Pre-Analysis Plan
Impact Evaluation of Online Program Delivery of Catholic Charities Fort Worth's LIFT Program

Michael Kofoed
University of Tennessee, Knoxville

Patrick Turner
University of Notre Dame

February 2025

I. Overview

The following is a pre-analysis plan for the randomized evaluation of program delivery method (online versus in-person) for Catholic Charities Fort Worth's (CCFW) LIFT program. We begin with the motivation for the program and the randomized evaluation in Section II. Section III lays out the target population, recruitment strategies, and eligibility criteria. Section IV describes our planned data sources, and Section V defines our primary and secondary outcomes. In Section VI, we present and discuss our power calculations relative to observational evidence and related studies. Finally, our empirical strategy is laid out in Section VII.

II. Background

Various studies using randomized design to explore the impact of virtual education in a college-length course have found significant negative impacts on learning outcomes relative to in-person instruction (Alpert et al., 2016; Bettinger et al., 2017; Figlio et al., 2023; Kofoed et al., 2024). Online students also struggled to concentrate in class and felt less connected to their instructors and peers (Kofoed et al., 2024).

Virtual instruction has differing effects across demographics. Negative impacts of online instruction are particularly strong for Hispanic students, male students, and lower-achieving students (Figlio et al., 2023). Remote peer mentoring has positive effects on motivation and studying behavior, particularly for the most able students (Hardt et al., 2022). Estimates imply that the availability of streaming services increases exam scores for high-ability students and decreases exam scores for lower-achieving students, though least-able students still benefit from live-streaming options because they save on attendance costs (Cacault et al., 2021). Other evidence suggests that students who select into the online classes perform better than they would, all other things constant, in a face-to-face class (Coates et al., 2004).

In the medical field, various virtual interventions have proven to be feasible, effective, and noninferior in achieving health outcomes, including telegram-based virtual education for adolescents with moderate-to-severe asthma, virtual obesity pharmacotherapy in adults, and Virtual House Calls for patients with multiple sclerosis (Faraji et al., 2020; Griebeler 2022; Robb et al., 2019). Virtual ultrasound instruction appears to be an effective alternative to traditional in-person instruction for pediatric residents (Gillon et al., 2024).

CCFW offers LIFT, a program that combines individualized case management, resource connections, financial coaching, and strategic financial assistance with the goal of advancing an individual's long-term financial freedom, labor market success, and well-being. Combining positive evidence from prior RCTs evaluating CCFW's case management program (Evans et al., 2025) and LIFT program data, it is likely that LIFT is an effective program, but it is vital to evaluate the effectiveness of various delivery methods (in-person and virtual) before expanding the program to the more rural areas of CCFW's 28 service counties. While most researchers have conducted randomized controlled trials comparing virtual and in-person content delivery methods in education or health care with varied results, there is little evidence on remote work in the workplace. There is a distinct need for more evidence regarding the effects of online support in the non-profit case management setting, which requires a unique one-on-one relationship. The significance of researching the benefits of virtual versus in-person case management extends far beyond North Texas. By conclusively determining if virtual services are as effective as in-person services, this RCT not only enables CCFW to serve more clients and expand their reach, but also demonstrates to other nonprofits and government entities that they, too, can broaden their services to assist more families in need, regardless of location.

III. Evaluation Design

A. Research Questions

Does virtual case management delivery, in comparison to in-person case management, impact outcomes including program engagement, goal attainment, self-reported trust in case managers, and financial outcomes including credit, debt, employment, and earnings?

B. Eligibility

LIFT applicants are eligible for both the program and study enrollment are 18 and above, speak English or Spanish, are ready and able to work, and live or work within the 28-county service area. Anyone with a financial coaching need is eligible; there are no income limitations on the sample.

C. Randomization

Starting in February 2025, eligible LIFT clients who provide study consent will be randomized during an intake call with an Operations Specialist on CCFW staff. Then, a CCFW LIFT Program Manager will assign them to a virtual or in-person navigator based upon their randomized group assignment. If an individual does not consent to be part of the study, the outreach specialist will refer him or her to Padua, CCFW's holistic case management services.

We plan to enroll LIFT clients into the study for three years, with a total of about 1,300 study participants. We estimate that 650 clients will be randomized into virtual program delivery, and 650 clients will be randomized into in-person program delivery over the course of the study. These numbers may vary depending on the take-up rate for program participation; for example, if a higher rate of clients originally assigned to the virtual delivery group turn down the program, more clients will need to be randomized, perhaps at a different randomization ratio, into virtual delivery in order to fill the spots that become available.

D. Intervention

The LIFT program combines individualized case management, resource connections, financial coaching, and strategic financial assistance with the goal of advancing an individual's long-term financial freedom, labor market success, and well-being. In one-on-one consultations, paired navigators focus on three areas contributing to self-sufficiency: financial resilience, resource stability and emotional resilience, the ability to plan and cope. A tenet of the LIFT intervention is that a client's readiness for change is essential to her or his success, and to achieve financial independence, individuals must actively contribute to creating their goals, objectives, and service plans.

Since LIFT's inception in 2016, the program has helped more than 500 clients in the Dallas-Fort Worth area. An earlier RCT conducted collaboratively by CCFW and LEO evaluated a similar case management program by CCFW called Padua. The results of the study show that Padua led to a significant improvement in labor market and housing stability for those in the Padua program. CCFW's strong culture of evaluation extends into their Research & Analytics Department, which tracks program data suggesting that LIFT similarly increases earnings, employment, and overall well being. Combining the evidence from the Padua RCT and LIFT program data, it is likely that LIFT is an effective program, but it is vital to evaluate the effectiveness of various delivery methods (in-person and virtual) before expanding the program to the more rural areas of CCFW's 28 counties. Therefore, all study participants will take part in the LIFT program and have access to the same resources, but randomization will determine if their case management sessions with their navigator occur in person or online.

IV. Data Sources

A. Catholic Charities Fort Worth

CCFW gathers baseline information on new program participants during the application and intake phases of enrollment. Program records from CCFW will provide engagement measures including details of coaching sessions and goal attainment which will allow for descriptive work of the treatment and treatment on the treated analysis. As part of internal program evaluation for clients in LIFT, CCFW invites their clients to complete a set of surveys covering topics of emotional resilience, resource stability, trust in navigator, and financial knowledge. These surveys are collected as part of regular contact between navigators and clients, and these responses will be used in the construction of study outcomes. LEO has a data sharing agreement with CCFW.

B. Ray Marshall Center (RMC) and the Texas Workforce Commission and Texas Health and Human Services Commission

RMC at the University of Texas at Austin facilitates the use of Texas administrative data sources for researchers. Employment and benefits usage data will come through RMC. LEO has an existing relationship and data sharing agreement with RMC.

C. National Student Clearinghouse

This source will provide the research team with post secondary enrollment records and degree completion data nationwide. This allows us to observe degree completion for the whole sample. LEO has already established a data use agreement with NSC.

D. Infutor Data Solutions

We plan to measure housing stability using data from Infutor Data Solutions. Infutor collects data on consumers' living addresses in the United States. We will use this data to track individuals' address histories, allowing us to gauge housing stability based on how frequently an individual's address changes over time. LEO has an existing relationship with Infutor which will allow us to collect this data on study participants.

E. Experian

We plan to measure the impact of the intervention on participants' credit score, use of credit, and total balance in delinquent accounts using data from Experian. LEO has an existing relationship with Experian which we will use to link records in this study with Experian's credit data.

V. Study Outcomes

CCFW's LIFT program is designed to allow clients to prioritize their goals and take an active role in working toward them. The program aims to support clients toward self-sufficiency through consistent contact and trust-building with program navigators. These essential program elements are why this evaluation will primarily focus on the program outcomes of participants, pulled from CCFW program data, to evaluate whether the virtual and in-person delivery methods deliver the same core components that have shown to best support clients. Preliminary results will be updated and reported for each subsequent year of the study. Two-year results will be available for the full sample in 2030.

A. Primary Outcomes

- Program engagement: indicator for whether a client is continuously engaged or has already collaboratively closed (successfully completed) their enrollment in the program. Measured at 6, 12, and 18 months following randomization.
- Goal attainment index: Rather than analyze multiple subgroups and outcomes in separate regressions, we will construct an index of goal attainment based upon study participants' primary goal as stated at baseline. The index will be measured at 6, 12, and 18 months following randomization. The index will combine standardized measures of goal attainment with z-scores using the measure's observed value for a given participant, the mean of the measure across the sample, and the standard deviation of the measure across the sample. To ensure comparability across different outcomes, we will recalibrate the direction of z-scores so that positive values consistently represent progress toward goals (e.g., a decrease in debt will be transformed to result in a positive z-score). The index will include the following measures, where the stated goal is listed and followed by the measure of attainment:

    - Increasing savings: standardized measure for increase in self-reported savings
    - Decreasing debt: standardized measure for decrease in Experian-reported total balance on delinquent accounts
    - Building credit: standardized measure for increase in Experian-reported credit score
    - Housing stability: standardized measure for count of Infutor-reported address changes
    - Career/job opportunities: standardized measure for measure of UI earnings
    - Education/certification: standardized measure for NSC-reported initiated enrollment or degree/certificate completion
    - Emotional health/stress reduction: standardized measure for increase in self-reported emotional resilience

We will also report average treatment effects for each component of this index separately as secondary outcomes.

- Credit score: Experian-reported credit score measured quarterly following randomization
- Employment: state administrative data-reported UI benefits receipt measured quarterly following randomization

B. *Secondary Outcomes*

- Program engagement: Alternative measures of program engagement include number of sessions attended, time spent enrolled in program, total time spent in sessions, number of 'no shows' at sessions, and an indicator of premature disengagement.
- Debt: Experian-reported total balance on delinquent accounts measured quarterly following randomization
- Degree completion: NSC-reported enrollment initiation or degree/certificate completion measured at 6, 12, and 18 months following randomization
- Housing stability: Infutor-reported address changes within 6, 12, and 18 months following randomization
- Benefits usage: state administrative data-reported lack of benefits usage measured quarterly following randomization at 12 and 18 months following randomization
    - Alternative measure if study population is likely to benefit more from connection with necessary resources: state administrative data-reported take-up of benefits measured at 6, 12, and 18 months following randomization
- Self-reported trust in navigator: We will construct a standardized score of a client's responses to a series of Likert-scale statements regarding their trust in their program navigator. Measured at 6, 12, and 18 months following randomization.
- Resource stability: We will construct a standardized sum of a client's responses to a series of "yes/no" questions regarding their ability to utilize resources or pay for standard costs. Measured at 6, 12, and 18 months following randomization.

- Emotional resilience: We will construct a standardized score of resilience according to a client's responses to a series of Likert-scale statements as a part of the BRIEF (Behavior Rating of Executive Function) questionnaire. Measured at 6, 12, and 18 months following randomization.
- Financial knowledge and stability: We will construct a standardized score of a client's responses to a series of "True/False" and Likert-scale statements. Questionnaire based upon CFPB Financial Empowerment Self-Assessment, CFPB Financial Well-Being Scale, and The Financial Management Behavior Scale. Measured at 6, 12, and 18 months following randomization.

VI. Statistical Power and Sample Size

We plan to enroll a study sample of 1,300 individuals, with approximately half of these assigned to the treatment group (virtual coaching), and a program take-up rate of 100%. We assume a high take-up rate due to the fact that randomization occurs after they have initiated program contact with CCFW staff but before any randomized activities begin. We are powered to detect a 6.7 percentage point change in the rate of continued engagement or program completion at 12 months following randomization. If we are unable to reject the null hypothesis that there is no significant difference between the outcomes of the randomized groups, we will be able to conclude that virtual program delivery has no impact, and is, in terms of outcomes, the same as in-person program delivery.

VII. Empirical Strategy

A. Main Specification

We will estimate intent-to-treat (ITT) treatment effects by OLS using the following regression:

$$(1) \qquad Y_i = \theta_1 T_{1i} + X_{ist}\theta_3 + \mu_{1s} + \lambda_{1t} + \epsilon_{1ist}$$

where Y_i is the outcome. T_{1i} is an intent-to-treat dummy indicating the random assignment of person *i* to the virtual program delivery group. The vector X_{ist} includes a set of person-level characteristics collected at baseline. We will also control for service area fixed effects (μ_{1s}) and year-quarter fixed effects (λ_{1t}), and ϵ_{1ist} is an error term. The coefficient on the treatment dummy, θ_1 will give us the difference in means between the treatment (virtual) and control (in-person) groups, or the estimated impact of virtual delivery of the program. If θ_1 is not statistically different from zero, it may be concluded that virtual delivery has no impact on specified outcomes.

B. Treatment on Treated Specifications

In addition to the reduced-form estimates obtained in the equations above, we are also interested in estimating the causal impact of virtual program delivery, also known as the *treatment-on-treated* (TOT) effect. In this case, not all clients will uniformly comply with their group assignment (e.g., someone in the online group may be adamant they will not attend virtual sessions). To this end, we will estimate the TOT by instrumenting for intervention participation with treatment assignment using a two-stage least squares regression framework.

We can examine the impact of the receipt of virtual program delivery of sessions on outcomes. In this case, let V_i be a dummy that equals 1 if the person participated in at least one virtual coaching session. The equation of interest in this case can be described by the equation

$$(2) \qquad Y_i = \beta_0 + \beta_1 V_i + X_{ist}\beta_3 + \mu_{2s} + \lambda_{2t} + \epsilon_{2ist}$$

As V_i is endogenous, we would need to use the assignment to the virtual program delivery treatment group (T_{2i}) where the first-stage regression is then

$$(3) \qquad V_i = \Phi_0 + \Phi_1 T_{2i} + X_{ist}\Phi_3 + \mu_{3s} + \lambda_{3t} + \epsilon_{3ist}$$

In all models, we will use heteroskedasticity-robust standard errors.

C. Treatment Effect Heterogeneity and Subgroup Analyses

Given that this study will recruit a broad range of clients, virtual program delivery may have different effects within different sub-groups. Understanding whether the program delivery method is significant broadly, for some sub-population of policy interest, or is most beneficial for some surprising sub-sample provides crucial information to CCFW, similar nonprofits, and governments on how they might scale programs in the event of an important finding.

The study will estimate the impact of virtual program delivery across several outcome categories and subgroups. The research team is interested in determining whether virtual program delivery seems to be a viable substitution for certain populations relative to others.

Primary subgroup

- The central research question surrounds the ability for CCFW and similar nonprofit organizations to expand services to more rural areas where resources may be more difficult to access. We will explore heterogeneity by participants whose home address at baseline is in a census tract that is above vs. below the median census tract population density per square mile.
    - Alternatively, this subgroup split may be better characterized by above vs. below the median distance from a CCFW hub location.

Other Primary Subgroups

- To measure level of poverty and resources at baseline: consisting of household income, employment status, household structure, housing stability, and benefits usage (SNAP/TANF, etc.). We will split the sample at the median of the baseline index for poverty level and family resources.

- Alternatively, this subgroup may be constructed using employed vs unemployed at baseline.
- Client gender
- Client age: above vs. below the median age of participating clients

Secondary Subgroups

- Client race
- Primary goal at baseline
- Above vs. below the median standardized score of responses to resource stability, emotional resilience, and financial knowledge/stability surveys at baseline
- Navigator caseload

D. Exploring Heterogeneity using Machine Learning

We may draw on an emerging literature that leverages machine-learning methods to explore heterogeneity of causal effects (Chernozhukov et al., 2018; Athey and Imbens, 2015, 2019; Davis and Heller, 2017). This methodology will enable us to learn as much as possible from our data using a disciplined and data-driven approach. Since the "state of the art" is still evolving, we cannot commit to a particular approach at present. However, we plan to pre-specify our approach prior to running this analysis and will interpret our results as suggestive.

E. Multiple Hypothesis Testing

Testing multiple hypotheses raises the likelihood that any one hypothesis is found to be statistically significant purely by chance. We will supplement our results by reporting summary indices that aggregate multiple outcome variables within a common outcome domain. Aggregation not only improves the statistical power within a given domain but also vastly reduces the number of hypotheses examined. This plan pre-specifies what data will be collected, primary and secondary outcomes, the main specification, and subgroups of interest. By committing to a set of analyses in advance, we avoid concerns about data-mining and specification searching, and credibly commit to a few hypotheses that, together, comprise the central test of CCFW's program. Classic p-values will be reported for all outcomes, which will provide a reader with full information that they can use to make multiple hypothesis testing corrections if they desire. We will also conduct non-parametric permutation tests and report permuted p-values for the main sets of analyses following Chetty et al. (2016).[1]

---

[1] This approach entails randomly re-assigning treatment status to students in the main sample and running the main specification thousands of times to simulate a counterfactual distribution of T-statistics. Relative to this counterfactual distribution, we can then compute permuted p-values as likelihood of observing our realized T-statistic. The same approach can be applied to sets of hypotheses to calculate the likelihood of observing by chance the magnitudes of treatment effects observed in the study.

VIII. References

Alpert, William T., Kenneth A. Couch, and Oskar R. Harmon. 2016. "A Randomized Assessment of Online Learning." *American Economic Review: Papers & Proceedings*, 106(5): 378–382.

Athey, S. and G. W. Imbens. 2015. "Machine Learning Methods for Estimating Heterogeneous Causal Effects." *stat,* 1050(5), 1–26.

Athey, S. and G. W. Imbens. 2019. "Machine Learning Methods that Economists Should Know About." *Annual Review of Economics,* 11: 685–725.

Bettinger, Eric P., Lindsay Fox, Susanna Loeb, and Eric S. Taylor. 2017. "Virtual Classrooms: How Online College Courses Affect Student Success." *American Economic Review,* 107 (9): 2855–2875.

Cacault, Paula M., Christian Hildebrand, Jeremy Laurent-Lucchetti, and Michele Pellizzari. Forthcoming. "Distance Learning in Higher Education: Evidence from a Randomized Experiment." *Journal of the European Economic Association*.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2018. "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India." National Bureau of Economic Research Working Paper 24678.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review*, 106 (4): 855-902.

Coates, Dennis, Brad R. Humphreys, John Kane, and Michelle A. Vachris. 2004. "No Significant Distance Between Face to Face and Online Instruction: Evidence from Principles of Economics." *Economics of Education Review*, 23 (5): 533–546.

Davis, J. and Sarah B. Heller. 2017. "Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs." *American Economic Review,* 107 (5), 546–50.

Evans, William, Shawna Kolka, James X. Sullivan, and Patrick S. Turner. 2025. "Fighting Poverty One Family at a Time: Experimental Evidence from an Intervention with Holistic, Individualized, Wrap-Around Services." *American Economic Journal: Economic Policy*, 17 (1): 311-61.

Figlio, David, Mark Rush, and Lu Yin. 2013. "Is it Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning." *Journal of Labor Economics*, 31(4): 763–784.

Gillon, Jason T., E. Liang Liu., Valerie Dutreuil, Stephanie G. Cohen, and Lekha A. Shah. 2024. "Comparison of In-person Versus Virtual Ultrasound Instruction for Pediatric Residents." *BMC Med Educ,* 24, article 203.

Griebeler, Marcio L., W. Scott Butsch, Paloma Rodriguez, Laure Lomeli, Matthew Kampert, Vinni Makin, Ula Abed Alwahab, Elena Borukh, Erin Daigle, James Bena, Kevin M. Pantalone, and Bartolome Burguera. 2022. "The Use of Virtual Visits for Obesity Pharmacotherapy in Patients with Overweight or Obesity Compared with In-person Encounters. *Obesity (Silver Spring)*, 30 (11): 2194‑2203.

Hardt, David, Markus Nagler, and Johannes Rincke. 2022. "Can Peer Mentoring Improve Online Teaching Effectiveness? An RCT During the COVID-19 Pandemic." *Labor Economics,* 78.

Kofoed, Michael S., Lucas Gebhart, Dallas Gilmore, and Ryan Moschitto 2024. "Zooming to Class?: Experimental Evidence on College Students' Online Learning during COVID-19." *American Economic Review: Insights, 6*(3): 324-340.

Robb, Jessica F., Megan H. Hyland, and Andrew D. Goodman. 2019. "Comparison of Telemedicine Versus In-person Visits for Persons with Multiple Sclerosis: A Randomized Crossover Study of Feasibility, Cost, and Satisfaction." *Multiple Sclerosis and Related Disorder*, 36.