link. Intervention effects at 6 months and 12 months will be estimated and tested for significance using linear contrasts.

5.4 EXPLORATORY MEDIATOR/MODERATOR ANALYSIS (AIM 2B)

Pre-specified subgroup analyses will be conducted to understand the intervention effect, and to identify subgroups of patients for whom the intervention was particularly beneficial and/or harmful. Pre-specified key moderators/subgroups are: baseline HbA1c ≤9% vs. >9%, patient uses insulin at baseline (Y/N), sex (man/woman), whether patient and SP live together (Y/N), whether SP has diabetes or pre-diabetes (Y/N), time since diabetes diagnosis (<1 year vs. ≥1 year), patient low health literacy (Y/N), patient level of SP support at baseline (high vs low), and patient enrollment relative to COVID pandemic start (before vs after March 2020).

Additional exploratory moderators and mediators of outcomes analysis will be selected for analyses based on the study conceptual model.

To examine moderation in intervention effectiveness, we will include an interaction of the variable of interest with the group*time term in our analytic models. A significant interaction term would indicate moderation, and the moderated effect estimates will be reported from the interaction model. If significant moderation was found, estimates and confidence intervals of the intervention effect within each subgroup defined by the moderator will be reported.

To examine mediation effects, we will apply the natural effect models (6) based on a counterfactual framework (7). We will use a two-way decomposition of the total effect of the intervention on A1c into an average direct effect and average indirect effect (i.e., other than through the mediator). In brief, generalized linear models with appropriate links will be fitted for each mediator, adjusting for the design variables (baseline HbA1c and whether the patient and SP live together), baseline insulin use, and age. Next, we will construct an expanded dataset by repeating each patient data in the original data set two times and including a hypothetical treatment arm variable representing FAM-ACT and I-DSMES. Each replicate in the expanded dataset will then be weighted by the ratio-of-mediator-probability (RMPW) calculated by dividing the probabilities corresponding to the values actually observed for the mediator given the actual exposure by that given the hypothetical exposure. Finally, a marginal structural model will be fitted under a GLMM framework for the outcome. Data expansion and calculation will be performed using the R package medflex (8).

5.5 BASELINE DESCRIPTIVE ANALYSES

We will compare the distribution of baseline variables between study arms to assess randomization success. However, no formal statistical hypothesis testing will be conducted to avoid unnecessary testing. We will examine comparability visually (histograms, boxplots, bar graphs) and by summary statistics: mean, standard deviation, median, minimum, and maximum for continuous variables and frequency and proportion for categorical variables.

6.0 ANALYTICAL ISSUES

6.1 MISSING DATA

We will describe the extent and reasons that data are missing, summarizing the proportion of patients with missing data for each outcome and by study arm and by site. We will compare baseline patient characteristics between those who have complete outcome data and those who do not.

A complete-case analysis may lead to loss of power, and bias may be introduced when there is substantial missing data and it is not completely at random. Our primary analytic approach (mixed models) can validly handle missing-

at-random data. However, as appropriate and consistent with our approach to handle ICEs (section 6.2) we will employ multiple imputations to conduct the ITT analyses.

6.2 INTERCURRENT EVENTS

We identified and evaluated all plausible strategies for each ICE in this study in line with the ICH E9 R1 guidelines. For each ICE, Table 2 below shows the strategy we selected (primary highlighted in green, supplemental highlighted in yellow) based on what most aligns with our clinical question of interest.

Table 2. Strategies for handling intercurrent events

ICE	STRATEGY	DESCRIPTION
Home A1c (instead of clinic)	Composite	- Home A1c will be analyzed if clinic A1c is unavailable.
	Hypothetical	- Home A1c will be treated as missing outcome data. Mixed modeling will project what would have been the clinic A1c at the timepoint from patient's available clinic A1c (e.g., BL, 6mo) of from similar patients.
Phone survey (instead of in-person)	Composite	- Phone survey will be analyzed if in-person survey is unavailable.
	Hypothetical	- Phone survey responses will not be used and treated as missing data. Mixed modeling will project what would have been the in-person survey response at the timepoint from patient's available in-person survey responses (e.g., BL, 6mo) of from similar patients.
Wider outcome assessment window	Hypothetical	- This will be more applicable when time enters the model as continuous variable. - Mixed model will project what would have been the outcome at the exact timepoint it should have been measured based on existing outcome data or trajectory or from similar patients.
	Composite	- This will be more applicable when time is entered into the model as categorical variable. - Data collected within the expanded window will be analyzed along with those collected in the original window at the timepoint.
Noncompliance	Treatment policy	- Observed outcomes after non-compliance will be used in the analysis. - This is the pure ITT approach.
	Principal Stratification	- Baseline characteristics will be used to identify the latent subgroup of compliers (in both intervention and control groups). - Patients observed or predicted to be compliers will be used in the analysis to estimate the complier average causal effect (CACE).

ICE	STRATEGY	DESCRIPTION
Delay in intervention classes	Treatment policy	- Observed outcomes after treatment delay will be used in the analysis. - This is the pure ITT approach.
	Hypothetical	- Observed outcomes after treatment delay will not be used in analysis and treated as missing data. - Mixed modeling will project what would have been the outcome had the patient received treatment on time using data prior to treatment delay or from similar patients.
Intervention delivery change (Zoom sessions)	Treatment policy	- Observed outcomes regardless of the intervention delivery modality will be used in the analysis. - This is the pure ITT approach.
	Hypothetical	- Outcome data after group zoom will not be used in the analysis and treated as missing data. - Mixed modeling will project what would have been the outcome had the patient received in-person intervention using data from similar patients.
Change in SP	Treatment policy	- Observed outcomes based on the original SP will be used for timepoints after the SP change. - This is the pure ITT approach.
	Hypothetical	- Outcome data after SP change will not be used in the analysis and treated as missing data. - Mixed modeling will project what would have been the outcome had the patient not changed SP using data prior to the change in SP or from similar patients.

6.3 SENSITIVITY ANALYSES

We will assess the robustness of study findings to missing data. We will perform a missing-not-at-random (MNAR) sensitivity analysis via control-based imputation in which all data are imputed based on the control arm. This assumes that the unobserved outcome in the FAM-ACT arm would have been similar to what was observed in the individual-focused arm. It imputes an outcome almost certainly better than that assumed by MAR and is unlikely to positively bias the estimates.

As supplementary estimand to the ITT effect, we will use an instrumental variable approach to estimate the complier average causal effect CACE, which measures the impact of the intervention in the subgroup of the population that complies with the assigned intervention (9,10). The intervention assignment will be used as the instrument since due to randomization, its impact on outcome is expected to be entirely mediated through the receipt of the intervention (11). As alternative to instrumental variable approach, propensity scores can be used to balance background characteristics facilitating a fair comparison between groups. We will also conduct a propensity score approach to estimate the CACE. To build the propensity score model for compliance, we will consider all available baseline variables of patients. We will then adopt the principal causal effect estimation method to estimate CACE (12).

The complier subgroup will be identified using the following definition:

FAM-ACT

- Intro session: attended by both patient and SP
- Patient classes attendance: at least 4 (out of 6) attended by patient
- SP classes attendance: at least 3 (out of 6) attended by SP
- Follow-up CHW contacts: At least 1 attended by patient

I-DSMES

- Intro session: Attended by patient only (not SP)
- Patient classes: at least 4 (out of 6) attended by patient
- SP classes: no attendance by SP
- Follow-up CHW contacts: At least 1 attended by patient

In sensitivity analyses, complier subgroup will be identified using patient attendance criteria alone.

6.4 ASSESSING IMPACT OF COVID-19 PANDEMIC

We employ recommended strategies to address analytic issues associated with COVID-19 (13). Most of the events or changes in protocol associated with COVID-19 were considered ICEs, analyses of which are detailed in section 6.3. In addition to these, we will define pre-COVID period as prior to March 2020 and COVID period beginning March 2020, and conduct the following analyses:

- To determine whether the target population remained the same, we will assess changes in enrollment and in study population over COVID-19 periods.
    - We will compare the recruitment rate (as % of screened) pre- vs. during COVID.
    - We will compare the characteristics of patients and SPs enrolled pre- vs during COVID.
- To determine whether COVID has a differential impact on the intervention, we will assess heterogeneity of effects across the COVID periods by including a group*COVID interaction in the analytic model. If the interaction was significant, we will report effect estimates stratified by COVID period.

REFERENCES

1. Donohue MC. longpower: Power and sample size calculations for linear mixed models. 2021.

2. Two Feathers J, Kieffer EC, Palmisano G, Anderson M, Sinco B, Janz N, et al. Racial and Ethnic Approaches to Community Health (REACH) Detroit partnership: improving diabetes-related outcomes among African American and Latino adults. Am J Public Health. 2005 Sep;95(9):1552–60.

3. Spencer MS, Rosland AM, Kieffer EC, Sinco BR, Valerio M, Palmisano G, et al. Effectiveness of a Community Health Worker Intervention Among African American and Latino Adults With Type 2 Diabetes: A randomized controlled trial. Am J Public Health. 2011 Dec;101(12):2253–60.

4. Hirst JA, Stevens RJ, Farmer AJ. Changes in HbA1c level over a 12-week follow-up in patients with type 2 diabetes following a medication change. PloS One. 2014;9(3):e92458.

5. Tang TS, Funnell M, Sinco B, Piatt G, Palmisano G, Spencer MS, et al. Comparative effectiveness of peer leaders and community health workers in diabetes self-management support: results of a randomized controlled trial. Diabetes Care. 2014 Jun;37(6):1525–34.

6. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol. 2012 Aug 1;176(3):190–5.

7. Pearl J. Direct and indirect effects. In: Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2001. p. 411–20. (UAI'01).

8. Steen J, Loeys T, Moerkerke B, Vansteelandt S. medflex: An R package for flexible mediation analysis using natural effect models. J Stat Softw. 2017;76(11):1–46.

9. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. J Am Stat Assoc. 1996 Jun 1;91(434):444–55.

10. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health. 2000;21:121–45.

11. Ten Have TR, Normand SLT, Marcus SM, Brown CH, Lavori P, Duan N. Intent-to-treat vs. Non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. Psychiatr Ann. 2008 Dec;38(12):772–83.

12. Jo B, Stuart EA. On the use of propensity scores in principal causal effect estimation. Stat Med. 2009 Oct 15;28(23):2857–75.

13. Meyer RD, Ratitch B, Wolbers M, Marchenko O, Quan H, Li D, et al. Statistical issues and recommendations for clinical trials conducted during the COVID-19 pandemic. Stat Biopharm Res. 2020;12(4):399–411.