SIF Final Evaluation Report

Subgrantee: Mercy Ministries of Laredo

Project Title: Sí Three: Integration of 3-D Health Services

Submitted by:

SIF Grantee: Methodist Healthcare Ministries of South Texas, Inc.

Program Name: Sí Texas: Social Innovation for a Healthy South Texas

Prepared by:

Evaluator: Health Resources in Action, Inc.

October 2018

Methodist Healthcare Ministries' Sí Texas program is a proud recipient

of the Social Innovation Fund (SIF) program.

This material is based upon work supported by the Corporation for National and Community Service (CNCS) under Social Innovation Fund Grant No. 14SIHTX001. Opinions or points of view expressed in this document are those of the authors and do not necessarily reflect the official position of, or a position that is endorsed by, CNCS.

Sí Texas Subgrantee: Mercy Ministries of Laredo
Program Title: Sí Three: Integration of 3-D Health Services

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

EXECUTIVE SUMMARY

This final report provides an overview of progress and findings for the evaluation of the Sí Three project of Mercy Ministries of Laredo (Mercy), a subgrantee of the Social Innovation Fund (SIF) Grantee Methodist Healthcare Ministries (MHM) of South Texas, Inc. MHM is a member of the 2014 SIF cohort. The evaluation was conducted by external evaluation contractor, Health Resources in Action (HRIA), at Mercy and an external comparison site, Nuestra Clinica del Valle.

Program Background

Mercy Ministries of Laredo (Mercy) is a primary healthcare clinic located in Webb County, Texas, which provides healthcare and health education to some of the poorest neighborhoods and colonias in the U.S. Mercy began implementing the Sí Three program and the evaluation study in the Laredo clinic in January 2016. Mercy completed evaluation data collection in July 2017. The Sí Three program expanded Mercy's efforts to integrate behavioral health, including optional faith-based behavioral health services, into primary healthcare services. The Sí Three program, aimed to improve behavioral health conditions (e.g., depression, anxiety, and addictive behavior) and chronic disease conditions (e.g., hypertension, obesity, and diabetes) through interventions that addressed the physical, behavioral, and spiritual health of patients as well as overall quality of life.

Prior Research

The Sí Three program targeted a moderate level of evidence through combining components of the integrated care model studied by Druss et al. (Druss, Rohrbaugh, Levinson, & Rosenheck, 2001) with faith-based care discussed by Worthington et al. (Worthington, Hook, Davis, & McDaniel, 2011). The Druss model included patient education and prevention, nurse practitioners, and increased interaction among the care team. Worthington found that religious/spiritual counseling resulted in greater improvements in psychological and spiritual outcomes when compared with secular therapies (Worthington, Hook, Davis, & McDaniel, 2011).

Evaluation Design

Mercy conducted a quasi-experimental design (QED) study with two comparison groups: an internal primary comparison group comprised of patients receiving care from Mercy's Laredo clinic and an external secondary comparison group comprised of patients receiving care from Nuestra Clinica del Valle at its Edcouch and Alton locations. The primary comparison group minimized several threats to internal validity given that intervention and primary comparison group participants were more similar at baseline on demographic and outcome measures and the primary comparison group participants were patients in the same clinic as the intervention group. More specifically, the primary comparison group addressed the following threats to internal validity: regression to the mean, history, testing, John Henry, and expectancy effects. The secondary comparison group enhanced internal and external validity and served as a sensitivity analysis for the primary comparison group findings. More specifically, the secondary comparison group addressed the internal validity threats of selection bias and novelty. Also, the secondary comparison group addressed external threats to validity including applicability to other populations and applicability to other settings/locations.

Mercy's recruitment target was 205 participants in each of the three study groups (intervention, primary comparison at Mercy, and secondary comparison at Nuestra Clinica del Valle), totaling 615 participants. Mercy met its clinic recruitment of 410 targets for the intervention group and primary comparison at its Laredo clinic (207 in the intervention group and 203 in the primary comparison group) in July 2016. Enrollment in the secondary comparison group at Nuestra Clinica del Valle was completed in April 2017 and the enrollment target was met. Mercy's 12-month retention target was 492, with 164 in each study arm. The final 12-month sample was 550 participants: 142 in the intervention group, 151 in the primary comparison group, and 257 in the secondary comparison group.

The implementation evaluation focused on measuring the level of program services provided and the quality of services program participants received relative to what was proposed. In addition, the implementation evaluation assessed the extent to which the comparison groups received program services similar to services received by the intervention group.

Description of Measures and Instruments

Mercy collected data for the Sí Texas shared impact measures: Body Mass Index (BMI) (calculated from height and weight), HbA1C (obtained via blood test as clinically needed), blood pressure (taken by provider), depression (using the Patient Health Questionnaire (PHQ-9), and quality of life (as measured by the Duke Health Profile). In addition to the shared measures, Mercy collected data utilizing the CAGE-AID to measure addictive behavior, GAD-7 to measure anxiety, waist circumference to measure weight control, and the Spirituality Index to measure receptivity for faith-based counseling. The primary impact measures were improvement in depression and BMI.

Research Questions

The primary impact measures for Sí Three: Integration of 3-D Health Services were improvement in depression and BMI. Below are the confirmatory and exploratory research questions:

1) Do patients who participate in the Sí Three intervention experience improvements in depressive symptoms, as measured by PHQ-9, after 12 months compared to patients who do not participate? *This question is confirmatory.*
    a. In addition, do these improvements differ by type of behavioral health service received (medical/behavioral or faith-based services)? *This question is exploratory.*
2) Do patients who participate in the Sí Three intervention experience improvements in BMI after 12 months when compared to patients that do not participate in the intervention? *This question is confirmatory.*
3) Do patients who participate in the Sí Three intervention experience improvements in quality of life, as measured by the Duke Health Profile, after 12 months when compared to patients that do not participate in the intervention? *This question is exploratory.*
4) Do patients who participate in the Sí Three intervention experience improvements in anxiety symptoms, as measured by GAD-7, after 12 months compared to patients who do not participate? In addition, do these improvements differ by type of behavioral health service received (medical/behavioral or faith-based services)? *This question is exploratory.*
5) Do patients who participate in the Sí Three intervention experience improvements in addiction symptoms, as measured by CAGE-AID, after 12 months compared to patients who do not participate? *This question is exploratory.*

6) Do patients who participate in the Sí Three intervention experience improvements in blood pressure, when compared to patients that do not participate in the intervention? *This question is exploratory.*
7) Do patients with a history or diagnosis of diabetes who participate in the Sí Three intervention experience improvements in HbA1c measures after 12 months when compared to patients that do not participate in the intervention? *This question is exploratory.*

Implementation Questions

The following evaluation questions examined program implementation and patient and provider satisfaction.

1. Did the Sí Three program reach its intended target population?
2. What are the components of the Sí Three program and how do these components work "on the ground" at 6 and 12 months?
    a. Are these components different than what was planned? If so, why?
3. What level of integrated behavioral health did Mercy Ministries achieve as a result of implementing the Sí Three program?
    a. To what extent have providers and staff adopted the components of the Sí Three program at 6 and 12 months? What are the facilitators and barriers to adoption?
    b. To what extent do providers and staff buy in to the Sí Three program, and how has buy-in affected implementation?
4. To what extent did the comparison groups receive program-like components?
5. To what extent did the Mercy clinic implement the Sí Three model with fidelity?
6. How satisfied are Sí Three patients with the services they received? How satisfied are providers with the Sí Three program?

Additional implementation evaluation questions include the following:

7. What percent of patients who were seen by primary care providers for diabetes, obesity, and/or hypertension completed standardized assessments (depression, anxiety, addictive behavior, quality of life, and spiritual well-being) on their initial visit?
8. What percent of completed assessment results were recorded according to protocol? Were all staff able to implement standard measurement protocols?
9. What percent of patients with depression, anxiety, and addictive behavior were referred to the Licensed Professional Counselor (LPC) or other behavioral health provider?
10. What percent of patients were assessed for depression, anxiety, quality of life, and addictive behavior on a semi-annual basis?
11. What was the show rate for all patients in the Sí Three intervention for IBH services? Did the show rate differ by the type of behavioral health resource the patient used?
12. What percent of referred patients could explain their physical and/or behavioral health treatment plans?

Impact Analysis

This report presents descriptive statistics, analysis of baseline equivalence, and analyses of impact across the study groups. All analyses were conducted based on an intention-to-treat approach. The unit of analysis was the individual patient. Impact measures are treated as continuous variables. Propensity

score matching was explored but not considered an appropriate approach due to insufficient number of covariates to match on to meet propensity score analysis assumptions and the completeness of data from all study groups. Generalized regression analysis results are presented as final results of the modeling sequence starting with bivariate models and ending with multiple regression models. These multiple regression models are adjusted for key demographic factors, covariates, and baseline impact measures identified as relevant via review of the scientific literature or found non-equivalent at baseline. The possibility of effect modification of the intervention-outcome relationship by patients' characteristics was also explored. Specifically, interaction terms of study group and baseline impact measures as well as age were included to understand whether there were differences in intervention effect by these characteristics. Stratified linear regression models were subsequently estimated for any model that found statistically significant effect modification.

Program implementation was assessed by reviewing collected measures at the pre-determined time points to identify any opportunities to improve implementation fidelity or need for statistical adjustments in impact analysis due to problems with implementation fidelity.

Key Findings

Evaluation of Mercy's implementation of the Sí Three program shows that the program was implemented in alignment with the program logic model and that there was strong fidelity in implementation. Facilitators to program implementation included communication among staff, staff experience with using an electronic medical record, moving staff offices to facilitate communication among physical and behavioral health staff, hiring staff who had specific roles that supported IBH, leadership and staff buy-in to the program, and clinic workflow adjustments to ensure patient needs for services could be met. For patients, additional factors that facilitated their participation included the low cost of services, clinic staff flexibility to meet their needs, strong rapport between patients and staff, and support for patient transportation services.

Study results indicate that the Sí Three program improved behavioral health among intervention participants. Consistent improvements were noted in behavioral health outcomes between the intervention participants and two comparison groups with intervention participants primarily receiving faith-based counseling services. More specifically, the study showed that, when controlling for baseline measures and other covariates, the intervention participants had significantly greater improvements when compared with the primary comparison group participants in the depression outcome over time (reduced depression as measured through PHQ-9 over the study period which includes baseline, 6-month and 12-month, β=-1.76, p=0.001). Over time results refers to analysis of the trajectory of change from baseline to 6 months, and to 12 months between the intervention and comparison groups and 12-month, or end of the study, results refers to analysis of the differences between intervention and comparison group at the 12-month assessment only.

Also, intervention group participants had significantly greater improvements when compared with primary comparison group participants on additional outcomes identified in the logic model (increased Duke General Health score at 12 months (the end of the study) β = 4.01, p=0.02, Cohen's d 0.24; increased Duke Physical Health Score at 12 months β=6.69, p=0.004; increased Duke General Health Score over time β=5.35, p=0.03; decreased GAD-7 at the end of the study β=-0.79, p=0.03, Cohen's d 0.22; and decreased GAD-7 over time β=-1.58, p=0.002).

Further, the study showed that, when controlling for baseline measures and other covariates, the intervention participants had significantly greater improvements when compared with the secondary comparison group participants in the depression outcome over time (reduced depression as measured through PHQ-9, β=-2.78, p=0.001) and additional outcomes identified in the logic model (increased Duke General Health Score over time β=5.96, p=0.001; decreased GAD-7 over time β=-3.05, p=0.001; decreased diastolic blood pressure at 12 months (end of study) β=-2.99, p= 0.001, Cohen's d 0.33; decreased HbA1c at 12 months (the end of the study) β=-0.51, p=0.01, Cohen's d 0.27; decreased HbA1c over time β=-0.35, p=0.05; decreased female waist circumference at 12 months (end of study) β=-2.31, p=0.001, Cohen's d 0.37; and decreased female waist circumference over time β=-2.13, p<0.001).

Conclusion and Next Steps

The evaluation was implemented as intended except for a deviation in the original timeline. Mercy conducted enrollment on a rolling basis between January 2016 and July 2016. Six-month follow-up began in June 2016 and ended in January 2017. Twelve-month follow-up began in November 2016 and ended in July 2017. This timeline represents a slightly longer timeline for enrollment and data than what was discussed in the SEP. A detailed timeline of the study can be found in Appendix A. Mercy did not have any changes to the budget or to their program team.

This evaluation study achieves a preliminary level of evidence. This evaluation study uses a QED design with primary and secondary comparison groups which were designed to mitigate major threats to internal validity such as selection bias. The program was implemented to fidelity, and the evaluation was conducted as intended. The study also meets the criteria for effective evidence because it demonstrates positive, significant findings for several exploratory outcomes. There were no negative intervention effects on confirmatory outcomes. Both the Duke General Health score and GAD-7 exploratory outcomes achieved small effect sizes (Cohen's d > 0.2) for the primary analysis comparing intervention participants with the primary comparison group. Three exploratory outcomes achieved small effect sizes for the secondary analysis (intervention compared to secondary comparison group).

The QED impact study demonstrated that the Sí Three integrated care model with faith-based behavioral health services had a significant association with physical and behavioral health improvements among intervention participants. After 12 months in the program, intervention participants were more likely than primary comparison group participants to experience significant improvements in depression over time, quality of life at 12 months and quality of life over time, and anxiety at 12 months and anxiety over time, when controlling for age, sex, and baseline characteristics. After 12 months in the program intervention participants were more likely than secondary comparison group participants to experience significant improvements in depression over time, quality of life over time, anxiety over time, diastolic blood pressure at 12 months, HbA1c at 12 months and over time, and female waist circumference at 12 months and female waist circumference over time, when controlling for age, sex, and baseline characteristics. Given the strength of the study design, there is considerable evidence that the intervention contributed to the improvements in health outcomes among participants. However, there was no significant change in obesity, waist circumference among males, or systolic blood pressure in the intervention group compared to either comparison group. Although similar improvements were observed in some health outcomes between the intervention and the two comparison groups, the findings are not consistent across all health outcomes. This is likely due to the different characteristics across the three groups.

Compared with the primary comparison group, intervention group participants were more likely to have higher scores on depression and anxiety measures and lower scores on quality of life measures at baseline assessment. Intervention and primary comparison group participants had similar demographics and baseline physical health measures, except for employment status. Intervention group participants were less likely to be employed than primary comparison group participants. According to Mercy staff, this reflects a difference in formal employment between the two groups where participants in either group would only self-assess as being employed if they had a full-time job.

In contrast, the secondary comparison group differed from the intervention group on many demographic measures and had poorer physical health measure scores at baseline. Statistical analyses procedures were used to control for these differences in analyses models. Despite the inherent differences between these two groups, results indicate that the Sí Three program improved behavioral and physical health outcomes among intervention participants.

This study contributes to our understanding of the impact of integrating behavioral health services with an option for faith-based counseling in the primary care setting. To our knowledge, this is one of the first studies examining the impact of an integrated care model featuring faith-based behavioral health counseling with a primarily Hispanic population. Lessons learned include the importance of leadership support, staff buy-in, and communication among staff. Staff rapport with participants and reworking clinic flow to meet staff and participant needs also were critical for program implementation and effectiveness.

In August 2017 Mercy implemented the Sí Three program throughout the clinic. The primary challenge to sustaining the model is securing financial resources to meet the needs of the low-income population.

INTRODUCTION

This final report reviews the methods implemented to evaluate Mercy Ministry of Laredo's program model according to the SEP, notes deviations and/or changes to the SEP, and describes final findings from the impact and implementation evaluations (including baseline data, six-month data, and twelve-month data). This report also provides a description of the reporting timeline discussed in the SEP and revised in Appendix A: Revised Project Timeline. The intended audience of this report is the Social Innovation Fund (SIF), although excerpts will also be used by Methodist Healthcare Ministries program staff and leadership and internal leadership and staff at Mercy Ministries of Laredo and Mercy Health.

Program Definition and Background

Residents of Webb County along the U.S.-Mexico border suffer from health disparities which stem from extreme poverty, lower levels of educational attainment, and inadequate access to basic health care. As identified in numerous region-specific assessments and reports, the scarcity of primary care and behavioral health service providers is a key factor influencing higher-than-average disease prevalence and unfavorable disease management. In addition, Laredo, TX (located in Webb County) and surrounding communities continue to serve increasing numbers of behavioral health cases with limited personnel and service-based resources to match the need. Reports estimate that the population to primary care provider ratio is 2,945:1—nearly double that for the state overall (1,893:1)—and the behavioral health provider ratio is similar, 3,500:1 (University of Wisconsin Population Health Institute, 2015). It is estimated that in the overall population of Webb County, 31% are obese, 27% are physically inactive, and 19% are classified as excessive drinkers (University of Wisconsin Population Health Institute, 2015). Estimates of the proportion of the population with diabetes vary; however, in 2016 at the Mercy clinic, 677 unduplicated patients were diabetic or pre-diabetic with 176 (26%) having an HbA1c above 7.0%.

Mercy is an integrated primary healthcare clinic located in Webb County, Texas, which provides healthcare and health education to some of the poorest neighborhoods and colonias in the U.S. The Texas Office of the Secretary of State defines a colonia as a residential area along the Texas-Mexico border lacking essential living infrastructures such as potable water and sewer systems, electricity, paved roads, and safe and sanitary housing. Mercy began implementation of the Sí Three program in the Laredo clinic in January 2016. The program expanded Mercy's efforts to integrate behavioral health, including optional faith-based behavioral health services, and physical health services. The Sí Three initiative aimed to improve behavioral health conditions (e.g., depression, anxiety, and addictive behavior) and chronic disease conditions (e.g., hypertension, obesity, and diabetes) through interventions that impact the physical, behavioral, and spiritual health of patients as well as overall quality of life. More specifically, Mercy improved work flow between primary care and behavioral health, increased communication between primary care and behavioral health and improved staff understanding of roles and integrated behavioral health culture. Also, to facilitate the integration of clinic services, a "care coordinator" served as a liaison between patients and clinic staff to promote Sí Three patient program participation with services and follow-ups. Additional personnel included a data entry clerk, nurse practitioners/navigators, a licensed professional counselor, an exercise coach, and a nurse educator.

Laredo, TX has a population of approximately 240,524 residents accounting for 94% of the population residing in Webb County (U.S. Census Bureau, 2013). Ninety-five percent of the population is

Hispanic/Latino of Mexican Descent (U.S. Census Bureau, 2013). In addition, as can best be captured by the U.S. Census, 35% of the population is foreign born, irrespective of citizenship status and 42% of the population is at or below the 200% federal poverty level (FPL). Because resources are limited in Laredo and, in particular among the population that uses Mercy's clinic, Mercy's Sí Three program combined multiple approaches to offer as many resources as possible to patients. The Sí Three program included education, exercise, nutrition, and both medical and faith-based behavioral health counseling. All Mercy patients received a care plan, which was updated regularly to meet patient needs. Mercy's EPIC Electronic Medical Record (EMR) system was used to document both physical and behavioral health visits and measures, whether a patient was in the intervention or primary comparison group, and whether the primary comparison group participants received any program-like services. The program did not deviate from the logic model as described in the August 2016 SIF evaluation plan (SEP) apart from removing the dietician role and replacing that role with a qualified nurse educator.

The enrollment target, based on depression as the confirmatory outcome, was 410 participants total across the intervention and primary comparison groups. The enrollment target was 366 participants for the secondary comparison group at Nuestra Clinica del Valle (NCDV). Mercy enrolled 411 participants in the intervention and primary comparison groups, while NCDV enrolled 366 participants in the secondary comparison group.

Overview of Prior Research

Based on prior research, Mercy's program model was assessed to have an incoming preliminary level of evidence. There is a preliminary level of evidence supporting the effectiveness of integration of behavioral health into primary care settings for improved patient health outcomes and cost effectiveness. The Sí Three: Integration of 3-D Health Services (Sí Three) model was based on a collaborative care model—and supported by evidence on the effectiveness of collaborative care models (Guide to Community Preventive Services, 2010). While the collaborative care model can take many different forms, it is defined as "a multicomponent, healthcare system-level intervention that uses case managers to link primary care providers, patients and mental health specialists." (Guide to Community Preventive Services, 2010). Case managers (navigators) are integral to the model and perform various functions, such as patient education and patient follow-up to track depression measures and adjustment of treatment plans. The Community Guide review found that collaborative care models produced more favorable results when compared to usual-care models for depression outcomes, including depression symptoms, adherence to treatment, response to treatment, remission/recovery, quality-of-life and functional status, and satisfaction with treatment. These results were supported for adults, older adults, women, men, Caucasian, African-American, Latino, and mixed-race populations in a diverse range of organizations and settings.

In addition, the Sí Three program was based on components of the integrated care model studied by Druss, Rohrbaugh, Levinson, & Rosenheck (2001). The Druss model which is an integrated model comprised of patient education and prevention, nurse practitioners, and increased interaction among the care team, found that patients in the integrated care model were significantly more likely to have received preventive care and had significantly greater improvement in health as compared to those who received the standard, stand-alone medical services. To these components, the Sí Three program added the option of faith-based/spiritual counseling to meet behavioral health needs. Worthington and colleagues conducted a meta-analysis of randomized control trials (RCTs) of religious/spiritual counseling and found greater improvements in psychological and spiritual outcomes as compared to

alternate secular therapies (Worthington et al., 2011). Worthington and colleagues reported that accommodating (adding to existing secular services) patient preferences for religious/spiritual services enhanced treatment outcomes and decreased premature termination of treatment by one-third. Further, a review by Koenig found that among eight RCTs, five showed that religious-based psychological interventions resulted in faster symptom improvement for depression compared to secular-based therapy or with control subjects (Koenig, 2012). Similarly, Koenig's review found that six-out-of-seven RCTs of religious interventions reduced anxiety levels more quickly than secular interventions or control subjects.

Program Components

Mercy's program theory of change was that if Mercy enhanced and strengthened their integrated practice (Heath et al., 2013), providers would gain an in-depth understanding of the roles of their team members and the culture of the clinic and partners. Patients, in turn, would express greater satisfaction with the team and would achieve their health goals. Moreover, healthcare providers would be more likely to refer patients to appropriate integrated services and track their participation and progress, facilitating patients' abilities to improve their physical and behavioral health. The logic model in Appendix B outlines the inputs, activities, outputs, and outcomes for the Mercy program. The Sí Three program combined components of the integrated care model studied by Druss et al. (2001) with faith-based care discussed by Worthington et al. (2011) The Druss model involves patient education and prevention, nurse practitioners, and increased interaction among the care team. Worthington found that religious/spiritual counseling resulted in greater improvements in psychological and spiritual outcomes as compared to alternate secular therapies (Worthington, Hook, Davis, & McDaniel, 2011). The only change in the program components from the SEP are changes in the program personnel. Instead of using a dietician for nutritional counseling, Mercy used a nurse educator. The dietician role has been removed from the logic model (Appendix B: Program Logic Model).

Inputs: The Mercy logic model had the following inputs which included a variety of existing and new internal program personnel as well as external program partners.

Internal personnel, who were involved in screening, referring, tracking, and providing in-house services, include:

- Principal Investigator – provided leadership for Sí Three program
- Program Manager – oversaw the Sí Three program and all personnel involved; was also one of the navigators/Nurse Practitioners (NPs)
- Contracted Evaluators (2) – consulted on data capture and reporting
- Data Entry – entered, maintained, and managed data for Sí Three program
- Navigators (3) – provided nursing care and referrals for patients
- Care Coordinator – followed up on patients' referrals and adherence to treatment plans
- Full-Time LPC – provided medical and faith-based counseling for behavioral health concerns
- Exercise Coach Part-Time – provided education and group fitness for patients and families
- Nurse Educator Part-Time – provided general education to patients regarding chronic disease, addictive behavior, and nutrition

External program partners, to whom intervention patients were potentially referred, included:

- Laredo Health Department – provided STD, HIV, TB or other services for mandatory reporting health conditions

- SCAN – provided child abuse or neglect services
- Border Region Behavioral Health Center – provided emergency and inpatient behavioral health services
- Faith-based Counselors – provided faith-based behavioral health services for participants who have mild or moderate anxiety and depression
- Other organizations aligned with patient needs, e.g. self-help groups for addictive behavior (AA, NA)

Activities: The activities section of the logic model provides an overview of Mercy programmatic activities at the patient and clinic levels.

- Clinic level: Providers and staff engaged in phone conferences, case conferences, face-to-face interactions, EPIC (EMR) training, and ongoing feedback and mentoring. All data were entered into the EMR, and the care coordinator tracked, monitored, and reminded patients of appointments. Providers and staff completed satisfaction surveys biannually.
- Patient level: Patients were assessed, diagnosed, and referred to primarily internal Mercy resources for physical and behavioral health needs, including the option of faith-based behavioral health services. Participating patients were reassessed quarterly for biological and behavioral health measures. Participating intervention group patients completed satisfaction surveys.

Outputs: Through implementation of program activities, outputs expected included:

- Recruit 205 participants into each arm of the study (intervention, primary, secondary comparison groups)
- Develop individual patient care plans
- Refer patients to appropriate in-house services and/or community resources
- Administer patient and staff satisfaction surveys
- Providers and staff understand and buy into integration model
- Providers use of standard measurement protocols
- Ongoing quality improvement among clinic staff

All activities and outputs identified in the logic model were evaluated as part of the implementation evaluation and were expected to influence the expected **short-**, **intermediate-**, and **long-term outcomes**. Short-term outcomes, intermediate, and long-term outcomes are presented in this final report.

Short-Term Outcomes: Short-term outcomes are the changes that are expected to occur during the five months of patients enrolling in the program and receiving Sí Three services. By working with the navigators, patients were expected to improve their knowledge of and skills for self-management and actively participate in their treatment plans. In addition, providers gained an in-depth understanding of the roles of their team members and the culture of the clinic and partners. The expected short-term outcomes are outlined below. These were assessed qualitatively in the study via focus groups and interviews.

- Clinic level: scheduling of patient follow-up appointments with appropriate internal or external community resources; entering data and tracking in EPIC, promoting, and monitoring patient use of services; improved communication across providers; adherence to program model.

- Patient level: improved patient knowledge of and skills for self-management; patients take a more active role in and can explain their treatment plans.

Intermediate Outcomes: Intermediate outcomes are the expected changes during the first 12 months of enrolling in the program and receiving Sí Three services. All intermediate outcomes are outlined below and were reported on during the study.

- Clinic level: improved workflow alignment across providers and services; improved clinic efficiency; increased rate of successful referrals, greater staff and provider satisfaction and data sharing.
- Patient level: patients participate in and are satisfied with referred resources; patients show improvements in waist circumference, BMI, HbA1c, blood pressure, depression, anxiety, addiction, and quality of life.

Long-Term Impact: Long-term outcomes are the changes that are expected to occur towards the end of the 12 months of the program or after the program ends. These long-term impacts are presented in the Implementation and Impact sections of this report.

- Clinic level: Clinic will improve integration of physical and behavioral health care.
- Patient level: Patients will meet physical and behavioral health targets and improve quality of life.

Overview of Impact Study

The impact evaluation used a non-randomized quasi-experimental design (QED) to evaluate the Sí Three program's impact. This study targeted a moderate level of evidence with a QED based on the incoming level of preliminary evidence. The Sí Three program combined components of the integrated care model studied by Druss et al. with faith-based care discussed by Worthington et al. The Druss model involves patient education and prevention, nurse practitioners, and increased interaction among the care team. Worthington found that religious/spiritual counseling resulted in greater improvements in psychological and spiritual outcomes as compared to alternate secular therapies (Worthington, Hook, Davis, & McDaniel, 2011).

The QED allowed for the identification and controlling for participant characteristics that may affect impact measures of interest. Two comparison groups were used for this study: a primary internal clinic comparison group of potentially similar patients who chose not to participate in Sí Three intervention activities and a secondary external comparison group, comprised of patients from Nuestra Clinica del Valle's Edcouch and Alton clinics. Analyses with the primary comparison group are considered the main study, while analyses with the secondary comparison group are included to enhance the external validity and generalizability of the primary comparison group results. The analyses between the intervention and secondary comparison group serve as sensitivity analyses aimed at an increased understanding of the intervention effects and how they may or may not differ when compared to a secondary comparison group under different conditions. The inclusion of patients from the Alton clinic is a change from the approved SEP which specified that the secondary comparison participants would only be recruited from the Edcouch clinic. Participants also were recruited from the Alton clinic to ensure an adequate sample size. Propensity score matching was found to not be appropriate to ensure patient equivalence in evaluating the program impact, which is a change from the approved SEP.

Research Questions

Mercy's evaluation plan included both implementation and impact research questions, as stated below. These questions have not changed since the approval of the SEP.

Implementation Questions

The following evaluation questions examined program implementation and patient and provider satisfaction. The final implementation evaluation included qualitative focus groups and interviews as well as assessment of quantitative implementation data.

1. Did the Sí Three program reach its intended target population?
2. What are the components of the Sí Three program and how do these components work "on the ground" at 6 and 12 months?
    a. Are these components different than what was planned? If so, why?
3. What level of integrated behavioral health did Mercy Ministries achieve as a result of implementing the Sí Three program?
    a. To what extent have providers and staff adopted the components of the Sí Three program at 6 and 12 months? What are the facilitators and barriers to adoption?
    b. To what extent do providers and staff buy in to the Sí Three program, and how has buy-in affected implementation?
4. To what extent did the comparison groups receive program-like components?
5. To what extent did the Mercy clinic implement the Si Three model with fidelity?
6. How satisfied are Sí Three patients with the services they have received? How satisfied are providers with the Sí Three program?

Additional implementation evaluation questions include the following:

7. What percent of patients who were seen by primary care providers for diabetes, obesity, and/or hypertension complete standardized assessments (depression, anxiety, addictive behavior, quality of life, and spiritual well-being) on their initial visit?
8. What percent of completed assessment results were recorded according to protocol? Were all staff able to implement standard measurement protocols?
9. What percent of patients with depression, anxiety, and addictive behavior were referred to the LPC or other behavioral health provider?
10. What percent of patients were assessed for depression, anxiety, quality of life, and addictive behavior on a semi-annual basis?
11. What was the show rate for all patients in the Sí Three intervention for IBH services? Does the show rate differ by the type of behavioral health resource the patient used?
12. What percent of referred patients could explain their physical and/or behavioral health treatment plans?

Impact Questions

The primary impact measures for Sí Three: Integration of 3-D Health Services were depression and improvement in BMI. Below are the confirmatory and exploratory research questions. The impact findings are presented later by Impact Question.

1. Do patients who participate in the Sí Three intervention experience improvements in depressive symptoms, as measured by PHQ-9, after 12 months compared to patients who do not participate? *This question is confirmatory.*
    a. In addition, do these improvements differ by type of behavioral health service received (medical/behavioral or faith-based services)? *This question is exploratory.*
2. Do patients who participate in the Sí Three intervention experience improvements in BMI after 12 months when compared to patients that do not participate in the intervention? *This question is confirmatory.*
3. Do patients who participate in the Sí Three intervention experience improvements in quality of life, as measured by the Duke Health Profile, after 12 months when compared to patients that do not participate in the intervention? *This question is exploratory.*
4. Do patients who participate in the Sí Three intervention experience improvements in anxiety symptoms, as measured by GAD-7, after 12 months compared to patients who do not participate? In addition, do these improvements differ by type of behavioral health service received (medical/behavioral or faith-based services)? *This question is exploratory.*
5. Do patients who participate in the Sí Three intervention experience improvements in addiction symptoms, as measured by CAGE-AID, after 12 months compared to patients who do not participate? *This question is exploratory.*
6. Do patients who participate in the Sí Three intervention experience improvements in blood pressure, when compared to patients that do not participate in the intervention? *This question is exploratory.*
7. Do patients with a history or diagnosis of diabetes who participate in the Sí Three intervention experience improvements in HbA1c measures after 12 months when compared to patients that do not participate in the intervention? *This question is exploratory.*

Contribution of the Study

The Sí Three evaluation contributes to the body of evidence regarding integrated behavioral health services in clinics serving predominantly low-income, Hispanic communities. The Sí Three evaluation targeted a moderate level of evidence by adapting components from two independent RCTs (Druss, Rohrbaugh, Levinson, & Rosenheck, 2001; Worthington, Hook, Davis, & McDaniel, 2011), ensured these models were culturally relevant and appropriate for their population (i.e., utilizing nurse practitioners and targeting an under-served, minority population), and examined the effects of the intervention through a quasi-experimental design (QED). It is recognized that a QED is not as rigorous as a randomized control trial (RCT); however, an RCT was not feasible because Mercy preferred to allow patients to choose group assignment due to policies of providing care to any patient who needs and requests it, as long as the clinic has the capacity to provide these services. It was possible, however, to recruit a primary comparison group of Mercy patients who chose not to participate in the intervention and identify another clinic in the region with patients who could serve as a secondary comparison group. Use of two comparison groups minimized threats to internal validity controlling for observed characteristics that could have affected impact measures of interest. The primary comparison group allowed for the examination of observed improvements in the intervention group as they relate to patients who received services at the same clinic but chose not to receive the intervention (patients with similar demographics and disease characteristics in the same setting). A secondary comparison group allowed for the examination of observed improvements in the intervention group as they relate

to patients who used a different clinic (factors related to being part of a different population in the same region).

Further, the primary comparison group minimized several threats to internal validity given that intervention and primary comparison group participants were more similar at baseline on demographic and outcome measures and the primary comparison group participants were patients in the same clinic as the intervention group. More specifically, the primary comparison group addressed the following threats to internal validity: regression to the mean, history, testing, John Henry, and expectancy effects. The secondary comparison group enhanced internal and external validity and served as a sensitivity analysis for the primary comparison group findings. More specifically, the secondary comparison group addresses the internal validity threats of selection bias and novelty. Also, the secondary comparison group addresses external threats to validity including applicability to other populations and applicability to other settings/locations.

The evaluation of the Sí Three program advances the evidence base related to integrated care models at clinics serving predominantly low-income, Hispanic communities. Also, the Sí Three evaluation examined the potential effect of faith-based behavioral health services for participants who met criteria for depression and/or anxiety. The evaluation study achieves a preliminary level of evidence given that an evidence-based intervention was adapted and evaluated using a QED with a primary and secondary comparison group. The use of a secondary comparison group (Nuestra Clinica del Valle's Edcouch and Alton Clinics) enhanced external validity or generalizability beyond the Mercy clinic. The use of a primary comparison group within the same clinic minimized threats to internal validity that were attributable to site characteristics. The program was implemented to fidelity, and the evaluation was conducted as intended. The study also meets the criteria for effective evidence. First, the study demonstrates positive, significant findings for several exploratory outcomes. As discussed in the Impact Study section of this report, positive and statistically significant results were demonstrated for the exploratory outcomes of quality of life and anxiety when comparing the intervention group to both the primary and secondary comparison groups. All statistically significant results achieved small effect sizes (Cohen's $d > 0.2$). Further, favorable and statistically significant results were found for the exploratory outcomes of blood pressure, HbA1c, and female waist circumference when comparing the intervention group to the secondary comparison group. There were no negative intervention effects on confirmatory or exploratory outcomes across all outcome analyses. Therefore, this study and its related findings are compelling and contribute to the field of the impact of an integrated model with faith-based behavioral health services.

SIF Evaluation Plan Updates

The evaluation plan was updated as follows.

- Recruitment was extended for four months to enroll sufficient numbers of participants in the primary comparison group. Six and 12-month follow-up assessments were collected as close as possible to the anniversary date of enrollment.
- The SEP logic model included a Part-Time LPC. The Part-Time LPC is not included in the final report as this provider did not provide services to the intervention group.
- The inclusion of patients from the NCDV Alton clinic is a change from the approved SEP which specified that the secondary comparison participants would only be recruited from the Edcouch clinic. Participants also were recruited from the Alton clinic to ensure an adequate sample size.
- Propensity score matching was found to not be appropriate to ensure patient equivalence in evaluating the program impact, which is a change from the approved SEP.

IMPLEMENTATION STUDY: STUDY APPROACH, METHODS, AND FINDINGS

Implementation Study Design

The implementation study aimed to understand how Mercy's program was implemented. As described in the SEP, two methods were used: 1) analysis of qualitative data collected through key informant interviews and focus groups, and 2) analysis of quantitative implementation data (e.g., patient visits, administrative data).

Qualitative Data Collection Methods and Analysis

The program's evaluator, Health Resources in Action (HRiA), conducted qualitative data collection at two-time points for the implementation study. Across the two-time points, a total of 8 staff members were interviewed, and 22 study participants were involved in focus groups.

For the mid-point interviews (September 2016), a total of 8 staff interviews were conducted in person. Mid-point interviews were intended to be conducted approximately 6 months after initial study enrollment. Due to logistical challenges, these interviews instead were conducted approximately 8 months after initial study enrollment, a deviation from the SEP. After the study concluded, 8 interviews were conducted (in mid-December 2017, approximately 4 months after the study ended). Interview participants included clinical providers (both primary and behavioral care) and other relevant clinical and nonclinical personnel.

The goal of the interviews was to assess program fidelity and understand in greater depth the context, facilitators, and challenges to program implementation. Program fidelity was assessed with clinic personnel interviewees by asking questions about program implementation from a clinic staff, program, and organizational level:

- ***Clinic staff level:*** The implementation evaluation measures programmatic implementation including clinic staff perceptions, attitudes and perceived barriers in care delivery for the target population. Clinic staff members were asked about their perceptions regarding the degree to which integration of primary care and behavioral health services has or has not been achieved at the mid- and end-point of the program, and their engagement with each other and aspects of the program.

- ***Program and organizational level:*** Interviews were also conducted with program managers and staff to obtain information about the operational level workflow and adherence to the original design of the program, and facilitators and barriers to implementation.

The interviews also aimed to capture information on clinical and administrative staff members' perceptions of barriers and facilitators to the program adoption, perceptions of program successes, challenges and opportunities for improvement, and perceived staff and patient satisfaction. Staff members were asked about their experiences with the program and perceptions of patient satisfaction both with the process of participating in the program as well as the outcomes. Appendix C and Appendix D present the semi-structured interview guides used to conduct the interviews at the mid-point and final data collection periods.

In addition to these semi-structured interviews, HRiA conducted two focus groups with intervention group participants after study implementation concluded (in mid-December, approximately 4 months after the study ended). The goal of the focus groups was to better understand the influence the program has had on participants' physical and behavioral health and wellbeing. Appendix E presents the semi-structured focus group guide used to conduct the focus groups at the final data collection period. Appendix F presents all implementation program components/activities, outputs, and outcomes that were measured using the qualitative data collection.

Table 1 describes participant demographics for the two focus groups. There were 22 intervention participants across the two focus groups, ranging from 8 to 14 participants per focus group. All participants resided in Webb County (100.0%), and self-identified as female (100.0%) and Hispanic (100.0%). A majority of participants were between the ages of 45 and 64 (62.0%), self-identified their race as White (57.1%), spoke Spanish as a primary language (86.4%), and had less than a high school diploma (63.6%). Ninety-five percent of participants did not have health insurance.

Table 1. Mercy Pre-Focus Group Demographics Survey

Measure	Mercy (n=22) n	%
County		
Webb	21	100.0
Missing	*1*	--
Sex		
Female	22	100.0
Missing	--	--
Age		
≤ 34	4	19.1
35-44	3	14.3
45-54	9	42.9
55-64	4	19.1
65+	1	4.8
Missing	*1*	--
Ethnicity		
Hispanic/Latino	22	100.0
Missing	--	--
Primary Language		
Spanish	19	86.4
English and Spanish	3	13.6
Missing	--	--
Education		
Less than a high school diploma	14	63.6
High school degree or equivalent (e.g., GED)	4	18.2
Some college, junior college, or vocational school	4	18.2
College degree or more	0	0.0
Missing	--	--
Health Insurance		
I don't have health insurance	19	95.0

Medicaid, Medical Assistance*	1	5.0
Missing	*2*	--

*Mercy Ministries only serves persons who do not have insurance. Given that the Laredo community is very dynamic and move in and out of eligibility for insurance, it is possible that an intervention participant became eligible for insurance after completing the intervention.

All interviews and focus groups were conducted by experienced and trained qualitative researchers from the HRiA evaluation team. A lead moderator conducted the interviews and focus groups and a research assistant took detailed notes. The interviews were conducted in English, and the focus groups were conducted in Spanish to match the primary language spoken at home by the majority of participants.

All interviews and focus groups were recorded digitally and transcribed. Detailed notes from mid-point interviews were coded by one coder using NVivo software. The mid-point interviews were analyzed with this approach due to the importance of expediency to complete the interim report and to provide findings to the subgrantee quickly for continuous quality improvement. Mid-point data were not re-coded for the summative analysis, but themes from the mid-point and summative data collection were synthesized together, and findings were summarized in narrative descriptions organized by theme with illustrative quotes. For the summative interviews and focus groups, two trained team members – who did not conduct interviews or focus groups - initially reviewed transcripts to develop a mutually-agreed upon codebook using a grounded theory approach. They then independently coded each transcript for themes using NVivo qualitative data analysis software (NVivo qualitative data analysis software; QSR International Pty Ltd. Version 11) and met to discuss concordance and discordance between their coding schemes. Differences were reconciled through discussion until a consensus on the first-level of coding was reached (average kappa=0.96). Differences were reconciled through discussion, and themes were identified by discussion frequency and intensity. If qualitative findings changed from mid-point data collection to summative data collection, it is noted.

Quantitative Data Collection Methods and Analysis

Implementation data of patient participation in the Sí Three program were analyzed. These mainly comprised of de-identified patient records from Mercy's Electronic Medical Record (EMR) system that included information on intervention and primary comparison group participants' behavioral health and primary care visits. Descriptive statistics on these services are provided in this section, including the mean, median, and range of number of completed and missed visits related to behavioral health and primary care for both groups. This information provides insight into fidelity and dose of the intervention.

Implementation Study Findings

The following presents the implementation study findings by research question as presented in the SEP.

Question 1. Did the program reach its intended target population?

All patients who met eligibility criteria and voluntarily consented to participate in the Sí Three program were offered the opportunity to participate in the intervention research study at the time of baseline data collection.

As described in the SEP, all Mercy clinic adult patients were eligible for the intervention study if any one or more of the following criteria were met:

- PHQ-9 ≥ 5
- GAD-7 ≥ 5
- CAGE-AID ≥ 2
- Waist circumference ≥ 40 in men and ≥ 35 in women
- BMI ≥ 30
- (Hypertension) Blood Pressure ≥ 140/90
- (Diabetes) A1C ≥ 7.0%

Mercy enrolled 411 participants into the intervention (n = 207) and primary comparison groups (n = 203). Participants were primarily female (87.0%) and Hispanic (99.8%) whose primary language was Spanish (87.8%). The mean age of participants at enrollment was 44 years, a majority (51.8%) reported not being employed, and the majority (53%) reported being married. All participants reported Webb County as their residence. All participants met the study eligibility criteria; therefore, the program reached the intended audience. The demographic characteristics and prevalence of the study outcomes among the Mercy Clinic population and enrolled intervention sample are provided in **Table 2.**

Table 2. Demographic Characteristics and Prevalence of Outcomes of Mercy Clinic Population and Intervention Group Participants[a]

	Mercy Clinic	Intervention Group
Total Adult Patient Count	1,864	207
Age		
Mean age	43	43.8
% 18-34	26.1	18.3
% 35 – 64	71.1	79.7
% > 65	2.7	1.9
Race/Ethnicity		
% Hispanic	98.4	99.5
% Non-Hispanic	1.6	.5
Gender		
% Female	83.4	87
% Male	16.6	13
Blood Pressure		
% of Patients with Elevated blood pressure (>150/90)	2.55	13
HbA1c		
% of Patients with Elevated A1c (Hb A1c > 7.0)	39.6	40.4
Body Mass Index		
% Obese	50.7	66.2

[a] Tests to identify statistically significant differences were not performed between the Mercy clinic population and intervention participant group because individual data on the Mercy clinic population were not available.

Question 2. What are the components of Mercy's Sí Three program and how do these components work "on the ground" at 6 and 12 months?

Question 2a. Are these components different than what was planned, and why are they different?

Overall, Mercy implemented the Sí Three program as planned. Patients who met eligibility criteria and consented to participate in the intervention research study were offered appropriate IBH services according to needs identified through physical, behavioral, and spiritual health assessments. The nurse practitioner was the referring provider for primary and external services and served as the patient navigator for referred services. The care coordinator scheduled follow-up appointments and reviewed patient attendance at identified services (physical and medical/faith-based behavioral) and contacted patients who did not attend referred services to increase patient participation. Services included 1) behavioral and community health services to address depression, anxiety, and/or addictive behavior; 2) primary care services to manage diabetes, obesity, and hypertension; and 3) health and nutrition education and exercise classes. A licensed professional counselor provided either traditional or faith-based behavioral health services that aligned with a patient's score on the spirituality assessment and assessment by the behavioral health consultant.

How Components Work "On the Ground"

Staff interviews delved deeper into how the program was being implemented. When asked about how primary care and behavioral health services were coordinated and connected, interview participants highlighted communication practices, data systems, workflows, and faith based behavioral health services as the key components of Mercy's Sí Three program. These were also mentioned during mid-point interviews. The physical clinic space was also discussed extensively, and will be explored below in the section entitled, "Implementation as Planned."

Communication
According to interviewees at the mid-point and summative interviews, communication was a core component of Mercy's integration strategy. Both in-person and electronic communication strategies were mentioned as essential elements of clinic integration. Interdisciplinary committee meetings, weekly team meetings, frequent emails, and impromptu in-person huddles among staff were described as increasing interaction and collaboration among program staff. Weekly meetings and frequent emails among primary care and behavioral health staff, explained interviewees, allowed the staff space and time to discuss patients holistically (patient visits, medications, and care plans) as well as address the clinic's integrated systems.

Data Systems
In addition to communication practices discussed above, the primary form of electronic communication for Mercy's Sí Three program was its data system, EPIC. Interview participants noted that, as part of its Sí Three integration efforts, Mercy updated its electronic medical record (EMR) to better integrate physical and behavioral health data. Clinical staff interviewees described how they were already well-versed in the EMR, but the system updates allowed for sharing of data between primary care and behavioral health staff. As one primary care clinical staff noted, *"We're blessed to have a system that's been in place before [Sí Three]. Most of my co-workers were already trained in the system so were very fluent in doing the charting, the documentation. It was just a matter of primary care providers looking into the behavioral health part and vice versa.* According to interviewees, this access to data and other

providers helped improve integration and coordination of care, particularly through the sharing of care plans. As one primary care clinical staff interviewee summarized, *"It's been a big, big help. EPIC has been fundamental in the success of the program."*

Workflow

Workflow, or how patients and clinic staff move within the clinical space, was seen as a key component of integration and closely tied to Mercy's communication practices and physical clinic space. These are discussed further in the adoption facilitators section. From the clinic staff perspective, clinic operations were adapted to implement the IBH model. Interviewees described how workflows were modified continuously to enhance internal referrals and reinforce clinic staff communication. For example, a clinic staff interviewee shared, *"I assess the patient and I identify what the patient needs on my referral form. The patient is then routed to that referral and the provider knows what I've covered and what the patient needs."*

Spirituality

While not discussed extensively in the mid-point interviews, according to interview and focus group participants in the summative data collection, the intentional incorporation of the opportunity for spiritual care as an approach to behavioral health care was foundational to Mercy's Sí Three program. Interviewees described how this approach was implemented through the provision of holistic care by one provider with training in spiritual, physical, and behavioral health. As one administrative staff interviewee summarized, *"we had a psychologist who was also a pastoral counselor and also a hospital chaplain. She has credentials in all three areas."* Focus group participants also recognized that the opportunity for spiritual care was at the core of the Sí Three program. *"That is the purpose [of the program], physically and spiritually complete treatment,"* a focus group participant explained.

Implementation as Planned

Overall, Mercy implemented the Sí Three program as planned. The most significant alteration to the plan was changing clinic workflow. As described above, Mercy made several changes to clinic workflow, specifically related to its physical space, which resulted in moving staff locations in the clinic. Clinic staff interviewees shared that throughout the program, staff shifted offices and changed the order in which patients saw providers to better meet the needs of patients. *"If we noticed it was not functioning with the workflow, we were able to identify what was not working and go back to redesign the workflow,"* described one primary care clinical staff person. "*We rearranged [the office] so all of the clinical side is together now, so it's more conducive to having instant consults and less running around for the patient,"* shared another primary care clinical interviewee.

Aside from these changes in clinical space and workflow, several other minor implementation revisions were made within the first three months of program implementation. For example, reinforcing what was noted in mid-point interviews, telepsychiatry was intended to be offered, but interviewees stated there was not a need for those services. Further, staff explained that they had planned to make external referrals for spiritual care but did not provide these referrals because of the "*unique, comprehensive credentials of the Licensed Professional Counselor*," who could provide services internally instead of having to refer patients out. Finally, due to challenges in hiring a qualified dietician, a qualified nurse educator provided nutritional counseling services.

Otherwise, according to interviewees during the mid-point and summative evaluations, Mercy implemented their IBH program with strong fidelity and this fidelity was observed within the first three months of implementation. Summarizing, an administrative staff interviewee said, *"I believe there was a lot of success in meeting the goals that were set out at the beginning of the program."*

Question 3. What level of integrated behavioral health did Mercy achieve as a result of implementing the Sí Three program?

Question 3a. To what extent have providers and program staff adopted the components of Mercy's Sí Three program at 6 and 12 months? What are the facilitators and barriers to adoption?

Implementation of Integrated Behavioral Health

According to the World Health Organization (2008), behavioral health integration encompasses the management and delivery of health services so that individuals receive a continuum of preventive and restorative mental health and addiction services, according to their needs over time, and across different levels of the health system. Quality integrated care requires a well-functioning, well-organized primary care practice as well as key behaviors at the organizational, practice, interpersonal, and individual clinician levels (Cohen et al. 2015).

There are many ways to assess how components of IBH are practiced in different settings. The Advancing Integrated Mental Health Solutions (AIMS) IBH checklist was developed by IBH experts to assess five core principles of collaborative care (AIMS Center, 2011). These principles include: (1) patient-centered care, (2) population-based care, (3) measurement-based treatment to target, (4) evidence-based care, and (5) accountable care. The checklist details core components and tasks for each of these principles that are self-assessed on a scale of "None," "Some," or "Most/All." Appendix K: Patient-Centered Integrated Behavioral Health Care Checklist presents the core descriptions of the Patient-Centered Integrated Behavioral Health Care Principles and Tasks Checklist as defined by the AIMS Center.

Mercy completed the AIMS IBH checklist in December 2015 (pre-intervention implementation) and June 2018 (post-intervention implementation). **Table 4** present Mercy's data from these assessments. Mercy reported improving IBH principles on nearly all measures with the exception of two of the Systematic Psychiatric Case Review and Consultation measures. Although Mercy had initially planned for psychiatric consultation to be part of the Sí Three program, Mercy determined that psychiatric services through telepsychiatry were not needed for its patient population. Pre-intervention Mercy assessed that it provided the checklist principles to none or some of its patients. Post-intervention Mercy assessed that it provided nearly all of the principles to most or all of its patients. Exceptions included the Accountable Care principle (some patients) and the previously mentioned Psychiatric principles (none of the patients). Regarding the Accountable Care principle, Mercy may only provide this principle to some patients because it serves patients without insurance and thus principles regarding reimbursement are not applicable.

Table 3. IBH Checklist Baseline to 12 months: Core Principles

We apply this principle in the care of (none, some, most/all) of our patients.	None	Some	Most/All
Patient-Centered Care Primary care and behavioral health providers collaborate effectively using shared care plans.		•	✓
Population-Based Care Care team shares a defined group of patients tracked in a registry. Practices track and reach out to patients who are not improving, and mental health specialists provide caseload-focused consultation, not just ad-hoc advice.	•		✓
Measurement-Based Treatment to Target Each patient's treatment plan clearly articulates personal goals and clinical outcomes that are routinely measured. Treatments are adjusted if patients are not improving as expected.		•	✓
Evidence-Based Care Patients are offered treatments for which there is credible research evidence to support their efficacy in treating the target condition.		•	✓
Accountable Care Providers are accountable and reimbursed for quality care and outcomes.	•	✓	

• Response at baseline ✓ Response at 12 months

Table 4. IBH Checklist Baseline to 12 months: Core Components and Tasks

We apply this principle in the care of (none, some, most/all) our patients.	None	Some	Most/All
Patient Identification and Diagnosis			
Screen for behavioral health problems using valid instruments		•	✓
Diagnose behavioral health problems and related conditions		•	✓
Use valid measurement tools to assess and document baseline symptom severity		•	✓
Engagement in Integrated Care Program			
Introduce collaborative care team and engage patient in integrated care program		•	✓
Initiate patient tracking in population-based registry	•		✓
Evidence-Based Treatment			
Develop and regularly update a biopsychosocial treatment plan		•	✓

We apply this principle in the care of (none, some, most/all) our patients.			
	None	Some	Most/All
Provide patient and family education about symptoms, treatments, and self-management skills		•	✓
Provide evidence-based counseling (e.g., Motivational Interviewing, Behavioral Activation)		•	✓
Provide evidence-based psychotherapy (e.g., Problem Solving Treatment, Cognitive Behavior Therapy, Interpersonal Therapy)		•	✓
Prescribe and manage psychotropic medications as clinically indicated		•	✓
Change or adjust treatments if patients do not meet treatment targets		•	✓
Systematic Follow-up, Treatment Adjustment, and Relapse Prevention			
Use population-based registry to systematically follow all patients	•		✓
Proactively reach out to patients who do not follow-up		•	✓
Monitor treatment response at each contact with valid outcome measures		•	✓
Monitor treatment side effects and complications		•	✓
Identify patients who are not improving to target them for psychiatric consultation and treatment adjustment		•	✓
Create and support relapse prevention plan when patients are substantially improved		•	✓
Communication and Care Coordination			
Coordinate and facilitate effective communication among providers		•	✓
Engage and support family and significant others as clinically appropriate		•	✓
Facilitate and track referrals to specialty care, social services, and community-based resources		•	✓
Systematic Psychiatric Case Review and Consultation			
Conduct regular (e.g., weekly) psychiatric caseload review on patients who are not improving	• ✓		
Provide specific recommendations for additional diagnostic work-up, treatment changes, or referrals	•		✓
Provide psychiatric assessments for challenging patients in-person or via telemedicine	• ✓		
Program Oversight and Quality Improvement			
Provide administrative support and supervision for program		•	✓

We apply this principle in the care of (none, some, most/all) our patients.			
	None	Some	Most/All
Provide clinical support and supervision for program		•	✓
Routinely examine provider- and program-level outcomes (e.g., clinical outcomes, quality of care, patient satisfaction) and use this information for quality improvement	•		✓

• Response at baseline ✓ Response at 12 months

Program Adoption
Interview and focus group participants were asked what facilitated or hindered program adoption, including patient participation in the program. The following section presents adoption facilitators and barriers identified by these participants.

Adoption Facilitators

During summative interviews and focus group discussions, adoption facilitators included increased communication, adapted data systems, the physical space of the clinic, and hiring and retention. Increased communication and coordination were also identified during the mid-point interviews as adoption facilitators.

Communication
Communication was the most frequently identified facilitator of program adoption from staff. Clinic staff mentioned numerous ways in which communication facilitated Sí Three program adoption. As described previously, the interdisciplinary committee meetings were highlighted as bringing together behavioral health and primary care staff to share information and develop care plans for patients. Interviewees also expressed how emailing information provided easy, quick ways to touch base with other staff, allowing them to make efficient adjustments to program implementation, such as rerouting of patients. Other staff described how open communication among staff, specifically among nurses, fostered teamwork across program staff.

Data Systems
Interviewees also highlighted how Mercy's data system (EPIC) facilitated program adoption. In addition to the EMR being a key component of how the program worked, interviewees described how staff experience using EPIC during years prior to the Sí Three program allowed them to take full advantage of the revised data system, such as giving primary care and behavioral health providers access to patient data, medication lists, provider notes and referrals. Several administrative and clinical staff interviewees also appreciated the technical support from the Mercy Health [system] office in helping tailor the clinic's data system for the Sí Three program, allowing the clinic to "*do pretty much whatever we want to do with our data*," explained one administrative staff interviewee.

Clinic or Physical Space
Interview participants highlighted that the physical clinic space at Mercy supported adoption of the Sí Three program. A clinical staff interviewee shared, *"I think one of the major advantages to be near primary care providers is they see me more, they ask more questions. I think it's much better."* Other clinical staff also noted the convenience of being *"next door"* to each other, which facilitated internal

communication and referrals. Another clinical staff highlighted the advantages of provider office proximity for patients stating, *"Now they see their [primary care] provider, then go next door to see their behavioral health [provider]."*

Staffing: Hiring and Retention
Interviewees identified staff hiring and retention as integral to adopting Mercy's Sí Three program. Despite some early hiring challenges identified in the mid-point interviews, such as the inability to hire a dietician, the planned additions of an exercise coach and an LPC/pastoral counselor were viewed as key to implementation of the program. Interviewees shared that as implementation progressed there was *"great cooperation by the entire staff to make the project a success" (administrative staff interviewee).*

Adoption Barriers

At the mid-point, interviewees noted several challenges to program adoption, including space limitations and layout, new data collection for the evaluation study adding time to patient visits, difficulty hiring a dietician and a referral psychiatrist, and changing clinic work flow to reduce wait times. Clinic space limitations and staffing were mentioned during summative interviews as well as communication challenges.

Clinic or Physical Space
As discussed previously, there were numerous adjustments to the clinic space and patient flow throughout the life of the Sí Three program implementation. In mid-point and summative interviews, several staff mentioned that the clinic space presented challenges in terms of layout and amount of space, as the Mercy clinic was not originally designed for IBH. Interviewees noted that it would have been helpful to have more space that allowed for closer interaction of staff providing behavioral health and primary care services. As the program concluded, the existing space and staff were rearranged to cluster clinical (behavioral and physical) staff together in one part of the building. According to one administrative staff interviewee, *"We do have limited space and limited funds to do something extravagant with the space."*

Communication
While most interviewees at both mid-point and summative interviews emphasized staff communication as a facilitator of adoption, several mentioned that there were a number of initial communication challenges to implementing the Sí Three program. For example, a few interviewees suggested that communication was not always clear regarding roles and responsibilities. As one staff interviewee explained, "*Communicating the roles and communicating what the patient needs between providers, that's been a very big challenge to me.*" Another clinical staff person shared that team huddles were less frequent during the second half of program implementation period, which decreased provider opportunities to communicate about patients.

Hiring and Staffing
Although focus group and interview participants emphasized the quality of the staff that Mercy was able to hire, the difficulty in finding qualified applicants for the dietician position was noted as a particular challenge during mid-point and summative interviews. *"To have a dietician or nutritionist as part of the program, I feel would have benefited clients,"* explained one staff interviewee. Interviewees noted the nurse educator stepped up to provide patient education regarding diet and nutrition; these interviewees, however, thought that having a dietician provide this education may have been more

beneficial to participants. As program implementation moved forward, summative evaluation interviewees noted that the staffing challenge had shifted to finding other providers, specifically hiring a behavioral health nurse practitioner, to meet patient needs. Note, although the need for this position has been identified, according to Mercy staff it is unlikely to occur in the near future as there are no other behavioral health nurse practitioners in Laredo.

Participant Facilitators

In addition to facilitators experienced by staff adopting the Sí Three program, focus group and interview participants were also asked to reflect on factors that facilitated patient participation in the program. Participant facilitators mentioned included cost, program staff flexibility, patient-staff relationships, and transportation.

Cost

Patient focus group participants recognized the Sí Three program, and Mercy's services in general, for being very affordable. Behavioral health and primary care services reportedly cost little to no money, as well as complimentary supplies at the clinic, such as blood pressure cuffs, eyeglasses, and medications. According to focus group participants, this low cost allowed patients to seek and receive care more readily than they could outside of Mercy. As one patient shared, *"They [the clinic] help you a lot on economic matters when you don't have money."*

Clinic Flexibility

The flexibility of the Mercy clinic patient care model was noted as assisting with patient participation in the Sí Three program. According to focus group participants, this flexibility was most pronounced in terms of advanced and on-the-spot scheduling, specifically for primary care. Further, as one patient commented, *"they [providers] are also good about sending you to another doctor and they make sure they give you an appointment with the dentist, the eye doctor, or other specialists."* [Note: It is Mercy's protocol to refer all patients to any needed service and to provide specialty referrals outside of the clinic.] In addition to flexibility of scheduling clinical visits, focus group participants shared that the clinic has extended hours to accommodate *"people who work and can't come during those hours."* Focus group participants also described how Mercy's staff demonstrated flexibility through the use of community health workers, or *promotoras*, who would conduct home visits. Extended hours and use of promotoras are not exclusive to the Sí Three program but were noted as facilitating patient participation in the program.

Patient-Staff Relationships

Interview and focus group participants described how patient relationships with staff and each other encouraged participation in the Sí Three program. As one interviewee remarked, *"I think what's worked well is the rapport with our clients and the social worker hired for the project."* It was also noted that the clinic's successful enrollment and retention of patients was due in large part to the relationships that staff have with patients. Patient focus group participants also highlighted that the relationships they have built at Mercy supported their participation, with one patient summarizing, "*I like it here [at Mercy]. It's like my second home.*"

Transportation
Mercy operates a mobile clinic, which provides services to 14 outreach sites. Additionally, some Mercy staff provide transportation services to some clients who have no other means to get to an appointment; this was generally seen as facilitating patient participation in the program.
One patient explained, *"So, they pick you up in the morning at whatever time, a half hour before your appointment."* And another patient continued, *"And they come with us, the van, and they leave us. Before, where I live, I had to get four buses to come here."* Mercy staff transportation services allowed some patients to get to the clinic in a timely and inexpensive way and reduced the *"stress of not knowing how to get to my appointment tomorrow."* This transportation service is provided on an as-needed basis.

Participant Barriers

In addition to barriers experienced by staff and providers adopting the Sí Texas program, focus group and interview participants were also asked to reflect specifically on barriers that patients faced while participating in the program. Barriers discussed included cost, the socio-political environment, and transportation.

Cost
While most focus group participants spoke of the minimal costs to participate in the Sí Three program at Mercy as facilitating participation, several administrative staff interviewees suggested that sliding scale copays prevented some patients from coming to their primary care visits.

Socio-Political Environment
Focus group participants and interviewees alike shared that the socio-political environment was a barrier for patients traveling to Mercy for services. Focus group participants and interviewees reported a perceived decline in patient visits to the clinic, particularly starting in March 2017. This was an issue that focus group participants and interviewees reported was experienced across the entire Mercy patient population, not only for Sí Three program participants. Although this reported barrier was noted as being particularly acute towards the end of the program period, it may have affected retention of participants in the Sí Three program study.

Transportation
While the transportation practices of Mercy staff were noted as a facilitator of program participation for some patients, many focus group participants described the difficulties of accessing and affording transportation to participate in the Sí Three program and its services, given the increased number of visits and classes associated with program participation. According to interview and focus group participants, many patients live far away from the clinic, and have limited public and private transportation options and limited income to spend on transportation. As one primary care clinical staff explained, "*If they use the local metro, it costs. It's not cheap for them to move around. And then we have individuals living at a distance. A lot of our patients live way out in the colonias. They don't have reliable transportation."*

Question 3b. To what extent do providers and staff buy in to the program, and how has buy-in affected implementation?

Clinical and administrative staff were asked about their support and buy-in for the Sí Three program as well as their perception of their colleagues' buy-in. Interviewees spoke about the culture of the clinic, as well as buy-in and satisfaction of frontline clinic staff as well as leadership and administration.

Clinic Culture

In general, interviewees perceived the clinic culture to be a supportive environment for adoption of the Sí Three program. Key to the cooperative clinic culture has been intentionally *"including everyone and recognizing all of the work that it takes for every individual to make the project work,"* shared one administrative staff. It was also noted that the staff take time to celebrate birthdays and program accomplishments, to enhance morale and "*make everyone feel appreciated and a part of it [the Si Three program],"* according to a primary care clinical staff interviewee. Several interviewees spoke of the strong teamwork and commitment of the clinic to addressing patients' needs for integrated behavioral health. Summarizing, one clinic staff interviewee shared, *"It's a team effort. It's leadership, the providers, everyone. They are very committed, and they love to do this because they see the need."*

Staff

Mercy staff interviewees expressed overall satisfaction with the program, citing increased access to services for their patients as well as initial positive health outcomes. "*Especially the nurses, they were happy to integrate the mental health with the medical,*" explained one staff interviewee. According to several staff interviewees during mid-point and summative interviews, however, there was some dissatisfaction among staff with the study design because the comparison group patients did not receive the same services and supports as the intervention group. As one staff interviewee shared, *"I felt kind of bad not being able to offer anything. That made the [primary comparison group] patient feel less of a patient because the intervention group was getting more."* Additionally, a few staff interviewees noted that there was a learning curve among the providers who did not have previous experience with an integrated model of care. While a few interview participants spoke of working through some initial tension around roles, responsibilities and expectations for integration, one staff interviewee explained, "*I wouldn't call it resistance; they just didn't know what it was because they had never done anything like this before."* Despite these concerns, the overall sentiment among staff was one of satisfaction as the study concluded.

Leadership and Administration

According to staff interviewees, Mercy has made financial and infrastructure investments to move forward with integrated care and Mercy's leadership, in particular the President/CEO, was seen as being very supportive of integrated care. Emphasizing this, one administrative staff interviewee stated, *"I don't know how many other [organizational] presidents go to all of those [Sí Texas] meetings, but she is very much involved in this [program]."*

Question 4. To what extent did the comparison groups receive program-like components?

Primary comparison group participants were monitored per usual Mercy clinic protocols on impact measures. To be consistent with Mercy's mission and values, all patients received needed care in the clinic or were referred to community resources to address care needs. Comparison group participants had access to healthcare providers, part-time behavioral health counselors, health education sessions with a BSN educator, and community exercise classes primarily through referrals. In contrast to the intervention participants, these services were not provided through a coordinated care model and there was limited availability of some of these services in the clinic.

"I have to say this, even if they're not in the intervention group, if somebody exhibits any symptomology of anxiety or depression, we score them with a PHQ-9 and GAD-7, and that's just professional. Like I'm going to do that for anybody," explained a primary care clinical interviewee.

Primary comparison group participants referred for BH services were scheduled for a co-located, but not integrated, visit with a part-time LPC who did not offer the option of faith-based behavioral health counseling. Co-located services occurred in the same building with little to no discussion between providers about patient needs and provision of behavioral and physical health services were not coordinated. In contrast, integrated visits involved discussion of patient needs among providers at the time of service and behavioral and physical health services were coordinated to occur at the same visit. Educational referrals for the primary comparison group patients were seen by a nurse educator at a visit that was not integrated or coordinated with other services.

The secondary comparison participants came from NCDV from either NCDV's Edcouch or Alton Clinic. Edcouch Clinic usual care for behavioral health entailed referring patients for behavioral health services when the patient had a PHQ-9 score ≥ 10 or when the PCP observed behavioral health distress, to a NCDV clinic with a behavioral health provider. The nurse called the Behavioral Health Care Manager to schedule an appointment for the patient to be seen by an LPC at the NCDV San Juan or Mercedes clinic dependent on the patient's discretion. In addition, if desired, the patient could ask the care manager to set up an appointment to be seen by the LPC. If the patient showed suicidal ideation with a plan to hurt him or herself or others, a call was made to the mental health authority, Tropical Texas Behavioral Health. Usual care regarding primary care involved a visit with the medical doctor.

Alton Clinic usual care for behavioral health entailed referring patients with a PHQ-9 score ≥ 10 or when the PCP observed behavioral health distress to outside behavioral health services or with an in-clinic visit with an LPC. The LPC was only at the Alton Clinic for one day every two weeks. Patients could call the care manager to set up an appointment to be seen by the LPC if she was there. If a patient did not have an appointment scheduled on the day the LPC was at the Alton Clinic, the patient was given the option to set up an appointment at the NCDV San Juan clinic or NCDV Mercedes clinic where he or she would be seen by a LPC. If the patient showed suicidal ideation with a plan to hurt him or herself or others, a call was made to the mental health authority, Tropical Texas Behavioral Health. Usual care regarding primary care involves a visit with the medical doctor.

During the study period, NCDV began offering patients at both clinics nutrition education information.

Question 5. To what extent did Mercy implement the Sí Three program with fidelity?

Mercy implemented the Sí Three program with strong fidelity. The program was based on the following collaborative care model components: assessing, diagnosing, and providing internal referrals for intervention participant physical and behavioral health needs, including the option of faith-based behavioral health services. Before, during, and after program implementation, clinic staff made changes to clinic workflow, specifically related to its physical space and consequent moving of staff in the clinic. Staff interviewees shared that throughout the program, staff shifted offices and changed the order in which patients saw providers to better meet the needs of patients. *"If we noticed it was not functioning with the workflow, we were able to identify what was not working and go back to redesign the workflow,"* described one staff interviewee. "*We rearranged [the office] so all of the clinical side is together now, so it's more conducive to having instant consults and less running around for the patient,"* shared another staff interviewee.

Aside from these changes in clinic space and workflow, several other minor revisions were made after the program began. For example, reinforcing what was noted in mid-point interviews, telepsychiatry was intended to be offered, but summative interviewees stated there was not a need for those services. Further, staff explained that they had planned to make external referrals for spiritual care but did not end up implementing them because of the "*unique, comprehensive credentials of the Licensed Professional Counselor*," who could provide services internally instead of having to refer patients out.

Otherwise, according to interviewees during the mid-point and summative evaluations, Mercy implemented their IBH program with strong fidelity. Summarizing, an administrative staff interviewee said, *"I believe there was a lot of success in meeting the goals that were set out at the beginning of the program."*

To support the qualitative data on program implementation, we summarize program implementation fidelity with the quantitative data from Mercy's EMR. All 207 patients in the intervention group were offered and received services from one of three nurse practitioners at the first visit (baseline) and at six- and twelve-month follow-up assessments with the exception of patients who were lost to follow-up or withdrawn from the study.

A component of Mercy's implementation was for intervention participants to come to the clinic quarterly, to monitor physical and behavioral health measures, in addition to the baseline, 6-month, and 12-month appointments for study assessments. Of those participants who completed a 12-month follow-up (n=142), 57.0% met this requirement (n=81). Of the 61 participants who missed at least one appointment (3-, 6-, or 9-month), a large majority (82%) missed only 1 visit. A total of 24 participants had 3 and 6-month appointments, but no 9-month appointment. A third (n=21) had 6- and 9-month appointments, but no 3-month appointment. Only 5 participants had 3 and 9-month appointments but missed their 6-month visit.

Table 5 summarizes the referrals made and closed among intervention and primary comparison group participants. Mercy uses a referral system to track the number of referrals made and completed. A referral is considered closed or completed when the provider to whom the patient is referred enters visit completion information in the EMR. Six hundred three referrals were made for the intervention group compared with 41 for the primary comparison group. Among intervention group participants, 98% of these referrals were closed compared with 85% of referrals for the primary comparison group.

The average number of referrals was 2 per patient in the intervention group compared with 1 referral per patient in the primary comparison group. The minimum number of referrals per patient in the intervention group was 1 and the maximum was 5 referrals. The primary comparison group had a maximum of 2 referrals per patient.

The large number of referrals made and completed for the intervention group reflects implementation of the intervention to fidelity. Services to which intervention patients were referred were provided at the Mercy clinic as integrated services. For example, intervention patients referred for behavioral health care received that care by a Mercy LPC who worked directly with primary care staff to address both behavioral health and physical health issues. Many of the services to which primary comparison group patients were referred were provided at the Mercy clinic through co-located services. Primary comparison group participants who were referred to behavioral health care received that care from a Mercy LPC who did not directly work with the primary care providers.

Table 5. Number of Referrals Made and Completed among Intervention and Primary Comparison Groups

Referral Type	Intervention		Primary Comparison		Total	
	Referrals Made	Referrals Closed	Referrals Made	Referrals Closed	Referrals Made	Referrals Closed
Behavioral Health	184	180	15	10	199	190
Community Resource	5	5	2	2	7	7
Diabetic Education	42	39	20	19	62	58
Fitness Center	190	188	1	1	191	189
Nutrition	182	179	3	3	185	185
Total Number	603	591	41	35	644	626
Total Percent	98		85		97	

Table 6 presents show rates by visit number and type for the intervention and primary comparison groups. For each visit type, the intervention group had a higher volume of visits, but lower show rate than the primary comparison group. Across all visit types, the intervention group also had a higher volume, but lower show rate than the primary comparison group (74% versus 84%). These data suggest that the intervention group had a lower participation rate in the recommended intervention services; however, with the high volume of average intervention group visits, an average of ~25 per participant over 12 months across visit types, a lower show rate might be expected among this group. As explained above, primary comparison group services were provided as co-located services at the Mercy clinic rather than the integrated care intervention participants received.

Table 6. Intervention and Primary Comparison Groups Visits Completed, Visits Scheduled, and Show Rates by Visit Type

Visit Type	Visits Completed	Visits Scheduled	Show Rate
Primary Care (Medical Visit)	**2997**	**3582**	**84%**
Integrated Care (Intervention)	1617	1988	81%
Usual Care (Primary Comparison)	1380	1594	87%
Behavioral Health	**798**	**1215**	**66%**
Integrated Care (Intervention)	578	914	63%
Usual Care (Primary Comparison)	220	301	73%
Health Education	**1158**	**1675**	**69%**
Integrated Care (Intervention)	984	1453	68%
Usual Care (Primary Comparison)	174	222	78%
Exercise Coach	**1082**	**1388**	**78%**
Integrated Care (Intervention)	1082	1388	78%
Usual Care (Primary Comparison)[a]	--	--	--
Total (overall study visits)	**6035**	**7860**	**77%**
Integrated Care (Intervention)	4261	5743	74%
Usual Care (Primary Comparison)	1774	2117	84%

[a]Data not reported n<5

As shown in **Table 7**, intervention participants completed 1,617 primary care visits. These ranged from 1 visit to a maximum of 20 visits, with the mean of 7.1 and median of 7 visits per participant in the intervention group. The intervention participants completed 554 behavioral health integrated care visits. These ranged from 1 visit to 17 visits, with the mean of 3.2 and median of 3 visits per participant in the intervention group. Participants in this group completed 811 health education integrated care visits. These ranged from 1 visit to 18 visits, with the mean of 4.7 visits per intervention participant. This group also completed 1,082 exercise coaching visits. These ranged from 1 visit to 115 visits, with the mean of 9.8 visits per intervention participant. Primary comparison group participants completed 1,380 primary care visits. These ranged from 1 visit to 17 visits, with a mean of 6.4 and median of 6 visits per primary comparison group participant. These participants completed 220 behavioral health and 174 health education co-located visits. No participant received more than 1 visit of either service type. Primary comparison group participants did not receive exercise coaching visits.

Table 7. Services Received by Group and Service Type

Service Type	Intervention					Primary Comparison					Total				
	Total	Mean	Median	Minimum	Maximum	Total	Mean	Median	Minimum	Maximum	Total	Mean	Median	Minimum	Maximum
Primary Care	1617	7.1	7.0	1.0	20.0	1380	6.4	6.0	1.0	17.0	2997	6.8	7.0	1.0	20.0
Behavioral Health	578	3.2	3.0	1.0	17.0	220	1.0	1.0	1.0	1.0	798	3.1	2.0	1.0	17.0
Health Education	984	4.7	4.0	1.0	18.0	174	1.0	1.0	1.0	1.0	1158	4.6	4.0	1.0	18.0
Exercise Coaching	1082	9.8	2.0	1.0	115.0	0	0.0	0.0	0.0	0.0	1082	9.8	2.0	1.0	115.0
Total	4261	37.3	32.0	2.0	198.0	5165	24.4	22.0	1.0	93.0	6035	30.9	27.0	1.0	198.0

Question 6. How satisfied are Sí Three patients with the services they have received? How satisfied are providers with the Si Three program?

Patient participants in focus groups were overwhelmingly satisfied with the Sí Three program, citing increases in the availability and quality of services provided, health knowledge, and ultimately health outcomes as reasons for being satisfied. As described above, in response to Question 3b., providers also were very satisfied with the program.

Services Provided
Patient focus group participants spoke highly about the quantity and quality of services received as part of the Sí Three program, including education services and exercise classes as well as the added dimension of spiritual care. *"The program is complete, it's good. I can say I'm complete because of that program, and that's the purpose – to provide complete physical and spiritual care,"* explained one patient. Several other focus group participants agreed, emphasizing the quality and thoroughness of the Sí Three services – *"This is the most complete because it includes all three things [physical, behavioral and spiritual]. So this is the best care I have seen."* In addition to the three dimensions of care, focus group participants expressed satisfaction with receiving self-help tools, which facilitated their participation. These tools included blood sugar monitors and strips, blood pressure monitors, scales, tape measures and exercise bands.

Health Knowledge
Program services, specifically the nutrition education and chronic disease management skills, were seen as increasing health knowledge, and a significant reason why patients were satisfied with the program. *"They help you with meals, nutrition too. The nutrition classes teach you about the vegetables, the meats,"* one patient shared. Another explained further saying, "*A lot of people learned how to portion control … what to eat, the exercises."*

Improved Outcomes
According to focus group participants, they were satisfied with the program because of the additional services provided as well as the improved health knowledge, which led to perceived improvement in health outcomes, both chronic physical disease and behavioral health issues. For example, one patient said, *"I lost like 80 pounds and started coming to the exercise class constantly. I began changing my health and my self-esteem as well."* Others emphasized how they were satisfied with the program because they had more energy and motivation and experienced improvements in their health. Staff interviewees also perceived that patients were satisfied with the program because of improved outcomes, citing many examples of patients who lost weight or controlled their diabetes. As one staff interviewee said, "*They [patients] see the difference [in blood sugar] from when they started in the 200s/300s and now they're 120/130 by simply cutting down on carbohydrates and exercising."*

Mercy provided intervention participants with the opportunity to provide feedback on services received from each type of provider (e.g., nurse practitioner, LPC, educator) through patient satisfaction surveys. Although the number of surveys received from participants was less than 50% of the intervention group, those participants who provided feedback reported being highly satisfied with all providers on both mid-point and endpoint feedback surveys. Please see Appendix R: Satisfaction Survey Results for data on patient satisfaction.

In addition, Mercy staff completed satisfaction surveys with the same items as the participants. Twenty eight of thirty-two staff completed the surveys. Overall, staff reported a high level of satisfaction with

the services that intervention participants received. For two items, patient wait times and patient understanding their care and treatment plan, the mid-point satisfaction ratings were below 60%. The ratings for patient wait times and patient understanding their care and treatment plan increased at the endpoint. Please see Appendix R: Satisfaction Survey Results for data on staff satisfaction.

Question 7. What percent of patients who were seen by primary care providers for diabetes, obesity, and/or hypertension complete standardized assessments (depression, anxiety, addictive behavior, quality of life, and spiritual well-being) on their initial visit?

All intervention group participants completed these standardized assessments on their initial visit.

Question 8. What percent of completed assessment results were recorded according to protocol? Were all staff able to implement standard measurement protocols?

Nearly all completed behavioral health assessment results were recorded according to protocols and staff, for the most part, implemented standard measurement protocols. As described elsewhere in this report, some staff initially did not record waist circumference. The protocol was clarified, and waist circumference was recorded according to protocols. Specific data on the actual recording of results according to protocol is not available; therefore, a completion percentage cannot be calculated.

Question 9. What percent of patients with depression, anxiety, and addictive behavior were referred to the LPC or other behavioral health provider?

Among intervention participants, 96.9% of these participants were referred to the LPC for depression, anxiety, or addictive behaviors. Among primary comparison group participants, 13.1% of the participants were referred to other behavioral health providers as described in the response to Question 5 above.

Question 10. What percent of patients were assessed for depression, anxiety, quality of life, and addictive behavior on a semi-annual basis?

According to Mercy's protocols, all intervention patients were assessed on behavioral health measures on a quarterly basis. From retention data, see **Table 16**, 169 or 81.6% of intervention participants were assessed on these measures at 6 months and 142 or 68.6% of intervention participants were assessed on these measures at 12 months following baseline.

Question 11. What was the show rate for all patients in the Sí Three intervention for IBH services? Does the show rate differ by the type of behavioral health resource the patient used?

The show rate for intervention participants was 74.2% for all visit types. Because nearly all intervention participants used the spiritual behavioral health services, comparisons cannot be made by type of behavioral health resource the patient used.

Question 12. What percent of referred patients could explain their physical and/or behavioral health treatment plans?

Mercy protocols included at least one intervention group participant contact, through telephone or provider visit, quarterly. As part of these contacts, participants were asked to explain their treatment plans. According to Mercy staff, protocols were implemented as planned; however, the implementation of the protocol was not recorded in Mercy EMR. Therefore, the exact percentage of implementation cannot be analyzed.

Additional Implementation Findings

In addition to data to answer the *a priori* implementation questions presented in the SEP, the qualitative implementation evaluation also yielded additional findings related to perceived success and impacts, sustainability, policy implications, program replication/scalability, and staffing. The following outlines key themes that emerged during the key informant interviews and focus groups not directly asked by the implementation research questions outlined above but that are still valuable to provide context for Mercy's program.

Program Successes and Impact

Patients and staff were asked to speak about their perceived successes and impacts of the Sí Three program at Mercy. Summative interview and focus group participants identified the program's impact on clinic integration as well as patients' health knowledge, chronic disease, and behavioral health. Successes identified at the mid-point included meeting the enrollment target quickly and patients' access to new services, such as nutrition education and exercise classes.

Clinic Integration
According to interviewees and focus group participants, one of the successes of the Sí Three program was that it integrated physical, behavioral/emotional, and spiritual care to treat patients holistically. Interviewees shared that this primary program impact was achieved through the reorganization of workspace within the Mercy clinic, or "relocation of staff" to fully integrate the primary care and behavioral health providers as well as the continuation of integrated practices, such as the use of the PHQ-9 and GAD-7 assessments for all patients.

Health Knowledge
As discussed related to participant satisfaction, the Sí Three program was perceived as increasing the health knowledge of patients. From sessions with the exercise coach to visits with the nurse educator, patient focus group participants shared that they learned many new skills, such as how to read food labels, cook healthy foods, and monitor their diabetes, blood pressure, and weight. Patient focus group participants and clinic staff interviewees explained that this education helped patients build a basic understanding of their health conditions to more effectively manage them over time. As one patient explained, *"In my case it was diabetes. They send you to the nutrition class and teach you about food portions. You are being educated to learn how to eat so that your numbers [A1C, blood pressure, weight] will go down."* Staff interviewees reinforced this – *"I think we empowered them more because now they know more about their cholesterol, what it is, and they're able to take better care of themselves."*

Chronic Diseases
Both focus group participants and interviewees alike discussed how the increased services (nutrition education, exercise, and behavioral health) designed to enhance patient health knowledge resulted in improved chronic disease management and outcomes for patients. Many interviewees and focus group

participants shared success stories of patients learning about and managing their diabetes, losing weight, and lowering their blood pressure. *"They started to treat me and bring it [blood pressure] down, and it has helped me a lot,"* described one patient. An administrative staff interviewee echoed this sharing, "*The patients are better… they came, and they really got invested in the nutrition class, the exercise class, monitoring their own health, taking charge of their own being. As they lost weight, they got better and felt better."*

Behavioral Health
Patient focus group participants as well as Mercy staff interviewees also spoke of the program's perceived impact related to patients' behavioral health, which included improvements to quality of life. Focus group and interview participants explained that they saw benefits to patients' behavioral health as a result of physical health improvements, and vice versa. One staff interviewee explained, *"They [patients] are trained [educated] now. They have the skills to reduce their anxiety and they see how it is also helping them with other chronic conditions."* Some patient focus group participants spoke of how the perceived improvements in physical and behavioral health have led to better self-esteem. For example, one patient shared, *"I had bad self-esteem. When I decided to come to counseling, they taught me how to cope with it, how to cure it. Then I wasn't just curing myself, but my family as well. I have seven children and my husband. So, when I started counseling here through the clinic, my life changed because they taught me how to treat my family, how to react to anything that was happening."*

Sustainability and Lessons Learned

Overall, results from interviews with Mercy staff as well as focus groups with patients indicated that implementation of Mercy's Sí Three program has been successful. Several lessons learned and opportunities for improvement emerged. At the mid-point, lessons learned related to clinic space, data systems/evaluation, leadership buy-in, patient barriers to care, and training. During the summative interviews and focus groups, lessons learned and opportunities for improvement focused on information-sharing, leadership buy-in, program replication and scalability, staffing, and training.

Information-sharing
Interview participants offered lessons learned and suggestions related to sharing information externally and internally. Externally, Mercy staff recommended information-sharing, specifically connecting with other individuals, such as psychiatrists, and organizations who are experts in integrated behavioral health when planning the implementation of integrated care. One staff interviewee explained, "*I would have liked to have some communication with a psychiatrist or an advanced mental health person when we were planning this."* Staff also recommended building on the existing nutrition education and physical activity services and extending those to offer groups in the evenings "*to get individuals more involved with different services*," according to one staff interviewee. In terms of internal information-sharing, other staff discussed the need to continuously share information within the clinic, so that "*everybody knows what's happening and keeps informed*" about what phase the program is in and where it is going in the future *(administrative staff interviewee)*.

Leadership Buy-in
Having initial and continued support from organizational leaders was seen as critical to Sí Three program success, according to interviewees. The Mercy Ministries of Laredo President, in particular, was seen as a champion of the Sí Three program, providing support for continuing to assess and provide integrated behavioral health services. "*She has been crucial to incorporating and sustaining the clinic's integration*,"

summarized one staff interviewee. Staff interviewees shared that the vision and leadership of the President has ensured success of the program.

Program Replication and Scalability

While the Sí Three program represents an expansion of Mercy's existing coordinated care between primary care providers and a part-time LPC, interviewees and focus group participants expressed hopes for additional program scale-up. As the Sí Three study ended, interviewees shared that Mercy has implemented a quality improvement initiative to integrate the Sí Three program into the clinic's entire patient population. As one administrative staff interviewee explained, "*For now, we have opened it [SÍ Three program] to the full clinic and every patient that comes through is screened and offered the same services."*

Staffing

As discussed throughout the interviews and focus groups, Mercy's staff were reported by these qualitative data collection participants as critical to the success of the Sí Three program. However, there were numerous lessons learned and opportunities for improvement around staffing, according to interview and focus group participants. The general suggestion from multiple staff interviewees was the need for additional behavioral health providers, specifically another full-time social worker as well as a psychiatrist. Acknowledging the scarcity of behavioral health providers in Laredo, one staff suggested, "*We don't have enough psychiatrists. We don't have enough psychologists. Take advantage of whatever is available in the community.*

In addition to more behavioral health providers, focus group participants would like to see specialists as part of Mercy's staff, specifically recommending a dentist, optometrist, and podiatrist to serve all patients. It was noted that some of these providers are available at Mercy, but infrequently and only for diabetic patients. "*They have the opportunity to do all of that,"* shared one patient. Interviewees noted that it will be important to hire staff for these behavioral health and specialist positions moving forward so that existing and new patients can access enhanced, integrated services.

Training

Although internal and external training of staff occurred, several interviewees suggested future training topics, such as integrated care and the integrated data system. At the mid-point, interviewees reported training was related to working in a shared [integrated] model of care as well as developing evaluation skills. As the Sí Three program concluded and staff interviewees looked back, they reinforced the need for training in integrated care, especially before program implementation begins. "*One of the challenges is that we as providers are not trained in this integration. When you go to school… you don't see the medical part as a behavioral health provider or the behavioral health as a medical [primary care] provider... So, I think we need to start from the beginning, from the education part,"* shared one staff interviewee. Others shared specifics, suggesting that, "*First, I would educate the staff in defining integrated health, why it is that we're doing it [integrated care], why it's important for the clinic,"* according to one staff interviewee, with another adding, "*Education on our roles in the integrated care program."* One administrative staff also highlighted the need for training related to the data system. "*I see a need when we ask staff to run a report or pull data to make sure they understand what they are pulling and why, and the right parameters to get those [data]."*

IMPACT STUDY – APPROACH AND METHODS

Overview of Impact Study Design

Mercy has implemented the Sí Three: Integration of 3D Health Services program in their clinic setting. This program is based on components of the integrated care model studied by Druss, Rohrbaugh, Levinson, & Rosenheck (2001). The Druss model—involving an integrated model of patient education and prevention, nurse practitioners, and increased interaction among the care team—found positive results in health outcomes for patients in the integrated care model compared to those receiving stand-alone medical services. The Sí Three program added faith-based/spiritual counseling to this model. Prior research includes meta-analyses of randomized control trials (RCTs) of religious/spiritual counseling which found greater improvements in psychological and spiritual outcomes as compared to alternate secular therapies (Worthington et al., 2011; Koenig, 2012).

An RCT was not feasible because Mercy preferred to allow participants to make the choice to participate in the intervention rather than randomly assigning participants to study group. Mercy has a policy of offering services to any patient who needs and requests them, as long as the clinic has the capacity to provide these services. A non-randomized QED was chosen to minimize threats to internal validity with the inclusion of a primary comparison group. A secondary comparison group was included in this study to assess the external validity and generalizability of the primary comparison group comparisons.

The primary comparison group minimized several threats to internal validity given that intervention and primary comparison group participants were more similar at baseline on demographic and outcome measures and the primary comparison group participants were patients in the same clinic as the intervention group. More specifically, the primary comparison group addressed the following threats to internal validity: regression to the mean, history, testing, John Henry, and expectancy effects. The secondary comparison group enhanced internal and external validity and served as a sensitivity analysis for the primary comparison group findings. More specifically, the secondary comparison group addresses the internal validity threats of selection bias and novelty. Also, the secondary comparison group addresses external threats to validity including applicability to other populations and applicability to other settings/locations.

Nevertheless, results may have limited external validity due to the setting of the study and the specific patient population being studied. Results may be generalized to other border communities but may have limited generalizability in settings outside of Southern Texas and/or other border communities in the United States. The study targeted a moderate level of evidence. The aim of the study is to contribute to the body of evidence associated with the understanding of integrated behavioral health services in clinics serving predominantly low-income, Hispanic communities.

Impact Study Design and Methods

Study Design

The study utilized a non-randomized QED to evaluate the Sí Three program's impact by comparing program participants to patients who did not participate in the program. The design allowed for identification and controlling for observed characteristics that could have affected impact measures of interest. For this study, two comparison groups were used. A primary clinic comparison group of

potentially similar patients who chose not to participate in the Sí Three program was recruited from Mercy's clinic population and enrolled in the study. Analyses with the primary comparison group are considered the main study, while analyses with the secondary comparison group are included to enhance the external validity and generalizability of the primary comparison group results. The analyses between the intervention and secondary comparison group serve as sensitivity analyses aimed at an increased understanding of the intervention effects and how they may or may not differ when compared to a secondary comparison group under different conditions. Patients from Nuestra Clinical del Valle's Edcouch and Alton clinics were recruited and enrolled in the study as a secondary comparison group. Participants enrolled in the study were followed for approximately 12 months. Quantitative program implementation data related to participation in intervention components is also provided in this report (see Implementation Evaluation section). This study did not deviate from the SEP in its methodology or design.

Assessment of Baseline Equivalence

At baseline, sociodemographic characteristic frequencies were analyzed for both intervention and primary comparison groups collected through a standardized set of questions developed by Mercy that were administered at the clinic. To assess baseline equivalence between these groups, the following sociodemographic characteristics were analyzed: sex, ethnicity, age, employment status, marital status, primary language, smoking, alcohol consumption, and Spirituality Index score. Baseline sociodemographic data were captured for all program participants; however, for marital status and alcohol consumption responses of "unknown" were recoded as missing as noted in **Table 8**. Baseline statistical comparisons were not performed for the following characteristics for the intervention and primary comparison groups: 1) county of residence because of the homogeneity of the study population on this characteristic, and 2) perceived spiritual strength (assessed by the LPC) because these data were only collected from intervention participants.

Among patient-level demographic characteristics, the intervention and primary comparison groups were statistically equivalent on all measures except for Spirituality Index score and employment status. The primary comparison group had a statistically significant higher average Spirituality Index score (50.2) compared to the intervention group (47.6). The intervention group had a statistically significant higher proportion of participants who noted they were not employed (56.5%) compared to the primary comparison group (47.1%).

Table 8. Tests of Baseline Equivalence for Demographic Measures: Intervention and Primary Comparison Groups

	Full Sample (n=410)		Intervention Group (n=207)		Primary Comparison Group (n=203)		p-value
Variables	N	%	N	%	N	%	
Sex							
Male	52	12.7	27	13.0	25	12.3	0.82
Female	359	87.3	180	87.0	178	87.7	
Ethnicity							
Hispanic	409	99.8	206	99.5	203	100.0	0.99
Non-Hispanic	1	0.2	1	0.5	0	0.0	

	Full Sample (n=410)		Intervention Group (n=207)		Primary Comparison Group (n=203)		p-value
Variables	N	%	N	%	N	%	
Age							
Mean	44.1	--	43.8	--	44.3	--	0.62
SD	10.8	--	11.3	--	10.3	--	
18-24	16	3.9	9	4.3	7	3.5	0.63
25-34	55	13.4	29	14.0	26	12.8	
35-44	147	35.9	75	36.2	72	35.5	
45-54	120	29.3	61	29.5	59	29.1	
55-64	67	16.3	29	14.0	38	18.7	
65+	5	1.2	4	1.9	1	0.5	
Employment Status[a]							
Employed	95	23.2	36	17.4	59	29.1	**0.03**
Not Employed	213	52.0	117	56.5	96	47.3	
Self Employed	99	24.2	53	25.6	46	22.7	
Student	3	0.7	1	0.5	2	1.0	
Marital Status							
Divorced	24	5.9	11	5.3	13	6.4	0.86
Legally Separated	26	6.4	11	5.3	15	7.4	
Married	218	53.3	113	54.6	105	51.7	
Significant Other	42	10.3	19	9.2	23	11.3	
Single	82	20.1	44	21.3	38	18.7	
Widowed	17	4.2	8	3.9	9	4.4	
Missing	1	--	1	--	--	--	
Primary Language							
English	50	12.2	26	12.6	24	11.8	0.82
Spanish	360	87.8	181	87.4	179	88.2	
County of Residence							
Webb County	410	100.0	207	100.0	203	100.0	--
Missing	--	--	--	--	--	--	
Smoking Status							
Current Smoker	38	9.2	23	11.1	15	7.4	0.43
Former Smoker	14	3.4	7	3.4	7	3.4	
Never Smoked	358	87.3	177	85.5	181	89.2	
Alcohol Consumption							
Yes	83	20.7	43	21.4	40	20.0	0.73
No	318	79.3	158	78.6	160	80.0	
Missing	9	--	6	--	3	--	
Spirituality Index							
Mean	48.9	--	47.6	--	50.2	--	**0.004**
SD	12.1	--	11.9	--	12.1	--	
Perceived Spiritual Strength							
Weak	--	--	16	10.6	--	--	--

	Full Sample (n=410)		Intervention Group (n=207)		Primary Comparison Group (n=203)		p-value
Variables	N	%	N	%	N	%	
Moderate	--	--	37	24.5	--	--	
Strong	--	--	98	64.9	--	--	
Missing	--	--	56	--	--	--	

Note: Bold denotes statistical significance (p-value < 0.05).
[a] Fisher's Exact test was used due to cells having expected count less than 5

Baseline equivalence was assessed for chronic disease status using the study impact measures (systolic blood pressure, diastolic blood pressure, HbA1c, BMI, waist circumference, PHQ-9, Duke General Health, GAD-7, and CAGE-AID) as noted in **Table 9**. Equivalence was assessed using t tests for continuous variables and Chi-square tests for categorical variables. For PHQ-9, Duke General Health, GAD-7, HbA1c, and CAGE-AID measures, nonparametric tests were employed due to non-normal distributions.

Examining baseline equivalence on the impact measures evaluates whether the two groups are statistically equivalent at that time point. For the nine impact measures in Mercy's primary comparison study, the intervention and primary comparison groups were statistically equivalent on all but three measures (PHQ-9, Duke General Health, and GAD-7). The primary comparison group began the study healthier on these measures than the intervention group with lower median PHQ-9 and GAD-7 scores and a higher median Duke General Health score.

Table 9. Tests of Baseline Equivalence for Impact Measures: Intervention and Primary Comparison Groups

	Full Sample (n=410) Mean (SD)	Intervention (n=207) Mean (SD)	Primary Comparison (n=203) Mean (SD)	p-value
BMI	32.9 (6.6)	33.2 (7.1)	32.5 (6.0)	0.26
Systolic blood pressure	124.6 (17.5)	125.3 (18.4)	123.9 (16.5)	0.41
Diastolic blood pressure	74.2 (9.8)	74.9 (10.1)	73.5 (9.5)	0.16
Waist Circumference: Males	42.2 (5.1)	41.5 (4.2)	43.0 (6.0)	0.31
Waist Circumference: Females	43.6 (5.6)	43.7 (6.0)	43.5 (5.1)	0.76
General Health	**71.1 (17.2)**	**67.7 (17.5)**	**74.5 (16.2)**	**<0.001**
Nonparametric Tests[a]	Median (SD)	Median (SD)	Median (SD)	*p*
PHQ-9	**4.0 (5.5)**	**5.0 (6.0)**	**3.0 (4.6)**	**<0.001**
GAD-7	**4.0 (5.3)**	**5.0 (5.6)**	**3.0 (4.7)**	**0.001**
HbA1c	6.6 (1.9)	6.5 (2.0)	6.7 (1.6)	0.76
CAGE-AID	0.0 (0.6)	0.0 (0.6)	0.0 (0.6)	0.66

Note: Bold denotes statistical significance (p-value < 0.05).)
[a] The Wilcoxon rank sum *test was used to examine non-normally distributed data*

Because this study used a quasi-experimental design and did not employ randomization to achieve baseline equivalence, adjusted regression analyses was proposed as the main analytic approach in the SEP to analyze the intervention effect accounting for potential confounders. It was not possible to employ matching in the study design phase as Mercy patients were allowed to choose whether to participate in the intervention or primary comparison group.

Similarly, at baseline, sociodemographic characteristic frequencies were analyzed for both the intervention and secondary comparison group. To assess baseline equivalence between these groups, the following sociodemographic characteristics were analyzed: sex, ethnicity, county of residence, age, employment status, marital status, primary language, smoking, and alcohol consumption. Baseline sociodemographic data were captured for all program participants; however, for marital status and alcohol consumption responses of "unknown" were recoded as missing as noted in **Table 10**.

Among patient-level demographic characteristics, again there were several statistically significant differences between the intervention and secondary comparison group. The intervention group was comprised of a higher proportion of females than the secondary comparison group, and the two groups lived in exclusively different counties (Note: county of residence was an expected difference between the two groups). The intervention group had a lower average age than the secondary comparison group. Also, a smaller percentage of participants in the intervention group were married and spoke English compared to participants in the secondary comparison group.

Table 10. Tests of Baseline Equivalence for Demographic Measures: Intervention and Secondary Comparison Group

	Full Sample (n=573)		Intervention (n=207)		Secondary Comparison (n=366)		p-value
Measure	n	%	n	%	n	%	
Sex							
Male	131	22.9	27	13.0	104	28.4	**<0.001**
Female	442	77.1	180	87.0	262	71.6	
Ethnicity							
Hispanic	570	99.5	206	99.5	364	99.5	0.92
Non-Hispanic	3	0.5	1	0.5	2	0.5	
County of Residence							
Cameron	1	0.2	0	0.0	1	0.3	**<0.001**
Hidalgo	365	63.7	0	0.0	365	99.7	
Webb	207	36.1	207	100.0	0	0.0	
Missing	--	--	--	--	--	--	
Age							
18-24	18	3.1	9	4.3	9	2.5	**<0.001**
25-34	50	8.7	29	14.0	21	5.7	
35-44	158	27.6	75	36.2	83	22.7	
45-54	180	31.4	61	29.5	119	32.5	
55-64	135	23.6	29	14.0	106	29.0	
65+	32	5.6	4	1.9	28	7.7	
Mean	47.8	--	43.8	--	50.1	--	**<0.001**
SD	11.9	--	11.3	--	11.6	--	
Employment Status							
Employed	222	38.7	89	43.0	133	36.3	0.12
Not Employed	351	61.3	118	57.0	233	63.7	
Marital Status							
Divorced	24	4.2	11	5.3	13	3.6	**<0.001**

Legally Separated	41	7.2	11	5.3	30	8.2	
Married	351	61.3	113	54.6	238	65.0	
Significant Other	19	3.3	19	9.2	0	0.0	
Single	104	18.2	44	21.3	60	16.4	
Widowed	31	5.4	8	3.9	23	6.3	
Missing	3	--	1	--	2	--	
Primary Language							
English	126	22.0	26	12.6	100	27.3	
Spanish	446	77.8	181	87.4	265	72.4	**<0.001**
SL	1	0.2	0	0.0	1	0.3	
Smoking Status							
Current Smoker	47	8.2	23	11.1	24	6.6	
Former Smoker	30	5.2	7	3.4	23	6.3	0.06
Never Smoked	496	86.6	177	85.5	319	87.2	
Alcohol Consumption							
Yes	120	21.2	43	21.4	77	21.0	
No	447	78.8	158	78.6	289	79.0	0.92
Missing	6	--	6	--	--	--	

Note: Bold denotes statistical significance (p-value < 0.05)

Baseline equivalence was assessed for chronic disease and behavioral health status using the study impact measures (systolic blood pressure, diastolic blood pressure, HbA1c, BMI, waist circumference, PHQ-9, Duke General Health, and GAD-7) as noted in **Table 11**. Equivalence was assessed using t-tests for continuous variables and Chi-square tests for categorical variables. For PHQ-9, Duke General Health, GAD-7, and HbA1c measures, nonparametric tests were employed due to non-normal distributions.

When examining baseline equivalence for the eight impact measures in Mercy's secondary comparison group, the intervention group began the study with lower average BMI and both systolic and diastolic blood pressure compared to the secondary comparison group. The intervention group was less healthy regarding behavioral health measures with higher median PHQ-9 and GAD-7 scores and a lower median Duke General Health score. Females in the intervention group began the study with a higher mean waist circumference than females in the secondary comparison group.

Table 11. Tests of Baseline Equivalence for Impact Measures: Intervention and Secondary Comparison

	Full Sample (n=573) Mean (SD)	Intervention (n=207) Mean (SD)	Secondary Comparison (n=366) Mean (SD)	p-value
BMI	**34.6 (7.5)**	**33.2 (7.1)**	**35.4 (7.6)**	**0.001**
Systolic	**128.8 (18.9)**	**125.3 (18.4)**	**130.8 (18.9)**	**0.001**
Diastolic	**79.0 (10.6)**	**74.9 (10.1)**	**81.3 (10.2)**	**<0.001**
Waist Circumference – Males	40.8 (6.2)	41.5 (4.2)	40.6 (6.6)	0.54
Waist Circumference – Females	**41.5 (5.8)**	**43.7 (6.0)**	**40.1 (5.4)**	**<0.001**
Non-Parametric Tests[a]	Median (SD)	Median (SD)	Median (SD)	p-value
PHQ-9	**2.0 (4.7)**	**5.0 (6.0)**	**1.0 (2.5)**	**<0.001**
General Health	**80.0 (17.5)**	**66.7 (17.5)**	**86.7 (15.4)**	**<0.001**
GAD-7	**1.0 (4.4)**	**5.0 (5.6)**	**0.0 (4.7)**	**<0.001**
HbA1c	6.7 (2.0)	6.5 (2.0)	6.8 (2.0)	0.43

Note: Bold denotes statistical significance (p-value < 0.05).

[a] The Wilcoxon rank sum test was used to examine non-normally distributed data; these results aligned with t test results.

Because this study used a quasi-experimental design and did not employ randomization to achieve baseline equivalence, adjusted regression analyses was proposed as the main analytic approach in the SEP to analyze the intervention effect accounting for potential confounders. Additionally, it was not possible to employ matching in the study design phase since the NCDV participants were also serving as a comparison group to another study in the Sí Texas portfolio. Therefore, statistical matching at the analysis phase was proposed in the SEP. The proposed matching method to evaluate the robustness of the main results was propensity score matching. In general, propensity score matching is typically used with a large set of covariates among large samples by matching cases with controls based on covariance of these covariates. It has been shown to reduce selection bias that may be present in observational and quasi-experimental design studies (Rubin and Thomas, 1996). Specifically, propensity score matching identifies close matches and removes participants from the analytic samples that have no appropriate match in the other group. This trade-off of reduced bias and reduced efficiency (due to discarded observations) tends to favor accuracy in large samples with many covariates (e.g., greater than 30 covariates), but can be challenging in terms of reduced precision and decreased statistical power in smaller sample evaluation studies with fewer number of covariates.

As proposed in the SEP, only a limited set of covariates were collected among intervention and comparison groups during the Sí Three study. The optimal matching algorithm within the nearest neighbor matching method was conducted and found that the propensity score matching reduced the total sample by 93 participants or 31.6% of the primary comparison group analysis sample. Discarding nearly a third of the study sample who completed an assessment at 12 months reduced statistical power. For the secondary comparison group analysis sample, a total of 208 or 50.2% were excluded using this matching method, also resulting in a reduction of power. This is in part due to a limited set of covariates and the inherent differences between the intervention group and comparison groups, particularly the secondary comparison group. Other matching methods (i.e., weighting, full matching, and sub-classification) require additional assumptions and weight assignment (either implicit or explicit), which are generally not as preferable as the optimal matching based on nearest neighbor method (Stuart 2010). Given the limitations of reduced analytic sample, a small number of covariates and properties of alternative matching methods, the adjusted regression approach accounting for available

covariates with model selection procedure, which have been properly conducted in the analyses, should be the most appropriate approach to ascertain the intervention effect.

Intervention and Comparison Group Conditions

Mercy patients who selected to participate in the intervention group were offered appropriate integrated behavioral health (IBH) services according to their needs identified through physical and behavioral health assessments. All study patients completed the Spirituality Index of Well-Being. The Sí Three LPC was able to access this assessment in the EMR, assess the patient's perceived spiritual strength during the visit, and provide behavioral healthcare that aligned with the patient's spirituality score and perceived spiritual strength. The SEP proposed that intervention patients who met the criteria for depression and anxiety and had a score greater than 50 on the Spirituality Index were to be offered the option of in-house secular or faith-based behavioral health services according to protocols. Patients who scored 50 or less on the Spirituality Index were to be offered secular behavioral health services only. However, the LPC for the intervention group was a pastoral counselor and hospital chaplain as well as an LPC. This LPC offered all intervention group participants faith-based or secular behavioral health services based on her assessment and patient preference as well as the patient's Spirituality Index of Well Being score.

Intervention patients who met the criteria for addiction were offered addiction services according to protocols. In addition, patients who met the physical criteria of high blood pressure, diabetes, or obesity received medical treatment, educational, dietetic, and exercise services at Mercy to improve their physical health according to protocols. A limited number of services were provided by community partners.

The nurse practitioners were the referring provider for internal and external services and served as the patient navigator for referred services. The care coordinator reviewed intervention group patient attendance at recommended services (physical and secular/faith-based behavioral) and contacted patients who do not attend referred services to increase patient participation in referred services.

When patients were referred for specific services (e.g., referral to behavioral health, referral to health education and/or referral to exercise class), the referring provider noted the referral in the EMR. Once a referred service was provided, that provider noted completion in the EMR. The care coordinator was proactive in follow-up to review open referrals. All patient encounters within and external to the clinic were tracked through EPIC EMR. Due to the nature of services, those patients who participated in Alcoholics Anonymous or other self-help support groups self-reported participation. The EPIC EMR system was used to track referrals.

The primary comparison group was monitored per usual Mercy clinic protocols for impact measures as described in the SEP. To be consistent with Mercy's mission and values, staff provided needed care to any patient in the clinic. This policy created the potential for contamination across groups in only a few cases. The primary comparison group may have received services from the same behavioral health providers who provided services to the intervention group if the services were needed immediately and the comparison group provider was unavailable. For example, in cases of emergency, primary comparison group participants received behavioral health counseling from the LPC who worked with intervention group participants. Mercy put in place several strategies to minimize contamination, including recording in the EMR whether a patient was in the intervention or primary comparison group.

This assisted providers to appropriately refer intervention participants for integrated services and primary comparison group participants to co-located services. Further, the Sí Three LPC for the intervention group was a separate provider from the part-time counselor who provided care to the comparison group participants. The counselor for the primary comparison group is not trained in pastoral counseling. Mercy's EPIC EMR system recorded whether a patient was in the intervention or primary comparison group and also whether they received any program-like services. Contamination was assessed and addressed as needed in the impact analyses presented later in this report.

In summary, the intervention services included 1) the use of a care coordinator for follow-up; 2) integrated behavioral health services provided by one individual who is trained as an LPC and pastoral counselor; and 3) an exercise coach and nurse educator (nutrition and general health) specifically hired for the project who only served the intervention group.

The secondary comparison participants came from NCDV from either NCDV's Edcouch or Alton Clinic. Edcouch Clinic usual care for behavioral health entailed referring patients for behavioral health services when the patient had a PHQ-9 score ≥ 10 or when the PCP observed behavioral health distress, to a NCDV clinic with a behavioral health provider. The nurse called the Behavioral Health Care Manager to schedule an appointment for the patient to be seen by an LPC at the NCDV San Juan or Mercedes clinic dependent on the patient's discretion. In addition, if desired, the patient could ask the care manager to set up an appointment to be seen by the LPC. If the patient showed suicidal ideation with a plan to hurt him or herself or others, a call was made to the mental health authority, Tropical Texas Behavioral Health. Usual care regarding primary care involved a visit with the medical doctor.

Alton Clinic usual care for behavioral health entailed referring patients with a PHQ-9 score ≥ 10 or when the PCP observed behavioral health distress to outside behavioral health services or with an in-clinic visit with an LPC. The LPC was only at the Alton Clinic for one day every two weeks. Patients could call the care manager to set up an appointment to be seen by the LPC if she was there. If a patient did not have an appointment scheduled on the day the LPC was at the Alton Clinic, the patient was given the option to set up an appointment at the NCDV San Juan clinic or NCDV Mercedes clinic where he or she would be seen by a LPC. If the patient showed suicidal ideation with a plan to hurt him or herself or others, a call was made to the mental health authority, Tropical Texas Behavioral Health. Usual care regarding primary care involves a visit with the medical doctor.

During the study period, NCDV began offering patients at both clinics nutrition education information.

Appendix I: Study Group Comparison Table provides additional details on the differences between processes and services for the intervention, primary comparison, and secondary comparison groups.

Study Sample

The following section describes the final data on the composition, eligibility, recruitment, enrollment, retention, and attrition of the study sample. Except where explicitly noted in subsections below, there were no deviations from the SEP in the Study Sample section, including no deviations from the SEP related to sample recruitment and retention, assessment and adjustment for non-response bias, or missing data.

Study Sample Composition

Table 12 presents participant demographics for the intervention and primary comparison groups at baseline. Intervention and primary comparison group study participants lived exclusively in Webb County. Most of the participants enrolled in these study groups were female (87.3%), Hispanic (99.8%), and spoke Spanish as their primary language (87.8%). The average participant age was 44.1 years. Over half of participants were not employed (52.0%) and were married (53.3%). The majority of participants reported they had never smoked (87.3%), and they did not consume alcohol (79.3%). The average Spirituality Index score was 48.9 out of 60. In addition, the LPC who provided services to the intervention group assessed perceived spiritual strength of participants; over half of participants reported strong perceived spiritual strength (64.9%).

Table 12. Participant Demographic Descriptive Statistics: Intervention and Primary Comparison Groups

	Full Sample (n=410)		Intervention (n=207)		Primary Comparison (n=203)	
Variables	N	%	N	%	N	%
Sex						
Male	52	12.7	27	13.0	25	12.3
Female	359	87.3	180	87.0	178	87.7
Ethnicity						
Hispanic	409	99.8	206	99.5	203	100.0
Non-Hispanic	1	0.2	1	0.5	0	0.0
Age						
Mean	44.1	--	43.8	--	44.3	--
SD	10.8	--	11.3	--	10.3	--
18-24	16	3.9	9	4.3	7	3.5
25-34	55	13.4	29	14.0	26	12.8
35-44	147	35.9	75	36.2	72	35.5
45-54	120	29.3	61	29.5	59	29.1
55-64	67	16.3	29	14.0	38	18.7
65+	5	1.2	4	1.9	1	0.5
Employment Status[a]						
Employed	95	23.2	36	17.4	59	29.1
Not Employed	213	52.0	117	56.5	96	47.3
Self Employed	99	24.2	53	25.6	46	22.7
Student	3	0.7	1	0.5	2	1.0
Marital Status						
Divorced	24	5.9	11	5.3	13	6.4
Legally Separated	26	6.4	11	5.3	15	7.4
Married	218	53.3	113	54.6	105	51.7
Significant Other	42	10.3	19	9.2	23	11.3
Single	82	20.1	44	21.3	38	18.7
Widowed	17	4.2	8	3.9	9	4.4
Missing	1	--	1	--	--	--

	Full Sample (n=410)		Intervention (n=207)		Primary Comparison (n=203)	
Variables	N	%	N	%	N	%
Primary Language						
English	50	12.2	26	12.6	24	11.8
Spanish	360	87.8	181	87.4	179	88.2
County of Residence						
Webb County	410	100.0	207	100.0	203	100.0
Missing	--	--	--	--	--	--
Smoking Status						
Current Smoker	38	9.2	23	11.1	15	7.4
Former Smoker	14	3.4	7	3.4	7	3.4
Never Smoked	358	87.3	177	85.5	181	89.2
Alcohol Consumption						
Yes	83	20.7	43	21.4	40	20.0
No	318	79.3	158	78.6	160	80.0
Missing	9	--	6	--	3	--
Spirituality Index						
Mean	48.9	--	47.6	--	50.2	--
SD	12.1	--	11.9	--	12.1	--
Perceived Spiritual Strength						
Weak	--	--	16	10.6	--	--
Moderate	--	--	37	24.5	--	--
Strong	--	--	98	64.9	--	--
Missing	--	--	56	--	--	--

Note: Bold denotes statistical significance (p-value < 0.05).
[a] Fisher's Exact test was used due to cells having expected count less than 5

Table 13 describes participant impact measures at baseline for the intervention and primary comparison groups. The intervention group began the study with a higher average BMI and systolic and diastolic blood pressures than the primary comparison group. The average waist circumference for males in the intervention was lower than for males in the primary comparison group. The mean waist circumference for females in both groups were similar at baseline. Intervention group participants had higher median PHQ-9 and GAD-7 scores and a lower median Duke General Health score and HbA1c level. As previously mentioned, in the assessment of baseline equivalence, there was a statistically significant difference between the study groups for PHQ-9, GAD-7, and Duke General Health scores.

Table 13. Baseline Primary Impact Measures: Intervention and Primary Comparison Groups

	Full Sample (n=410) Mean (SD)	Intervention (n=207) Mean (SD)	Primary Comparison (n=203) Mean (SD)
BMI	32.9 (6.6)	33.2 (7.1)	32.5 (6.0)
Systolic	124.6 (17.5)	125.3 (18.4)	123.9 (16.5)
Diastolic	74.2 (9.8)	74.9 (10.1)	73.5 (9.5)
Waist Circumference: Males	42.2 (5.1)	41.5 (4.2)	43.0 (6.0)
Waist Circumference: Females	43.6 (5.6)	43.7 (6.0)	43.5 (5.1)
General Health	71.1 (17.2)	67.7 (17.5)	74.5 (16.2)
Nonparametric Tests[a]	Median (SD)	Median (SD)	Median (SD)
PHQ-9	4.0 (5.5)	5.0 (6.0)	3.0 (4.6)
GAD-7	4.0 (5.3)	5.0 (5.6)	3.0 (4.7)
HbA1c	6.6 (1.9)	6.5 (2.0)	6.7 (1.6)
CAGE-AID	0.0 (0.6)	0.0 (0.6)	0.0 (0.6)

[a] The Wilcoxon Signed Rank test was used to examine non-normally distributed data.

Table 14 provides patient demographic and clinic characteristics and prevalence of study outcomes at the clinic level for the Mercy Clinic and the Edcouch Clinic, the secondary comparison clinic identified in the SEP. The Alton Clinic was added during study implementation to ensure an adequate secondary comparison group pool. (Note: Because only a small portion of the secondary comparison group comes from the Alton Clinic, the Edcouch Clinic demographics are most appropriate for this comparison. The Edcouch Clinic appears to differ from Mercy Clinic patient and clinic characteristics on all measures except level of integration.

Table 14. Comparison of Mercy and Edcouch Clinics on Patient Demographics, Study Outcomes, and Clinic Characteristics

	Mercy Clinic	**Edcouch Clinic (Secondary Comparison)**
Total Adult Patient Count	1,864	1,674
Age		
Mean age	43	52
% 18-34	26.1	20.3
% 35 – 64	71.1	70.9
% > 65	2.7	8.8
Race/Ethnicity		
% Hispanic	98.4	99.6
% Non-Hispanic	1.6	0.4
Gender		
% Female	83.4	68.9
% Male	16.6	31.1
Insurance Status		
% Self Pay	98.5	85.4
% Medicare/Medicaid/Other	1.5	14.6
Blood Pressure		
% of Patients with Elevated blood pressure (>140/90)	20.3	49.4

HbA1c		
% of Patients with Elevated A1c (Hb A1c > 7.0)	39.6	54.6
Body Mass Index		
% Obese	50.7	67.9
Provider Mix		
Number of PCPs	0	1
Number of Residents	0	0
Number of Nurse Practitioners	4	1
Number of Registered Nurses	2	1
Number of Licensed Vocational Nurses	0	1

Table 15 presents participant demographics for the intervention and secondary comparison groups at baseline. While the intervention participants lived exclusively in Webb County, those in the secondary comparison group lived almost exclusively in Hidalgo County. Most of the participants enrolled in these study groups were female (77.1%). Participants in both study groups were primarily Hispanic (99.5%). The majority of the two groups spoke Spanish as their primary language (77.8%). The average participant was 47.8 years old across both the intervention and secondary comparison groups. Over half of participants were not employed (61.3%) and were married (61.3%). The majority of participants reported they had never smoked (86.6%) and they did not consume alcohol (78.8%).

Table 15. Participant Demographic Descriptive Statistics: Intervention and Secondary Comparison Groups

	Full Sample (n=573)		Intervention (n=207)		Secondary Comparison (n=366)	
Measure	n	%	n	%	n	%
Sex						
Male	131	22.9	27	13.0	104	28.4
Female	442	77.1	180	87.0	262	71.6
Ethnicity						
Hispanic	570	99.5	206	99.5	364	99.5
Non-Hispanic	3	0.5	1	0.5	2	0.5
County of Residence						
Cameron	1	0.2	0	0.0	1	0.3
Hidalgo	365	63.7	0	0.0	365	99.7
Webb	207	36.1	207	100.0	0	0.0
Age						
18-24	18	3.1	9	4.3	9	2.5
25-34	50	8.7	29	14.0	21	5.7
35-44	158	27.6	75	36.2	83	22.7
45-54	180	31.4	61	29.5	119	32.5
55-64	135	23.6	29	14.0	106	29.0
65+	32	5.6	4	1.9	28	7.7
Mean	47.8	--	43.8	--	50.1	--
SD	11.9	--	11.3	--	11.6	--

Employment Status						
Employed	222	38.7	89	43.0	133	36.3
Not Employed	351	61.3	118	57.0	233	63.7
Marital Status						
Divorced	24	4.2	11	5.3	13	3.6
Legally Separated	41	7.2	11	5.3	30	8.2
Married	351	61.3	113	54.6	238	65.0
Significant Other	19	3.3	19	9.2	0	0.0
Single	104	18.2	44	21.3	60	16.4
Widowed	31	5.4	8	3.9	23	6.3
Missing	*3*	--	*1*	--	*2*	--
Primary Language						
English	126	22.0	26	12.6	100	27.3
Spanish	446	77.8	181	87.4	265	72.4
SL	1	0.2	0	0.0	1	0.3
Smoking Status						
Current Smoker	47	8.2	23	11.1	24	6.6
Former Smoker	30	5.2	7	3.4	23	6.3
Never Smoked	496	86.6	177	85.5	319	87.2
Alcohol Consumption						
Yes	120	21.2	43	21.4	77	21.0
No	447	78.8	158	78.6	289	79.0
Missing	*6*	--	*6*	--	--	--

Table 16 describes participant impact measures at baseline for the intervention and secondary comparison groups. The intervention group began the study with lower average BMI and systolic and diastolic blood pressures than the secondary comparison group. The average waist circumference for both males and females in the intervention group was higher than in the secondary comparison group. Intervention group participants had higher median PHQ-9 and GAD-7 scores and a lower median Duke General Health score and HbA1c level. As previously mentioned, in the assessment of baseline equivalence, there is a statistically significant difference between the study groups for BMI, blood pressure measures, female waist circumference, PHQ-9, GAD-7, and Duke General Health scores.

Table 16. Baseline Primary Impact Measures: Intervention and Secondary Comparison Groups

	Full Sample (n=573)	Intervention (n=207)	Secondary Comparison (n=366)
	Mean (SD)	Mean (SD)	Mean (SD)
BMI	34.6 (7.5)	33.2 (7.1)	35.4 (7.6)
Systolic	128.8 (18.9)	125.3 (18.4)	130.8 (18.9)
Diastolic	79.0 (10.6)	74.9 (10.1)	81.3 (10.2)
Waist Circumference – Males	40.8 (6.2)	41.5 (4.2)	40.6 (6.6)
Waist Circumference – Females	41.5 (5.8)	43.7 (6.0)	40.1 (5.4)
Non-Parametric Tests[a]	Median (SD)	Median (SD)	Median (SD)
PHQ-9	2.0 (4.7)	5.0 (6.0)	1.0 (2.5)
General Health	80.0 (17.5)	66.7 (17.5)	86.7 (15.4)
GAD-7	1.0 (4.4)	5.0 (5.6)	0.0 (4.7)
HbA1c	6.7 (2.0)	6.5 (2.0)	6.8 (2.0)

[a] The Wilcoxon Signed Rank test was used to examine non-normally distributed data.

Patient Flow Description

Patient flow diagrams for Mercy intervention, primary comparison, and secondary comparison participants, following the CONSORT structure, are presented on the following page in **Figure 1** and **Figure 2** (Schulz et al., 2010). This diagram depicts the study process from assessment of eligibility, to enrollment and group selection, ending with retention and analysis. Sample sizes are provided throughout to show where there was participant attrition. Qualitative reasons for any ineligibility, withdrawal, or lost-to-follow-up are provided where applicable. For the intervention and primary comparison groups, in the "enrollment" stage, the 106 participants who were excluded did not meet one or more of the eligibility criteria; an additional 83 patients were eligible but declined to participate. For the secondary comparison group, 37 were excluded due to lack of interest, time, or availability of transportation. In the "follow-up" stage, those participants categorized as "lost to follow-up" did not complete an assessment at that time point but did not withdraw from the study. Due to the lack of official withdrawal from the study, those who were lost to follow-up at 6 months remained in the study and were still eligible to complete a 12-month assessment.

Figure 1. Patient Flow Description, Intervention and Primary Comparison

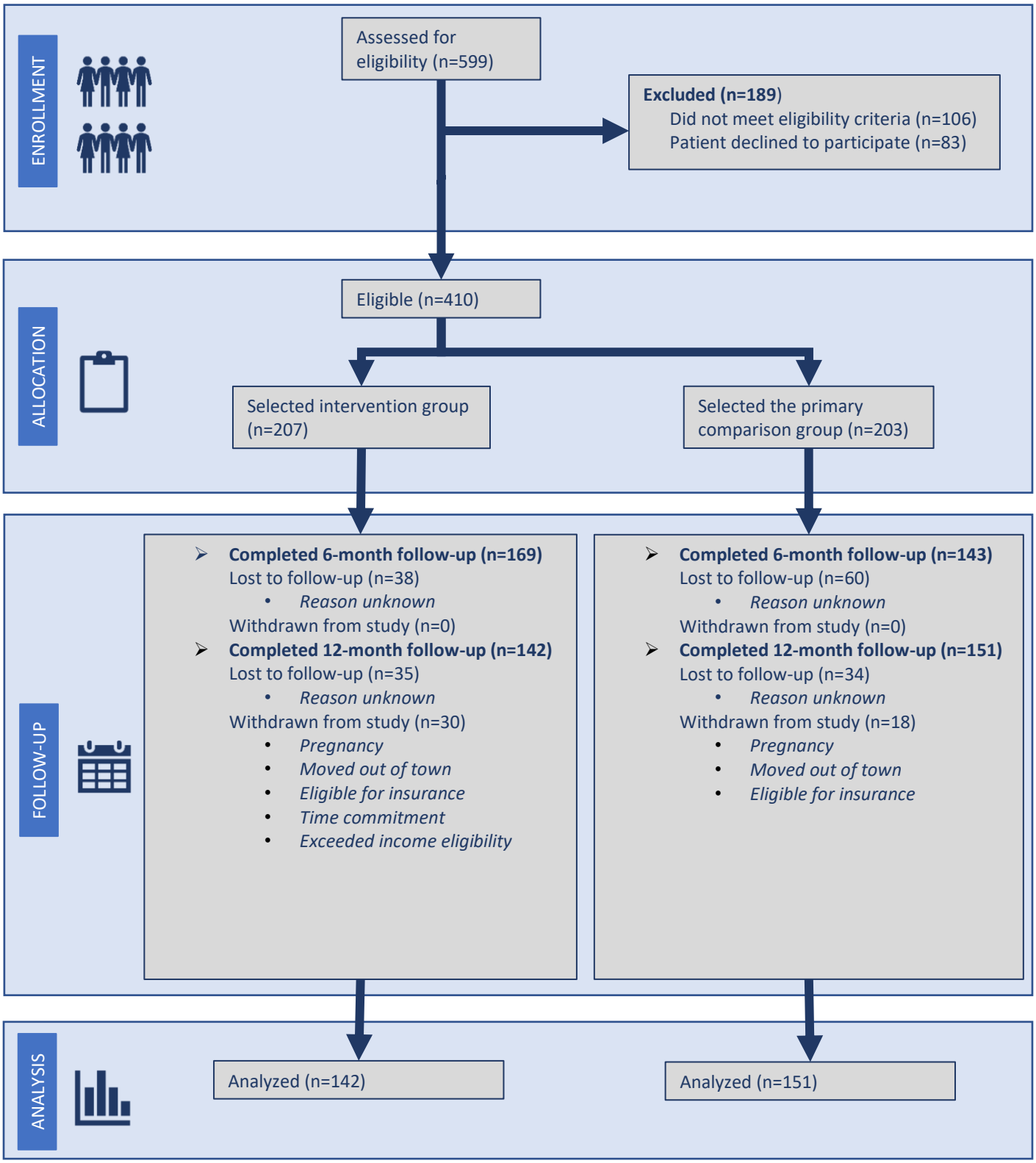
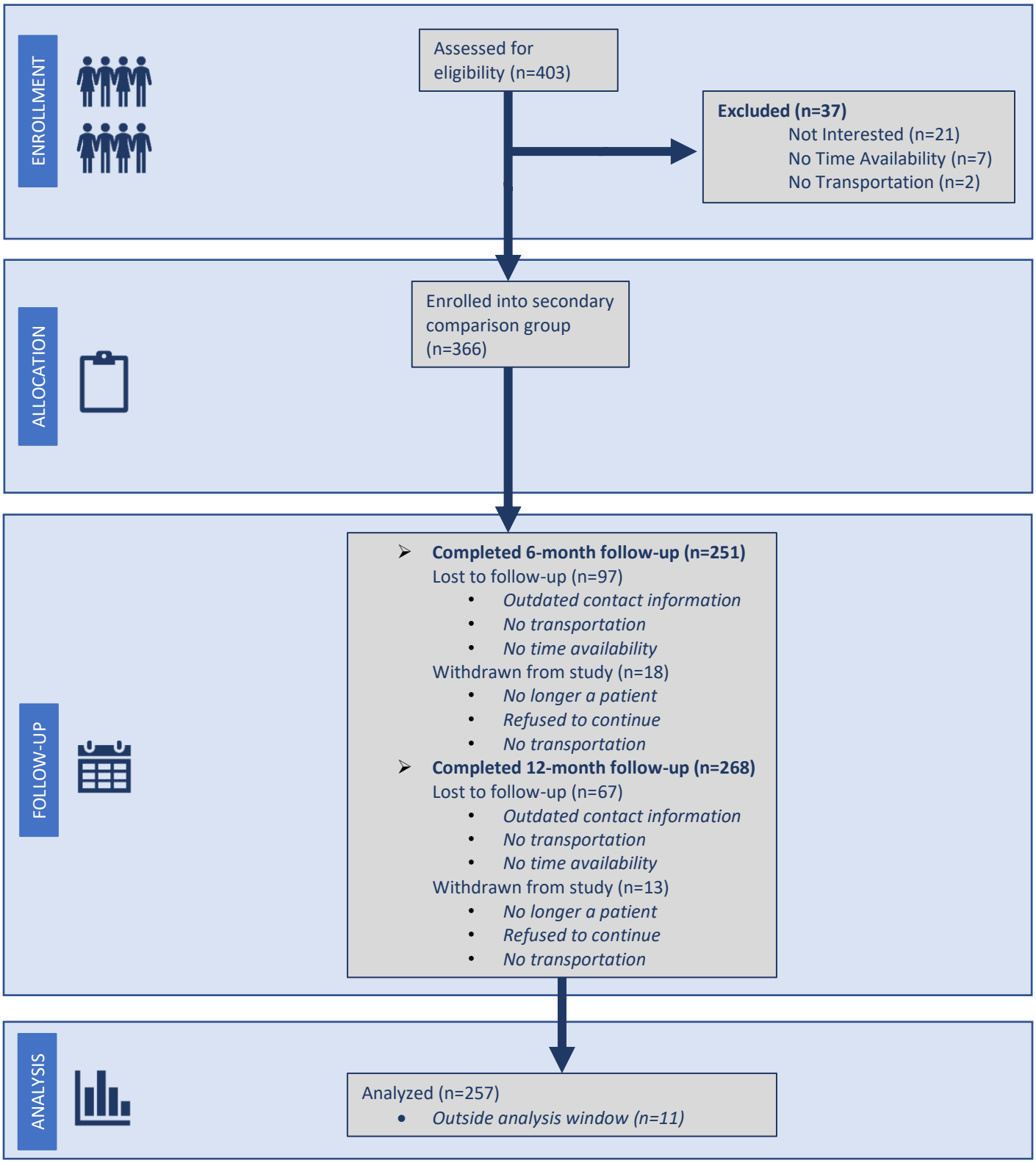


Figure 2. Patient Flow Description, Secondary Comparison



Sample Recruitment, Retention, and Attrition

Participant Eligibility and Recruitment

Patients for the Sí Three program were recruited from new and existing patients through Mercy's clinic and mobile van sites. At the time of the patient's clinic visit, the medical office assistant (MOA) took vital signs (height, weight, BP, waist circumference), and the care coordinator presented the surveys to the patient and patient self-administered the Sí Three surveys (instruments that measure depression, anxiety, quality of life, spirituality, and addiction). Patients at the mobile van site were given the surveys by the promotora, seen by the nurse practitioner, and referred to the clinic care coordinator for enrollment. Patients were then handed off from the care coordinator (with their assessments) to the program manager (navigator/NP) to discuss eligibility for the program. During the enrollment period, Mercy screened all adult patients for hypertension, obesity, diabetes, depression, anxiety, quality of life and/or addiction. Patients who met any of the following eligibility criteria were informed of the Sí Three program and offered an opportunity to participate in the program:

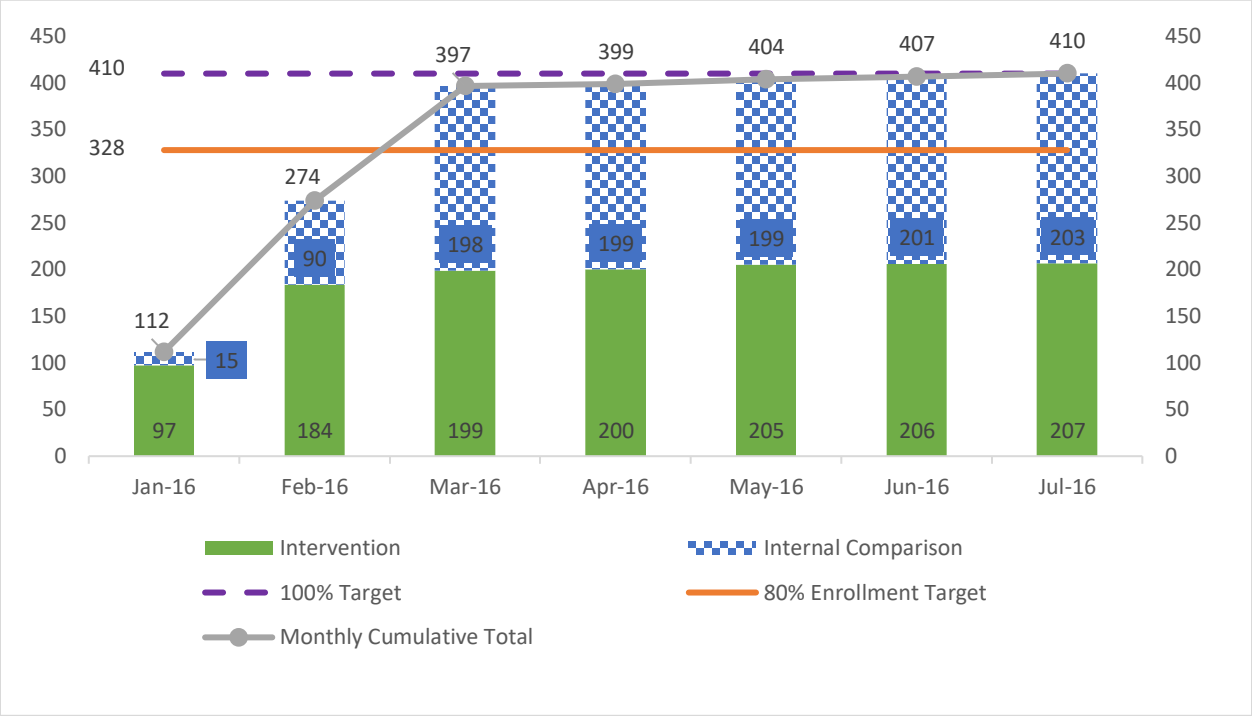
- PHQ-9 ≥ 5
- GAD-7 ≥ 5
- CAGE-AID ≥ 2
- Waist circumference ≥ 40 in men and ≥ 35 in women
- BMI ≥ 30
- Blood Pressure ≥ 140/90
- A1C ≥ 7.0%

If the eligible patient chose to participate, the program manager conducted the informed consent process. Consent procedures included explanation of the study and answering all questions that the participant had at the time of enrollment. The navigator read the consent form aloud to prospective participants, making sure they understood what participation entailed and their rights as participants.

Sample Enrollment and Retention

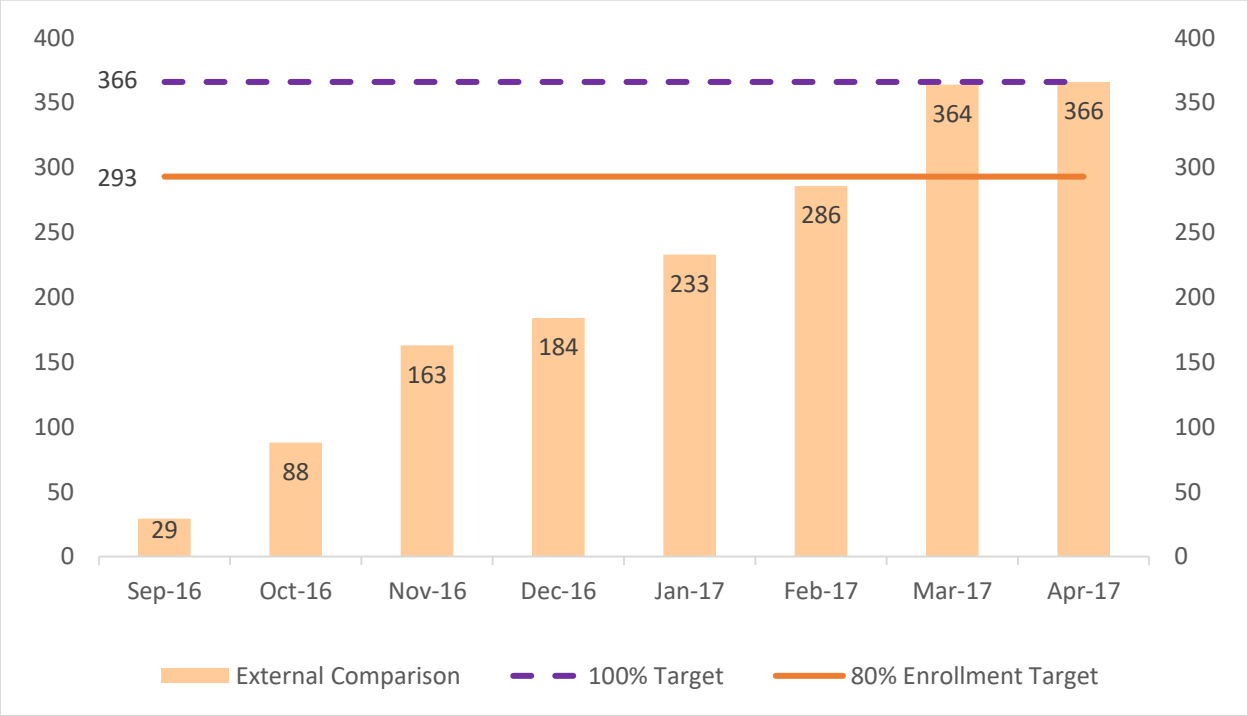
Participant enrollment began in January 2016 and continued through July 2016. This was a deviation from the planned timeline in the SEP, which presented an initial enrollment end date of March 2016. This change was to provide additional time to achieve initial enrollment targets. Enrollment extended past the original timeframe due to a revised sample size calculation during the SEP-approval and development phase (April-August 2016) and high interest of the target population in participating in the intervention rather than primary comparison group. The final timeline is presented in Appendix A: Revised Project Timeline. The enrollment target was 410 participants total across the intervention and primary comparison groups. Mercy successfully met the enrollment target by enrolling 410 into their intervention and primary comparison groups combined (see **Figure 3**). In terms of baseline study group enrollment, Mercy met their enrollment target for the primary comparison group (207 chose to be in the primary comparison group) and achieved 99% target enrollment for the intervention group (203 chose to be a part of the intervention group).

Figure 3. Cumulative Baseline Enrollment for Intervention and Primary Comparison Groups



Enrollment in the secondary comparison group at the NCDV clinics began in September 2016 and concluded in April 2017. The enrollment target at the clinic was 366 participants for the secondary comparison group, which Edcouch and Alton met exactly (see **Figure 4**). This inclusion of both Edcouch and Alton clinic patients is a deviation from the SEP. Patients from the Alton clinic were enrolled as members of the secondary comparison group due to concerns that the Edcouch clinic alone would not reach the enrollment target. These Alton clinic patients who were included in the secondary comparison pool had the required impact measures collected, aligning with Mercy's intervention group data collection.

Figure 4. Cumulative Baseline Enrollment for Intervention and Secondary Comparison Group



For 6-month follow-up data collection, Mercy collected data starting from 60 days before a participant's 6-month enrollment anniversary date up through 60 days after the anniversary date. A similar follow-up window was implemented for 12-month data collection. These data collection windows were developed during the evaluation study after the SEP was approved. Mercy began assessing participants for their 6-month follow-up assessments in June 2016 and completed the follow-up assessments in January 2017. Twelve-month follow-up assessments were collected between November 2016 and July 2017.

Table 17 presents subgrantee-reported information on the number of participants who returned for 6-month and 12-month follow-up through February and July 2017 respectively, by study arm. Mercy retained 91.8% of the 6-month target in the intervention group (169 out of 207 returned for a 6-month follow-up assessment, 184 needed to maintain power). The retention rate in the intervention group decreased further from 6-month follow-up, meeting 86.6% of the 12-month retention target (142 out of 207 returned for a 12-month follow-up assessment, 164 needed to maintain power). The primary comparison group reached 77.7% of the 6-month retention target (143 out of 203 returned for a 6-month follow-up assessment, 184 needed to maintain power). The retention improved for the primary comparison group at 12 months, with Mercy retaining 92.1% of the 12-month target (151 out of 203 returned for a 12-month follow-up assessment, 164 needed to maintain power). The final sample for all groups was 293 participants, 89.3% of the targeted 328 participants for sufficient power.

Table 17. Final Assessment of Follow-up Retention at 6 and 12 Months for Intervention and Primary Comparison Group

Group	Number Enrolled	Retention Target (assumes 10% attrition at 6 months and-20% attrition at 12 months)	Number Retained (i.e., completed assessment at 6 or 12 months)	Percent of Retention of the Enrolled Sample	Percent of Retention Target
6-month Retention					
Intervention Group	207	184	169	81.6%	91.8%
Primary Comparison Group	203	184	143	70.4%	77.7%
Total Sample	410	368	312	76.0%	84.7%
12-month Retention					
Intervention Group	207	164	142	68.6%	86.6%
Primary Comparison Group	203	164	151	74.4%	92.1%
Total Sample	410	328	293	71.4%	89.3%

Table 18 below describes the retention of the secondary comparison group. While the retention targets were set at a level to meet the needs of all three, shared comparison Sí Texas subgrantees, these targets exceeded the number of participants necessary for sufficient power in Mercy's secondary comparison group analysis by 58% (257 retained at 12 months, 164 needed).

Table 18. Final Assessment of Follow-up Retention at 6 and 12 Months for Secondary Comparison Groups

Group	Number Enrolled	Retention Target (assumes 10% attrition at 6 months and-20% attrition at 12 months)	Number Retained (i.e., completed assessment at 6 or 12 months)	Percent of Retention of the Enrolled Sample	Percent of Retention Target
6-month Retention					
Secondary Comparison Group	366	330	257	70.2%	77.9%
12-month Retention					
Secondary Comparison Group	366	293	257	70.5%	87.7%

Mercy used several study retention strategies to ensure sufficient power for analyses. First, the Sí Three care coordinator collected as many contact methods as possible from the study participant (for both intervention group and primary comparison group) during the enrollment process. Second, Mercy managed follow-ups using the patient care coordinator. The patient care coordinator contacted study participants on a monthly basis using any and all means of communication to reach participants, including telephone, text, voicemail, mail, and home visits. The care coordinator also used her

relationship with participants to locate and remind participants of their follow-up appointments. Appointments for study follow-up for both the intervention group and the primary comparison group were made for the same day as scheduled primary care, behavioral health care or other on-site appointments to minimize the number of return trips to the clinic for study participants. All intervention and primary comparison group patients were scheduled for quarterly follow-up clinic visits.

The Mercy SEP specifies data collection at baseline, 6 months, and 12 months. To ensure clinical needs were met by participants as patients at the clinic, MHM and the external evaluator provided additional guidance after SEP approval to Mercy that 6- and 12-month assessments could be collected within a 60-day window from anniversary date of the baseline assessment date.

This strategy accounted for normal variation in participant utilization of clinical services. The population served by Mercy faces many challenges in obtaining health care, so Mercy staff focused on ensuring clinical needs were met by participants as a priority. Assessments were scheduled to coincide with scheduled appointments to minimize burden on participants and maximize retention.

NCDV retained comparison group participants through a variety of strategies, which started with providers having a strong rapport with patients in a small clinic setting. In addition, specific staff were hired to work primarily on the Sí Texas project and more specifically the evaluation component. One of these staff scheduled data collection to coincide with existing provider appointments whenever possible. For those study participants who did not have a provider appointment during the data collection window, the staff made a specific appointment for those participants to come in for data collection. Staff contacted participants with an appointment reminder call during the week of the appointment. Also, comparison group participants received incentives in the form of a gift card of $10 at baseline, $15 at 6 months and $25 at 12 months. The amount of the incentive increased over time to offset the greater likelihood of attrition at later data collection dates (Grady, 2005) and to offset costs to patients.

Sample Attrition Analyses – Intervention and Primary Comparison Groups

The attrition for both the intervention and primary comparison group was greater than anticipated. The study anticipated 80% retention of the sample at 12 months. At the end of the study, there was an overall retention rate of 71%, with 74% in the primary comparison group and 69% in the intervention group. To examine whether this 5% difference was statistically significant, a Chi-square test was performed comparing the proportion of participants who were lost to follow-up in the intervention to those who were lost to follow-up in the primary comparison group. The results of this analysis were not statistically significant at the 0.05 level (p=0.22). Given these results, the two study groups did not have significantly differing attrition rates after 12 months of follow-up.

Of the Mercy participants who did not complete the study, there was a subpopulation who were unable to continue due to change in eligibility criteria (e.g. pregnancy, insurance status, or location – see **Table 19**). The remaining participants who did not complete the study were either lost to follow-up, completed their assessment outside the allowed follow-up windows, or withdrew for reasons not related to eligibility. **Table 20** presents the attrition data based on these categories. A bivariate comparison was conducted comparing the two different attrition groups to one another. The statistically non-significant results (p=0.42) indicate that the proportion of people who did not complete for eligibility reasons and those with different reasons are similar in the intervention and primary comparison groups. An additional bivariate comparison was run comparing these two attrition groups as

well as those who completed the study. The statistically non-significant results (p=0.31) indicate that the proportions in each group were similar in the intervention and primary comparison groups.

Table 19. Total Number of Participants Not Completing Study for Eligibility-based Reasons

	Intervention (n=26)	Primary Comparison (n=18)	Total (n=44)
Changed Study Group	0	1	1
Changed Provider	1	1	2
Exceeded Income	3	0	3
Moved Out of Town	5	5	10
Obtained Insurance	10	7	17
Pregnancy	7	4	11

Table 20. Breakdown of Study Attrition for Intervention and Primary Comparison Group

	Intervention n=207		Primary Comparison n=203		Total n=410	
	N	%	N	%	N	%
Completed Study	142	69%	151	74%	293	71%
Did Not Complete Study	65	31%	52	26%	118	29%
Withdrawn (change in eligibility)	26	13%	17	9%	43	11%
Lost to Follow-up or Withdrew[a]	39	19%	35	17%	74	18%

[a] Figure 1 provides reasons for withdrawal from the study.

Mercy used a variety of strategies to retain intervention and primary comparison group study participation. These strategies included making follow-up appointments during a visit, collecting and using multiple methods to contact participants to remind them of follow-up assessments, and providers and clinic staff developing a strong rapport with all clinic patients.

Although the difference in attrition between groups is not a concern for bias in the end-point analyses, the overall attrition rate was higher than anticipated. To explore the potential influence this may have had on results, bivariate analyses were conducted to examine whether participants who were lost to follow-up were significantly different than those who remained in the study, across all participants in the intervention and both comparison groups and within each study arm across demographic characteristics and baseline health measures. T-tests were used for continuous measures and Chi-square tests for categorical data. Fisher's Exact Test was utilized if the expected cell counts were less than 5 and non-parametric tests were performed on non-normally distributed data. Appendix G: Loss to Follow-Up/Attrition Tables presents all the results from these analyses.

Some differences were found and are detailed in the subsections below. A logistic regression model was then utilized to understand the influence of these differences in predicting a participant's likelihood to drop out of the study. In this model, intervention status did not have a statistically significant influence on the likelihood of being lost to follow-up. Of the differences detected, only age significantly predicted the likelihood of being lost to follow-up. Because the difference in attrition rates were not significant between groups and this characteristic was balanced at both baseline and 12 months, attrition bias is not of concern in interpreting the results of the study.

According to Mercy staff, as the health of participants improved, their financial situations improved through obtaining jobs or insurance, or they experienced other life events, such as becoming pregnant. Those who obtained insurance or became pregnant were no longer eligible to participate in the study and those who obtained jobs were less likely to have time to participate in the study.

Intervention Group

When examining the differences in health measures at baseline between those who were lost to follow-up and those who remained in the study at 12 months in the intervention group, there was a statistically significant difference in blood pressure. Those who did not complete the study had lower blood pressure at baseline measurement than those who remained in the study.

Primary Comparison Group

There were no statistically significant differences in health measures at baseline between those who were lost to follow-up and those who remained in the study at 12 months in the primary comparison group. Demographically, there were statistically significant differences in age and language in the primary comparison group; participants who did not complete the study were younger than those who did, and a higher proportion of the English speakers did not complete the study.

Intervention and Primary Comparison Group

While the difference in PHQ-9 score for those who completed the study and those who did not was not statistically significantly different when analyzing the primary comparison and intervention groups separately, when examining the differences by attrition status in the overall sample, those who did not complete the study were found to have a higher PHQ-9 score at baseline than those who completed the study.

Sample Attrition Analyses – Intervention and Secondary Comparison Group
At the end of the study, there was a retention rate of 71% in the secondary comparison group and 69% in the intervention group. To examine whether this 2% difference was statistically significant, a Chi-square test was performed comparing the proportion of participants who were lost to follow-up in the intervention group to those who were lost to follow-up in the secondary comparison group. The results of this analysis were not statistically significant at the 0.05 level (p=0.69). Given these results, the two study groups did not have significantly differing attrition rates after 12 months of follow-up.

Although the difference in attrition between groups is not a concern for bias in the end-point analyses, the overall attrition rate was higher than anticipated. To explore the potential influence this may have had on results, bivariate analyses were conducted to examine whether participants who were lost to follow-up were significantly different than those who remained in the study, for the entire sample and within each study arm across demographic characteristics and baseline health measures. T-tests were used for continuous measures and Chi-square tests for categorical data. Fisher's Exact Test was utilized if the expected cell counts were less than 5 and non-parametric tests were performed on non-normally distributed data**.** Appendix G: Loss to Follow-Up/Attrition Tables presents all the results from these analyses.

Some differences were found and are detailed in the subsections below. A logistic regression model was then utilized to understand the influence of these differences in predicting a participant's likelihood to withdraw or be withdrawn from the study. In this model, intervention status did not have a statistically significant influence on the likelihood of being lost to follow-up. Of the differences detected, only sex significantly predicted the likelihood of being lost to follow-up. Because the difference in attrition rates was not significant between groups, attrition bias is not of concern in interpreting the results of the study.

Secondary Comparison Group

There were no statistically significant differences in health measures at baseline between those who were lost to follow-up and those who remained in the study at 12 months in the secondary comparison group. Demographically, there was a statistically significant difference in sex in the secondary comparison group; those who did not complete the study were more likely to be male.

Intervention and Secondary Comparison Group

When looking at the differences between participants with differing attrition status in the overall sample of the intervention and secondary comparison group, there were no demographic differences, but those who did not complete the study had lower blood pressure than those who did.

Non-Response Bias and Missing Data

All data collected for the Mercy evaluation were recorded in Mercy's electronic medical record system which was enhanced to include all of the survey instruments used in the Sí Three evaluation study, including the Duke Health Profile. The EPIC team built the survey tools into the patient care platform so that data could be retrieved. Weekly meetings with the EPIC team prior to starting the project was key to prepare the platform and do practice data runs. Close communication with the EPIC team made the applications and modifications seamless. The data clerk was able to attend all of the meetings along with the project manager and prevent the missingness of data.

Missing data on covariates is a potential issue that could lead to biased results. The data collection team made all efforts to minimize missing data through training and use of standard practice measures within the clinic settings captured by the EMR. Imputation approaches were noted as an option if there were missing data on important covariates (Rubin, 1996). However, the data collected and submitted by Mercy were largely complete and therefore multiple imputation methods were not used in any analyses of Mercy's data.

Regarding the nine study impact measures for the primary end-point analysis, complete baseline data were collected for all Mercy participants for each measure except for blood pressure and waist circumference. There was only 1 participant missing blood pressure and there were 24 participants missing waist circumference at baseline. Early in the project waist circumference was missing for some of the patients on the mobile van due to confusion between van staff and clinic staff. Staff addressed the confusion and van staff collected waist circumference on the remaining mobile van patients. There was minimal missing demographic data. All demographic measures had complete data collected at baseline except for marital status (n=1), alcohol consumption (n=9), and perceived spiritual strength (n=56). Both the missing marital status data point and those missing for alcohol consumption were

reported as "unknown" and recoded to missing. The missing data for perceived spiritual strength was due to this measure being assessed by the LPC and only for participants that had been referred for anxiety and/or depression. This measure was only collected for the intervention participants and is not used in any final analyses of study data. At 12-month follow-up the following data were missing: 2 participants' blood pressure, 6 participants' BMI, 10 participants' waist circumference, 3 participants' PHQ-9 score, 3 participants' CAGE-AID, 3 participants' GAD-7 score, and 3 participants' Duke Health Profile scores.

There was no missing impact measure data for the secondary comparison group; however, there were 6 participants missing alcohol consumption and 3 participants missing marital status data. These missing data are due to responses of "unknown" in the NCDV clinic data system being recoded as missing for analysis.

Measures

The measures for the impact analysis aligned with the measures presented in the logic model depicted in Appendix B: Program Logic Model. The impact measures assessed for the Sí Three intervention program participants and the primary comparison group were HbA1c, blood pressure, BMI, depression score, anxiety score, addiction status, waist circumference, and quality of life measured through the DUKE Health Profile. These measures were the same for the secondary comparison group except for addictive behavior. There were no changes to the measures described in Mercy's SEP and interim report. Information on the number of respondents and tests of normality are provided here (see **Table 21** and **Table 22**). PROC UNIVARIATE in SAS was used to understand the distributions of these measures at baseline. QQ plots and histograms were used to determine if the measure should be treated as normal, be transformed, or treated as non-normal data. Descriptive statistics for each of these measures, including number of participants with or without the impact measures, are included in this final report.

Table 21. Impact Measure Sample Size by Follow-up: Intervention and Primary Comparison Group

Measure	Sample Size		
	Baseline	6-month	12-month
HbA1c	190	172	167
Systolic Blood Pressure	409	305	293
Diastolic Blood Pressure	409	305	293
BMI	410	303	289
PHQ-9	410	289	290
GAD-7	410	289	290
Duke Health Profile	410	289	290
Waist Circumference	394	294	285
CAGE-AID	411	289	290

Table 22. Impact Measure Sample Size by Follow-up: Intervention and Secondary Comparison Group

Measure	Sample Size		
	Baseline	6-month	12-month
HbA1c	465	346	342
Systolic Blood Pressure	573	413	399
Diastolic Blood Pressure	573	413	399
BMI	573	414	396
PHQ-9	573	415	397
GAD-7	573	415	397
Duke Health Profile	573	415	397
Waist Circumference	552	408	395

Nine clinical impact measures were measured during this study:

HbA1c: HbA1c levels are routinely measured in the monitoring of people with diabetes. HbA1c levels depend on the blood glucose concentration. The higher the glucose concentration in blood, the higher the level of HbA1c. Levels of HbA1c are not influenced by daily fluctuations in the blood glucose concentration but reflect the average glucose levels over the prior six to eight weeks. Therefore, HbA1c is a useful indicator of how well the blood glucose level has been controlled in the recent past (over two to three months) and may be used to monitor the effects of diet, exercise, and drug therapy on blood glucose in people with diabetes (American Diabetes Association, 2014).

It is Mercy's clinical practice only to recommend and subsequently collect HbA1C tests among patients who are: (1) known/self-reported to be diabetic, (2) have an elevated blood glucose at time of clinic visit or are suspected to be diabetic through other signs and symptoms. Therefore, only a portion of enrolled patients had HbA1C data available. Patients with an HbA1c greater than or equal to 7.0% were referred to the nurse educator, dietician and exercise coach. In addition, the navigator/NP determined the need/appropriateness of medication.

For HbA1c, there were 190 respondents with complete data at baseline, 172 respondents at 6 months, and 167 respondents at 12 months for the intervention and primary comparison group. There were 465 participants with complete data at baseline, 346 at 6 months, and 342 at 12 months for the intervention and secondary comparison group. The distribution of responses for HbA1c at baseline was determined to be non-normally distributed in the intervention and primary comparison sample as well as the intervention and secondary comparison sample. The log transformation was examined but did not normalize the distribution of HbA1c. Therefore, nonparametric tests were used in bivariate analyses.

Blood Pressure: Blood pressure is usually expressed in terms of the systolic pressure over diastolic pressure and is measured in millimeters of mercury (mm Hg). Blood pressure varies depending on situation, activity, age, and disease states (American Heart Association, 2015).

Blood pressure was measured by the MOA, manually using a stethoscope and sphygmomanometer and following clinically-established practice guidelines (National Guidelines Clearinghouse, 2011). Patients with a blood pressure greater than or equal to 140/90 were referred to the nurse practitioner, nurse educator, dietician, and exercise coach. In addition, the navigator/NP determined the need/appropriateness of medication.

For blood pressure, there were 409 respondents with complete data at baseline, 305 respondents at 6 months, and 293 respondents at 12 months in the intervention and primary comparison groups. There were 573 participants with complete data at baseline, 413 at 6 months, and 399 at 12 months for the intervention and secondary comparison groups. The distributions of responses for systolic and diastolic at baseline were both determined to be normally distributed in the intervention and primary comparison sample as well as the intervention and secondary comparison sample.

Body Mass Index (BMI) and Waist Circumference: BMI is a confirmatory outcome in this study. It is generally used as an indicator of body fat. The MOA (medical office assistant) recorded the patient's height and weight in the EPIC EMR and EPIC calculated patient's BMI. Patients with a BMI greater than or equal to 30 were referred to the nurse practitioner, nurse educator, dietician, and exercise coach.

For BMI, there were 410 respondents with complete data at baseline, 303 respondents at 6 months, and 289 respondents at 12 months in the intervention and primary comparison groups. There were 573 participants with complete data at baseline, 414 at 6 months, and 396 at 12 months for the intervention and secondary comparison groups. The distribution of responses for BMI at baseline was determined to be slightly skewed in the intervention and primary comparison sample as well as the intervention and secondary comparison sample. Using the log transformation of the BMI data for bivariate analyses led to a more normal distribution and therefore the parametric test was used.

Waist circumference can be used to estimate potential disease risk. Waist circumference was measured by the MOA manually using an appropriate length tape measure following clinically-established practice guidelines (National Guideline Clearinghouse, 2014). Patients with a waist circumference greater than or equal to 40 for males and greater than or equal to 35 for females were referred to the navigator/NP for review and referral to the nurse educator, dietician and exercise coach.

For waist circumference, there were 394 respondents with completed data at baseline, 294 respondents at 6 months, and 285 respondents at 12 months in the intervention and primary comparison groups. There were 552 respondents with completed data at baseline, 408 respondents at 6 months, and 395 respondents at 12 months in the intervention and secondary comparison groups. The distribution of responses for waist circumference at baseline was determined to be normally distributed in the intervention and primary comparison sample as well as the intervention and secondary comparison sample.

Depression: Depression is characterized by depressed or sad mood, diminished interest in activities which used to be pleasurable, weight gain or loss, psychomotor agitation or retardation, fatigue, inappropriate guilt, difficulties concentrating, as well as recurrent thoughts of death. Diagnostic criteria established by the American Psychiatric Association dictate that five or more of the above symptoms must be present for a continuous period of at least two weeks. In addition to being a chronic disease in its own right, the burden of depression is further increased as depression appears to be associated with behaviors linked to other chronic diseases (American Psychiatric Association, 1994).

- **Administration method:** Depression was measured using the self-administered PHQ-9 assessment tool, distributed and collected by the care coordinator. The PHQ-9 is a multipurpose instrument for screening, diagnosing, monitoring and measuring the severity of depression.
- **Administration time:** The PHQ-9 was given to patients as part of their assessment.
- **Intended respondent:** The PHQ-9 was administered to all adult patients who visited the clinic.
- **Potential score/response range:** The PHQ-9 has a total possible score of 27. The PHQ-9 scoring criteria are categorized as minimal (0-4), mild (5-9), moderate (10-14), moderately severe (15-19) and severe (20-27) depression (Kroenke & Spitzer, 2002). Patients with a score of 5 or higher were referred for secular and/or faith-based behavioral health services. Please note, faith-based services (to be discussed below) were offered as an option to those patients with a Spirituality Index score of 50 or greater.

PHQ-9 score is a confirmatory outcome in this study. There were 411 respondents with complete data at baseline, 289 respondents at 6 months, and 290 respondents at 12 months in the intervention and primary comparison groups. There were 573 participants with complete data at baseline, 415 at 6 months, and 397 at 12 months for the intervention and secondary comparison groups. The distribution of responses for PHQ-9 at baseline was determined to be non-normally distributed in the intervention and primary comparison sample as well as the intervention and secondary comparison sample. The log

transformation was examined but did not normalize the distribution of PHQ-9. Therefore, nonparametric tests were used in bivariate analyses.

Anxiety: Anxiety disorders are characterized by excessive and unrealistic worry about everyday tasks or events or may be specific to certain objects or rituals. In addition to being helped by pharmacotherapies, anxiety disorders are often treated by behavioral approaches.

- **Administration method:** Anxiety was measured via the self-administered GAD-7 assessment tool, distributed and collected by the care coordinator. The GAD-7 is a valid and efficient tool for screening for anxiety and assessing its severity in clinical practice and research (Sadock & Sadock, 2007).
- **Administration time:** The assessment was given to the patient as part of their assessment.
- **Intended respondent:** The GAD-7 was administered to all adult patients who visited the clinic.
- **Potential score/response range:** For the GAD 7, of a possible total point value of 21, anxiety is measured as mild (5-9), moderate (10-14) or severe (15 or above) (Spitzer, Kroenke, Williams, & Löwe, 2006). Patients with a score of five (5) or higher were referred for medical and/or faith-based services. Please note, faith-based behavioral health services (to be discussed below) were offered as an option to those patients with a Spirituality Index score of 50 or more.

For GAD-7 score, there were 410 respondents with complete data at baseline, 289 respondents at 6 months, and 290 respondents at 12 months in the intervention and primary comparison groups. There were 573 participants with complete data at baseline, 415 at 6 months, and 397 at 12 months for the intervention and secondary comparison groups. The distribution of responses for GAD-7 score at baseline was determined to be non-normally distributed in the intervention and primary comparison sample as well as the intervention and secondary comparison sample. Therefore, nonparametric tests were used in bivariate analyses.

Addiction: Addiction refers to excessive harmful or hazardous behavior, including use of psychoactive substances, alcohol, as well as gambling, sex, and food, etc. Addiction can lead to dependence syndrome - a cluster of behavioral, cognitive, and physiological phenomena that develop after repeated abuse and that typically include a strong desire to continue the behavior, difficulties in controlling it, persisting despite harmful consequences, a higher priority given to that behavior than to other activities and obligations, increased tolerance, and sometimes a physical withdrawal state.

- **Administration method:** Addiction was measured via the self-administered CAGE-AID assessment tool, distributed and collected by the care coordinator. The CAGE is an internationally recognized assessment instrument developed by Dr. John Ewing to identify alcoholism. CAGE-AID is an adapted version for use in clinical settings to include a wide range of addictive issues as part of a general health exam (Brown & Rounds, 1995). The CAGE-AID is a conjoint questionnaire where the focus of each item of the CAGE questionnaire was expanded from alcohol alone to include alcohol and other drugs. CAGE-AID questions ask about use, motivation to reduce substance use, and perceptions of use.
- **Administration time:** The assessment was given to the patient as part of their assessment.
- **Intended respondent:** The CAGE-AID was administered to all adult patients who visited the clinic.
- **Potential score/response range:** The four items of the CAGE-AID are scored zero (0) or one (1), and a finding of two (2) is clinically significant. As an option to those patients with a score of two (2) or higher, behavioral health services were offered by Mercy and other partners. Please note, faith-based behavioral health services (to be discussed below) were offered as an option to

those patients with a Spirituality Index score of 50 or more. The CAGE-AID assessment was not administered at the Edcouch clinic as clinic administrators were not comfortable assessing addictive behaviors due to a lack of internal or regional capacity to address any addiction issues that may be diagnosed.

For CAGE-AID score, there were 410 respondents with complete data at baseline, 289 respondents at 6 months, and 290 respondents at 12 months in the intervention and primary comparison groups. The distribution of responses for CAGE-AID score at baseline was determined to be non-normally distributed. Therefore, nonparametric tests were used in bivariate analyses.

Quality of Life (QOL): QOL is a broad multidimensional concept that usually includes subjective evaluations of both positive and negative aspects of life. Health serves as one of several domains for overall QOL. Aspects of culture, values, and spirituality are also key aspects of overall quality of life that add to the complexity of its measurement (CDC, 2011).

- **Administration method:** Quality of life was measured via the self-administered Duke Health Profile, distributed and collected by the care coordinator. The Duke Health Profile instrument contains six health measures (physical, mental, social, general, perceived health, and self-esteem), and four dysfunction measures (anxiety, depression, pain, and disability) (Parkerson, Broadhead, & Tse, 1990).
- **Administration time:** The Duke Health Profile assessment tool was given to the patient as part of their assessment.
- **Intended respondent:** The Duke Health Profile assessment tool was administered to all adult patients who visited the clinic.
- **Potential score/response range:** The Duke Health Profile has 11 scales, five of which measure function (physical health, mental health, social health, general health, perceived health, self-esteem) and six of which measure dysfunction (anxiety, depression, anxiety-depression, pain disability). Scores range from 0 to 100. For scales measuring function, the higher the score, the more functional the person being evaluated. For scales measuring dysfunction, the higher the score, the more dysfunctional the person being evaluated. For the purposes of this report, data analysis includes general health, physical health, and social health scales.

For Duke General Health score, there were 410 respondents with complete data at baseline, 289 respondents at 6 months, and 290 respondents at 12 months in the intervention and primary comparison groups. There were 573 participants with complete data at baseline, 415 at 6 months, and 397 at 12 months for the intervention and secondary comparison groups. The distribution of responses for Duke General Health score at baseline was determined to be normally distributed in the intervention and primary comparison sample; therefore, parametric tests were used in bivariate analyses. This measure was determined to be non-normally distributed in the intervention and secondary comparison sample. The log transformation was examined but did not normalize the distribution of the Duke General Health score in this sample; nonparametric tests were used in bivariate analyses.

Spirituality: Spirituality is a broad concept with room for many perspectives. In general, it includes a sense of connection to something bigger than oneself, and it typically involves a search for meaning in life. As such, it is a universal human experience and is associated with overall well-being, quality of life and health outcomes (Daaleman, Frey, Wallace, & Studenski, 2002).

- **Administration method:** Spirituality was measured via the self-administered Spirituality Index of Well-Being assessment tool, distributed and collected by the care coordinator. The Index

(Daaleman et al 2002) is a valid and reliable instrument that can be used in health-related quality-of-life studies.

- **Administration time:** The Spirituality Index was given to the patient as part of their assessment.
- **Intended respondent:** The Spirituality Index was administered to all adult patients who visited the clinic.
- **Potential score/response range:** The Spirituality Index of Well-Being includes 12 items in two subscales, life scheme and self-efficacy. Total scores across the two subscales range from 12 to 60. Patients with a score of fifty (50) or higher on the Spirituality Index were offered the option of faith-based behavioral health services, as suggested by Daaleman et al., 2002, given that these patients are categorized as having high spirituality and thus would be more responsive to faith-based behavioral health services. These services were only offered to patients who scored 5 or greater on the PHQ-9 or GAD-7 and 50 or greater on the Spirituality Index of Well-Being. Spirituality was not assessed among participants at the Edcouch clinic, as the Spirituality Index was used at Mercy solely as a screening mechanism for determining the type of behavioral health services most appropriate for an intervention group patient at Mercy.

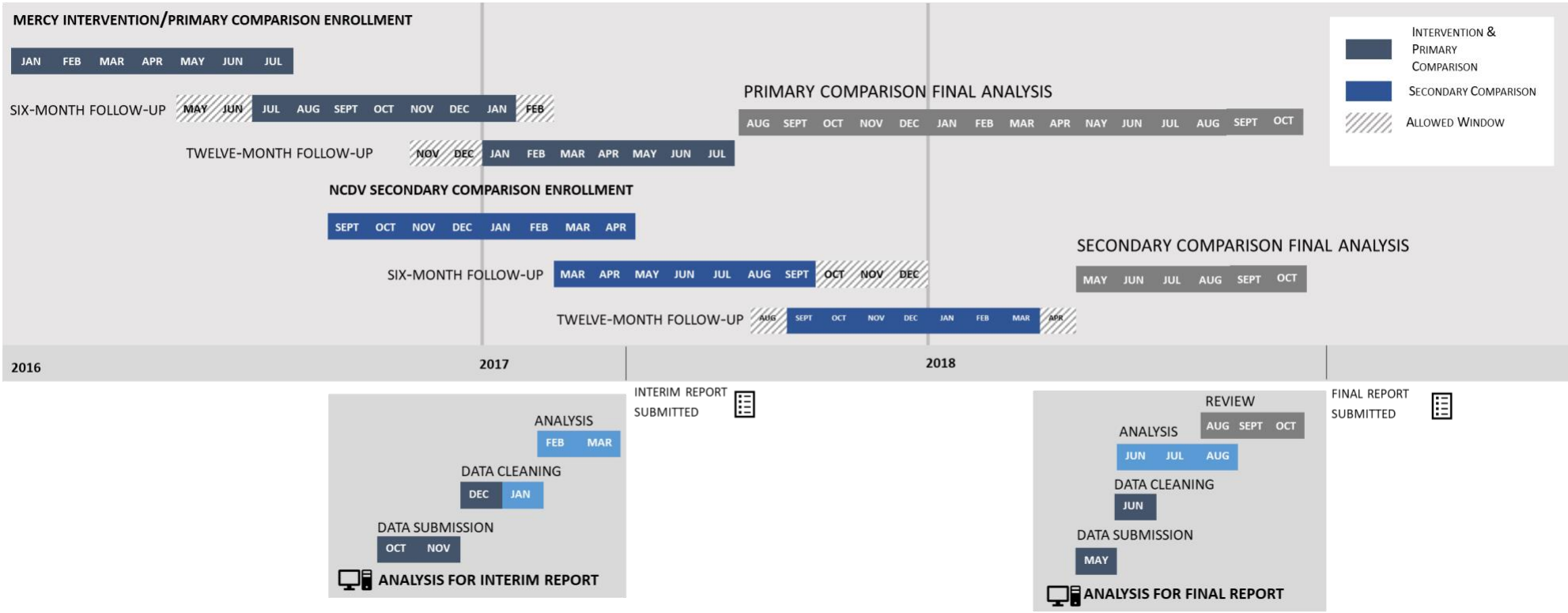
Data Collection Activities

Planned data collection activities were executed as described in the SEP without deviation. Baseline data collection for the intervention group, primary comparison group and secondary comparison group occurred at study enrollment. Within Mercy's clinic, the medical office assistant (MOA) and nurse navigators collected physical health measures on the intervention group and primary comparison group. The care coordinator distributed and collected the behavioral health assessments (PHQ-9, GAD-7, CAGE-AID and Duke Profile) and spirituality assessments. At the secondary comparison group clinics, the nurse collected data on all measures except addiction and spirituality.

Figure 5 depicts the data collection timeline as it relates to SEP approval and analyses completed for this final report. Mercy participant enrollment began in January 2016 and continued through July 2016. As previously noted, this was a deviation from the planned timeline in the SEP. This change was to provide additional time to achieve initial enrollment targets. Mercy began assessing participants for their 6-month follow-ups in June 2016 and completed follow-ups in January 2017. Twelve-month follow-ups began in December 2016 and concluded in August 2017.

Data from the study were submitted on a quarterly basis to HRiA by Mercy and then cleaned and assessed for quality. The data cleaning for these data required a manual process that is detailed later in this report.

Figure 5. Timeline for Data Collection and Analyses for the Final Report



IMPACT STUDY – ANALYSIS AND RESULTS

This section presents the final impact analyses and results for both comparisons groups included in this study. First presented are the results for the primary comparison group analyses. This group was designed to be the study's primary comparison group due to a greater ability to maintain internal validity with a comparison group from the same clinic population as the intervention group. Additionally, as this comparison group was derived from the same population, the primary comparison group was statistically equivalent with the intervention group at baseline on a majority of health and sociodemographic measures enabling unbiased comparisons of these two groups.

Following the results of the primary comparison group analyses, the results for the secondary comparison group are presented. As proposed in the SEP, this group was included in the design to enhance the external validity and generalizability of the primary comparison group results. However, the statistical nonequivalence of the intervention and the secondary comparison groups at baseline poses challenges. Because of this, analyses with the primary comparison group are considered the main study, while analyses with the secondary comparison group are included to enhance the external validity and generalizability of the primary comparison group results. The analyses between the intervention and secondary comparison group serve as sensitivity analyses aimed at an increased understanding of the intervention effects and how they may or may not differ when compared to a secondary comparison group under different conditions.

Primary Comparison Group Analysis and Results
Final impact study results for the intervention and primary comparison group are presented by research question. This section presents the primary results of this study and includes detail on the statistical methods used. Any deviations from what was planned in the SEP were based on field conditions and analytic judgment at the time of analysis.

Descriptive statistics for complete data are examined in this final report for the intervention and primary comparison group. These statistics include patients' sociodemographic characteristics and other key covariates. These covariates were examined to assist in identifying potential factors that may result in nonequivalence between the two groups. Chi-square tests, and Fisher's Exact Tests, when necessary based on cell counts, were used for comparison of categorical data to examine baseline equivalence. Two sample t-tests were used for continuous data that were normally distributed, and the Wilcoxon Signed Rank test was used for non-normally distributed data. Because a nonequivalent comparison group QED design is employed in the study, an intent-to-treat analysis was conducted with adjustment of potential nonequivalence of covariates and baseline impact measure. The decision was made not to perform secondary power calculations as the final sample size was just shy of the target and prior research indicated that these post-hoc power analyses are not necessarily helpful in the interpretation of observed results (Goodman and Berlin, 1994).

All descriptive, baseline equivalence, bivariate, multivariate, and longitudinal analyses reported in this final report were performed with SAS version 9.4 (Cary, NC). PROC GLM was utilized for the primary linear regression models. For impact measures that were assessed to be non-normally distributed, analyses were conducted using both PROC GLM and PROC GENMOD in order to assess any possible bias deriving from the non-normality. For linear regression models, using normal linear regression methods (e.g., PROC GLM) produced results consistent with those produced with methods accounting for the non-normality of these data (e.g., PROC GENMOD). Differences were considered statistically significant

at p<0.05. Effect sizes were calculated for both confirmatory outcomes regardless of statistical significance of model results and for any exploratory outcome with a statistically significant result. Results are presented in the "Findings" section under research questions when applicable. The statistic utilized for these calculations was Cohen's d using the following equation:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{\mu_1 - \mu_2}{s}$$

Unit of Analysis and Overview of Analyses Performed

The unit of analysis was the individual patient. An "end-point" analysis was the primary analytic approach. This "end-point" analysis approach is a conventional approach to analyze clinical trial data collected from individuals with both baseline data and end-point data of primary interest (Liebschutz, et al., 2017). Generalized regression analysis was used following a modeling sequence from bivariate models to multiple regression models adjusting for baseline levels of outcome measures and covariates assessed to be relevant based on review of the scientific literature or found to be unbalanced between the two groups at baseline. The parameter of interest was the dichotomous variable that differentiates the treatment status (i.e., intervention vs. primary comparison group). Between-group comparison of baseline and single follow-up outcomes were assessed by end-point analyses that accounted for the baseline level of impact measures. Additionally, because multiple follow-up impact measures form individual trajectories over the study period, longitudinal analyses were used to assess whether the impact measure trajectories differ by intervention status (Fitzmaurice et al., 2004). A time measure was developed and applied to denote baseline, 6 and 12-month follow-up measures.

In addition to adjusting for key covariates, potential collinearity and its impact on the standard error estimates for the covariates in the model was assessed by examining the variance inflation factor when necessary. A contingency plan was included in the SEP indicating that, in areas where multiple comparisons would be necessary, the p-value would be adjusted to account for multiple comparisons, such as the Bonferroni correction. This step was ultimately unnecessary for the executed analyses since there was no need to account for multiple comparisons. Also, the SEP included the potential use of generalized estimating equations to adjust for clustering by provider. However, this step was not taken in the executed analyses due to the structure of how services were provided within the study. All intervention participants saw the same group of providers and often saw more than one over the course of the study. The case is the same for the primary comparison group participants who saw the same clinical and different behavioral health providers than the intervention. Therefore, there was no concern about clustering by provider.

To evaluate the intervention effect, a multiple linear regression model approach was used following a sequence of interrelated models. The analysis sequence began by developing a bivariate model regressing the follow-up impact measure on intervention status (intervention vs. primary comparison group) followed by the estimation of an adjusted model accounting for the baseline measure of interest and further adjustment for key covariates. Parametric two sample t-tests were used for bivariate analysis of one of the confirmatory impact measures (BMI) as well as some of the exploratory study outcomes (blood pressure, waist circumference, and Duke General Health). One of the confirmatory variables (PHQ-9 score) and some exploratory outcomes (GAD-7, HbA1c, CAGE-AID) were found to be non-normally distributed. In these bivariate analyses, nonparametric Wilcoxon Rank Sum tests were conducted due to the increased sensitivity to detect a difference in non-normally distributed data. The

nonparametric results are presented throughout this report; however, additional parametric t-tests were performed for these measures to align with linear regression methods for the final analyses. Though the parametric results are not presented, both the nonparametric and parametric bivariate analyses produced consistent results.

Following bivariate comparisons, multivariate and longitudinal analyses were performed separately to answer each research question. As previously mentioned, multiple imputation methods were not necessary due to the complete nature of the submitted data. Propensity score matching was explored but was not included in the outcome analyses for the reasons described previously in the Methods section of this report. The primary adjusted multivariate analysis models the outcome of interest on intervention status with relevant covariates included. The longitudinal analysis evaluates whether the impact measure trajectories differ by intervention status across the 12-month study.

Effect modification of the intervention-outcome relationships were also examined. To explore the faith-based and spiritual elements of Mercy's services and the possible influence of the intervention's effectiveness on the health outcomes of interest, baseline Spirituality Index score was explored as a possible effect modifier. An interaction term was included in the models for each of the 12-month impact measures between study group and baseline Spirituality Index score, separated into quintiles. Quintiles, or five groups of equal sizes, were calculated to understand the possible effect modification of a change in the level of spirituality index a participant reports (i.e. lower, middle, or higher levels). Additionally, possible effect modification of baseline health condition was explored for the corresponding impact measure (e.g., baseline depression as an effect modifier for impact on PHQ-9 score at 12 months).

The SEP indicated a set of planned covariates for adjustment in the models. Of those listed, age (continuous and categorical), sex, employment, number of comorbidities, number of visits by type, and time were included in one or more of the analyses. Categorical age was operationally defined by the following categories: 18-24, 25-34, 35-44, 45-54, 55-64, and those who are 65 years or older. Employment was included as a dichotomized variable with categories of "employed", including employed and self-employed participants, and "not employed", including unemployed and student participants. As anticipated, the study population was fairly homogeneous on ethnicity thus this was not included in the final models. Additional data on characteristics of the study population--Spirituality Index score, primary language, smoking, alcohol consumption, and marital status--were included for possible selection in one or more of the analyses. Marital status was considered a dichotomous variable with categories "married", including only those who indicated they were married, and "not married", which includes all other categories for the marital status variable.

Raw data were cleaned manually prior to analysis using a systematic approach including quarterly verification of submitted data with study site staff. Data from Mercy's EMR could only be exported into a format that did not lend itself to immediate analysis. Assessment data for each participant spanned several rows of encounters which needed to be merged into a single row with columns for each variable that aligned with the time windows specified in the SEP. In many cases, Mercy entered input assessment data on a later day than the actual encounter. The EMR time stamped these data with the date of entry rather than the date of the encounter. Because the date stamp for an encounter could reflect a date after the actual date of assessment, the HRiA analyst manually reviewed each row to classify each encounter by assessment (e.g., determine which encounter belonged to the 12-month follow-up assessment). During the manual cleaning, each row was classified to a specific assessment (e.g. baseline,

6-month, 12-month) by its date. The first complete chronological encounter data was coded as the baseline assessment and defined the anniversary dates for the 6-month and the 12-month assessment. An encounter was assigned to the 6-month assessment if it occurred within the following window: 60 days before or 60 days after the true 6-month anniversary date. An encounter was assigned to the 12-month assessment if it occurred within the following window: 60 days before or after the true 12-month anniversary date. During the manual cleaning, cases that were on the line—meaning, that they occurred a few days outside the window—were examined to see if the projected anniversary encounter would have occurred on a weekend, holiday or other time when the clinic was closed. Those cases were recoded as 6-month or 12-month assessments instead of excluding the data point entirely. This re-coding exception occurred in less than 2% of participant cases.

A backward elimination modeling selection procedure was employed for the end-point analysis approach where covariates with a p-value larger than 0.15 were excluded from the final model for parsimony. In some cases, age and sex were selected for inclusion in statistical models a priori due to the known biological influence of these characteristics on health outcomes; this is noted where relevant under each research question. For some research questions, predictor variables were included that could be correlated with the outcome of interest. Where relevant, the variation inflation factor (VIF) is reported in the model selection process. Using PROC CORR, the range of correlation between the predictors included in the model and the outcomes of interest is -0.49, the Pearson coefficient for baseline PHQ-9 score and 12-month Duke General Health score, to 0.97, the Pearson coefficient for baseline BMI and 12-month BMI.

Results for each of the impact measures are presented separately by research question. At the end of this section, **Table 64** presents the mean scores for each of the outcome variables.

Depressive Symptoms

Question 1. **Do patients who participate in the Sí Three intervention experience improvements in depressive symptoms, as measured by PHQ-9, after 12 months compared to patients who do not participate? This question is confirmatory. In addition, do these improvements differ by type of behavioral-health service received (medical/behavioral or faith-based services)? This question is exploratory.**

Overview of Analysis

To answer this confirmatory question about intervention impact on depressive symptoms, data were collected using the PHQ-9 assessment tool. The exploratory question aimed at understanding whether there was effect modification by the type of behavioral health services received was not evaluated as the majority of intervention participants selected the faith-based services; therefore, there was no second group for stratified analyses. This is a deviation from the SEP; however, the decision was based on the data collected representing Mercy's clinical practice. The sample sizes for the presented analyses of PHQ-9 score are as follows: bivariate analyses (n=290), primary linear regression analyses (n=283), and longitudinal analyses (n=332). (Note: All intervention patients were offered faith-based services after talking with the patient and determining the patient's perceived spirituality even those patients whose Spirituality Index score was below 50, this was a deviation from the SEP).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean PHQ-9 data in each study period for the overall sample as well as the intervention and primary comparison groups. The overall sample had a mean PHQ-9 score of 5.5 at baseline. This decreased to 4.2 for participants who returned at 6-month follow-up and again to 2.7 for those who returned at 12-month follow-up. The intervention group had a higher PHQ-9 score of 6.7 at baseline compared to the primary comparison group mean score of 4.0, indicating higher levels of depressive symptoms. Across the study, for participants who completed a follow-up assessment, the intervention group mean PHQ-9 score improved to 4.9 at 6 months and 2.9 at 12 months. The primary comparison group also followed a similar improving trend in depressive symptoms with mean PHQ-9 score for those who completed a follow-up decreasing to 3.3 at 6 months and 2.5 at 12 months. As previously noted, the differences in PHQ-9 score between the intervention and primary comparison groups at baseline were statistically significant (**Table 9**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The improvements observed in PHQ-9 score from baseline to 12-month follow-up within both the intervention and primary comparison groups were found to be statistically significant.

Bivariate analyses were also performed between the intervention and primary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for PHQ-9 score when comparing the intervention and primary comparison group at 12 months, the null hypothesis cannot be rejected. The mean PHQ-9 score at 12-months was not significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, PHQ-9 score. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for PHQ-9 score were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline PHQ-9 score, baseline Duke General Health score, baseline GAD-7 score, the number of comorbidities at baseline, and baseline Spirituality Index score. The inclusion of baseline PHQ-9 score controlled for the statistical imbalance between intervention and primary comparison groups at baseline. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline BMI was included in the initial full model. This model was specified as follows:

$$Y_{(PHQ\text{-}9)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_PHQ\text{-}9 + \beta_{10} BL\_General + \beta_{11} BL\_GAD7 + \beta_{12} BL\_Comorbidities + \beta_{13} BL\_Spirituality + \beta_{14} BL\_BMI + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, a multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model for PHQ-9 score included those covariates with p-value of 0.15 or less: primary language, smoking, baseline PHQ-9 score, baseline Duke General Health score, baseline GAD-7 score, and baseline BMI. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(PHQ\text{-}9)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 Smoke + \beta_6 BL\_PHQ\text{-}9 + \beta_7 BL\_General + \beta_8 BL\_GAD7 + \beta_9 BL\_BMI + \varepsilon$$

Because baseline quality of life and anxiety measures were selected for inclusion in the final model of depressive symptoms, and quality of life and anxiety are known to be related to depressive symptoms, we conducted an additional test to quantify any multicollinearity between the Duke General Health and GAD-7 scores with PHQ-9 score. The variance inflation factor (VIF) of the Duke General Health score in the PHQ-9 score model was 2.2 and the VIF of the GAD-7 score was 3.1, both below the commonly accepted cutoff of 5, indicating minimal influence on the variance from the correlation of these variables (Belsley et al., 1980; O'Brien, 2007; Lasser, et al. 2017).

Findings

Estimates by covariate for the final model of PHQ-9 score are presented in **Table 23**.

Mean PHQ-9 score at 12 months did not differ significantly between the intervention and primary comparison group (p=0.06); the effect size (using Cohen's d) is 0.20. Below is the selected model with each covariate's effect estimate included:

$$Y_{(PHQ\text{-}9)} = 0.19 + -0.81(Intervention) + 0.02(Age) + -0.53(Male) + 1.27(English) + 2.18(Current\ Smoker) + 1.21(Former\ Smoker) + 0.19(BL\_PHQ\text{-}9) + -0.04(BL\_General) + 0.15(BL\_GAD7) + 0.08(BL\_BMI) + \varepsilon$$

Table 23. Effect of IBH Intervention on Twelve Month PHQ-9 Score, Intervention Compared to Primary Comparison Group

Variable	PHQ-9 (n=283)		
	Estimate (β)	Standard Error	p-value
Intervention	-0.81	0.43	0.06
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.02	0.02	0.24
Male[a]	-0.53	0.62	0.40
Female (ref)	--	--	--
English	1.27	0.70	0.07
Spanish (ref)	--	--	--
Current smoker	2.18	0.83	0.01
Former smoker	1.21	1.18	0.30
Never smoker (ref)	--	--	--
BL_PHQ-9	0.19	0.07	0.01
BL_General	-0.04	0.02	0.03
BL_GAD-7	0.15	0.07	0.04
BL_BMI	0.08	0.03	0.02

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).
[a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for PHQ-9 score (not shown). The models estimated included interaction terms between study group and baseline depression and baseline Spirituality Index quintile.

Additional Analyses

Longitudinal analyses were used to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate a linear mixed model accounting for multiple data points for each individual, the PROC MIXED procedure in SAS was used. There was a significant time by study group interaction for PHQ-9 score over the 12-month study (p<0.001), indicating that the PHQ-9 trajectories from baseline to 6 months, and then to 12 months were significantly different between the intervention group compared to the primary comparison group (**Table 23** and **Table 24**). For each 6 months in the study, the intervention group experienced a relative decrease of 1.76 points in PHQ-9 score compared to the primary comparison group. Adjusting for the covariates that were selected in the primary model—age, sex, language, smoking, baseline Duke General Health and GAD-7 scores—did not alter these results (not shown).

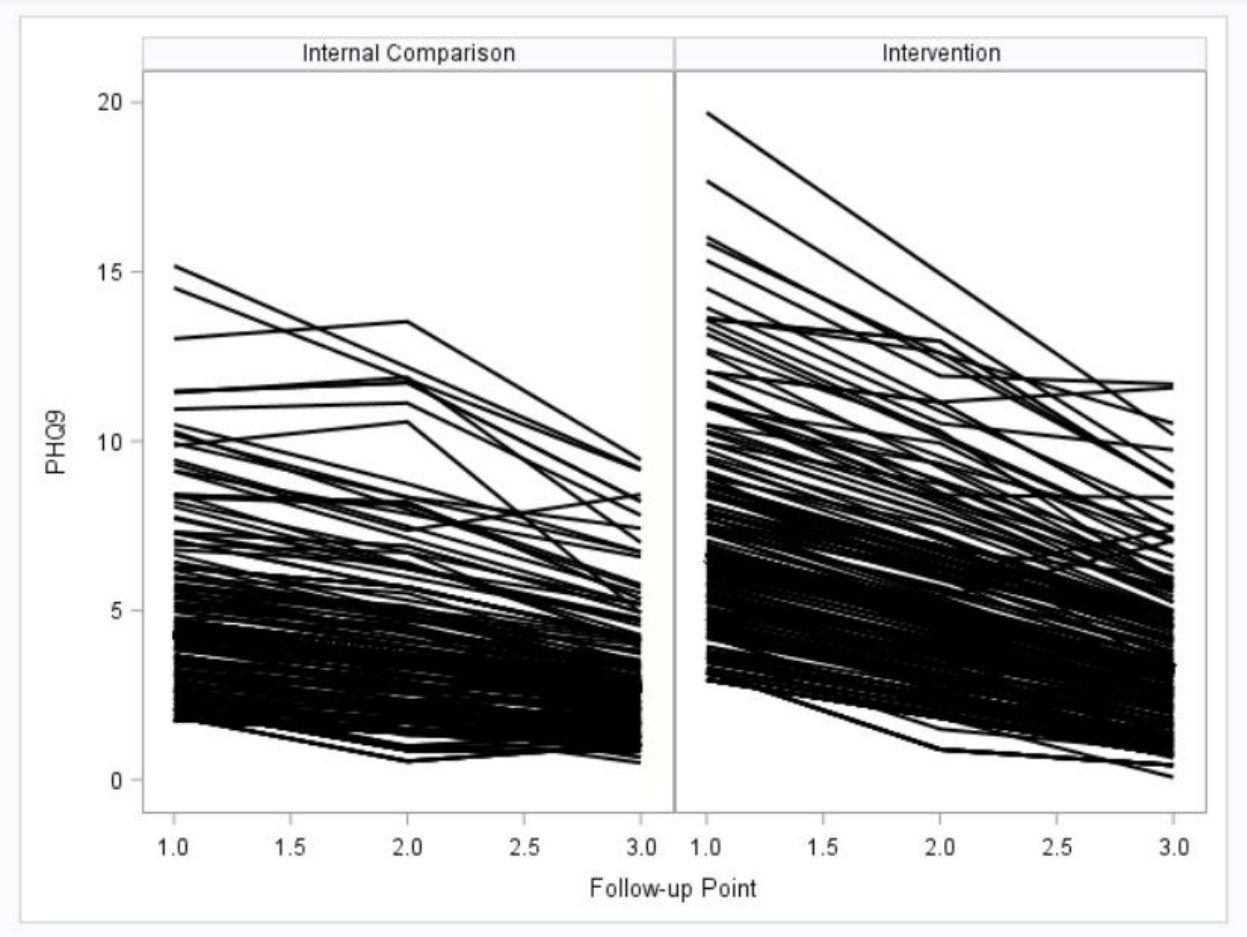
Table 24. Effect of IBH Intervention on Trajectory of PHQ-9 Score Across Twelve Month Study, Mercy Intervention Compared to Primary Comparison Group

Variable	PHQ-9 (n=332)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	**-1.76**	**0.54**	**0.001**
Time*Primary Comparison (ref)	--	--	--
Time	-1.75	0.38	<0.001
Intervention	2.28	0.52	<0.001
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To visualize the longitudinal effect of the intervention on PHQ-9 score, a two-panel spaghetti plot was produced using PROC SGPANEL. **Figure 6** displays the primary comparison group trajectory in the left panel and the intervention group trajectory in the right panel. The x-axis of the graph shows the study follow-up points with 1.0 representing baseline, 2.0 is the 6-month point, and 3.0 is the 12-month end-point. The trajectory figure visually displays the differences identified in the longitudinal statistical model, identifying a higher mean PHQ-9 score at baseline within the intervention compared to the primary comparison group and a steeper decrease in PHQ-9 score is seen in the trajectories from baseline to 12 months for those in the intervention than in the primary comparison group.

Figure 6. Individual Trajectories of PHQ-9 Score Across Twelve Month Study Period for IBH Intervention and Primary Internal Comparison Groups



To understand whether and how the different intervention components affected the relationship between the intervention and mean PHQ-9 score, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received. When including these four visit variables in the model for PHQ-9 score, the variable for number of behavioral health was selected based on our established selection criteria of p-value <0.15 in the initial model. The results of the model with this additional variable are presented in **Table 25**. Accounting for the number of behavioral health visits in the model, the effect on PHQ-9 score is statistically significant (p=0.01). On average, for participants in the intervention group, there is a 1.10 decrease in PHQ-9 score at 12 months holding all other variables in the selected model constant compared to participants in the primary comparison group.

Table 25. Effect of IBH Intervention on Twelve Month PHQ-9 Score Including Number of Visits, Mercy Intervention Group Compared to Primary Comparison Group

Variable	PHQ-9 (n=283)		
	Estimate (β)	Standard Error	p-value
Intervention	**-1.10**	**0.42**	**0.01**
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.02	0.02	0.31
Male[a]	-0.31	0.60	0.61
Female (ref)	--	--	--
English	1.23	0.68	0.07
Spanish (ref)	--	--	--
Current smoker	2.01	0.81	0.01
Former smoker	1.17	1.15	0.31
Never smoker (ref)	--	--	--
BL_PHQ-9	0.16	0.07	0.03
BL_General	-0.03	0.02	0.11
BL_GAD-7	0.15	0.07	0.04
BL_BMI	0.08	0.03	0.01
Number of behavioral health visits	0.23	0.06	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

There are no limitations specific to this measure to note.

Body Mass Index

Question 2. **Do patients who participate in the Sí Three intervention experience improvements in BMI after 12 months when compared to patients that do not participate in the intervention? This question is confirmatory.**

Overview of Analysis

To answer this confirmatory question about intervention impact on BMI, data on weight, height, and patient BMI were collected and evaluated. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed to analyze BMI data. The sample sizes for the presented analyses of BMI are as follows: bivariate analyses (n=287), primary linear regression analyses (n=281), and longitudinal analyses (n=338).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean BMI data in each study period for the overall sample as well as the intervention and primary comparison groups. The overall sample had a mean BMI of 32.9 kg/m^2 at baseline. Mean BMI was 33.0 kg/m^2 for those who returned at 6-month follow-up and

to 33.3 kg/m^2 for those who returned at 12-month follow-up. The intervention group began the study with a mean BMI of 33.2 kg/m^2 at baseline compared to the primary comparison group's mean BMI of 32.5 kg/m^2 at baseline. Aligning with the overall sample trend, for those who completed an assessment at follow-up, the intervention group mean BMI increased to 33.3 kg/m^2 at 6-month follow-up and to 34.0 kg/m^2 at 12 months. In the primary comparison group, mean BMI increased from baseline to 6 months to 32.7 kg/m^2 and remained stable at 12 months for those who completed a follow-up assessment. As previously noted, the two groups were considered statistically equivalent at baseline (**Table 9**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The increases observed from baseline to 12-month follow-up within both the intervention and primary comparison groups for BMI were not statistically significant.

Bivariate analyses were also performed between the intervention and primary comparison groups comparing mean BMI at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for BMI when comparing the intervention and primary comparison group at 12 months, the null hypothesis cannot be rejected. The mean BMIs at 12-months were not significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, BMI. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for BMI were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline BMI, the number of comorbidities at baseline, and baseline Spirituality Index score. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline PHQ-9 score was included for possible selection. This model was specified as follows:

$$Y_{(BMI)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5\ MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_BMI + \beta_{10} BL\_Comorbidities + \beta_{11} BL\_Spirituality + \beta_{12} BL\_PHQ\text{-}9P + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the completeness of the evaluated data.

The final model for BMI included those covariates with p-value of 0.15 or less: sex, primary language, baseline BMI, and the number of comorbidities at baseline. Age was maintained based on a priori selection. Age as a continuous variable was selected for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(BMI)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 BL\_BMI + \beta_6 BL\_comorbidities + \varepsilon$$

Findings

Estimates by covariate for the final model of BMI are presented in **Table 26**.

Mean BMI at 12 months did not differ significantly between the intervention and primary comparison group (p=0.87); the effect size (using Cohen's d) is 0.005. Below is the selected model with each covariate's effect estimate included:

$$Y_{(BMI)} = 1.46 + 0.03(\text{Intervention}) + -0.01(\text{Age}) + -0.55(\text{Male}) + 0.62(\text{English}) + 0.98(\text{BL\_BMI}) + -0.15(\text{BL\_Comorbidities}) + \varepsilon$$

Table 26. Effect of IBH Intervention on Twelve Month BMI, Mercy Intervention Group Compared to Primary Comparison Group

Variable	BMI (n=281)		
	Estimate (β)	Standard Error	p-value
Intervention	0.03	0.19	0.87
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.01	0.01	0.49
Male	-0.55	0.27	0.04
Female (ref)	--	--	--
English	0.62	0.32	0.05
Spanish (ref)	--	--	--
BL_BMI	0.98	0.02	<0.001
BL_Comorbidities	-0.15	0.08	0.06

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria

There were no statistically significant effect modifications for BMI (not shown). The models estimated included interaction terms between intervention group and baseline obesity and baseline Spirituality Index quintiles.

Additional Analyses

Longitudinal analyses were used to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. There was no significant time by intervention group interaction for BMI over the 12-month study (p=0.62), indicating that the BMI trajectories from baseline to 6 months, and then to 12 months did not differ statistically between the two study arms (**Table 27**). Adjusting for the covariates that were selected in the primary model—age, sex, language, and number of comorbidities at baseline—did not alter these results (not shown).

Table 27. Effect of IBH Intervention on Trajectory of BMI Across Twelve-Month Study, Mercy Intervention Group Compared to Primary Comparison Group

Variable	BMI (n=338)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	-0.09	0.18	0.62
Time*Primary Comparison (ref)	--	--	--
Time	0.20	0.12	0.11
Intervention	0.75	0.65	0.25
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To understand whether and how the different intervention components affected the relationship between the intervention and BMI, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received. When including these four visit variables in the model for BMI, the variable for number of primary care visits was selected based on the selection criteria of a p-value <0.15. The results of the model with these additional variables are presented in **Table 28**. Accounting for the number of behavioral health and primary care visits, the intervention effect on BMI does not meet criteria for statistical significance (p=0.74).

Table 28. Effect of IBH Intervention on Twelve Month BMI Including Number of Visits, Mercy Intervention Group Compared to Primary Comparison Group

Variable	BMI (n=281)		
	Estimate (β)	Standard Error	p-value
Intervention	-0.06	0.19	0.74
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.01	0.01	0.24
Male	-0.50	0.26	0.06
Female (ref)	--	--	--
English	0.60	0.31	0.06
Spanish (ref)	--	--	--
BL_BMI	0.98	0.02	<0.001
BL_Comorbidities	-0.17	0.08	0.03
Number of primary care visits	0.06	0.03	0.03

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

There are no limitations specific to this measure to note.

Functioning and Quality of Life

Question 3. **Do patients who participate in the Sí Three intervention experience improvements in quality of life, as measured by the Duke Health Profile, after 12 months when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on functioning and quality of life, data were collected using the Duke Health Profile, specifically the General Health score. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for the Duke Health Profile. Analyses were also conducted on the Duke Health Profile domains that comprise the General Health score: Physical Health, Mental Health, and Social Health scores. The sample sizes for the presented analyses of Duke General Health score are as follows: bivariate analyses (n=290), primary linear regression analyses (n=283), and longitudinal analyses (n=332).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean Duke General Health score data in each study period for the overall sample as well as the intervention and primary comparison groups. The overall sample had a mean Duke General Health score of 71.1 at baseline. This increased to 75.4 for participants who returned at 6-month follow-up and again to 79.3 for those who returned at 12-month follow-up. The intervention group had a lower mean Duke General Health score of 67.7 at baseline compared to the primary comparison group mean score of 74.5 at baseline, indicating better general health status. Across the study, for participants who completed a follow-up assessment, the intervention group mean Duke General Health score increased at both 6 and 12-month follow-up to 72.9 and 78.5 respectively. The primary comparison group also followed a similar trend in general health with mean Duke General Health score for those who completed a follow-up increasing to 78.6 at 6 months and 80.0 at 12 months. As previously noted, the differences in Duke General Health score between the intervention and primary comparison groups at baseline were statistically significant (**Table 9**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The improvements observed in Duke General Health score from baseline to 12-month follow-up within both the intervention and primary comparison groups were statistically significant.

Bivariate analyses were also performed between the intervention and primary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for Duke General Health score when comparing the intervention and primary comparison group at 12 months, the null hypothesis cannot be rejected. The mean Duke General Health scores at 12-months were not significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, Duke General Health score. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for Duke General Health score were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline Duke General Health score, baseline PHQ-9 score, baseline GAD-7 score, the number of comorbidities at baseline, and baseline Spirituality Index score. The inclusion of baseline Duke General Health score controlled for the statistical imbalance between intervention and primary comparison groups at baseline. Additionally, to further understand whether and how physical and behavioral health are associated in this study population, baseline BMI was included for possible selection. The model was specified as follows:

$$Y_{(DUKE\ General)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_General + \beta_{10} BL\_PHQ\text{-}9 + \beta_{11} BL\_GAD\text{-}7 + \beta_{12} BL\_Comorbidities + \beta_{13} BL\_Spirituality + \beta_{14} BL\_BMI + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, a multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model for Duke General Health score included those covariates with p-value of 0.15 or less, which were baseline Duke General Health score, baseline PHQ-9 score, and baseline BMI. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results. The final model specification was:

$$Y_{(DUKE\ General)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 BL\_General + \beta_5 BL\_PHQ\text{-}9 + \beta_6 BL\_BMI + \varepsilon$$

Because the baseline depression measure was selected this model of quality of life as proposed, an additional test to quantify any multicollinearity between the Duke General Health score with PHQ-9 score. The variance inflation factor (VIF) of PHQ-9 score in the Duke General Health score model was 2.1, below the commonly accepted cutoff of 5 indicating minimal influence on the variance from the correlation of these variables (Belsley et al., 1980; O'Brien, 2007; Lasser, et al. 2017).

Findings

Estimates by covariate for the final model of Duke General Health score are presented in **Table 29**.

On average, there is a 4.01 increase in Duke General Health score at 12 months for participants in the intervention group compared to those in the primary comparison group, holding all other variables in the model constant. This result is statistically significant with a p-value of 0.02; the effect size (using Cohen's d) is 0.24. Below is the selected model with each covariate's effect estimate included:

$$Y_{(Duke\ General)} = 73.7 + 4.01(Intervention) + -0.08(Age) + -2.24(Male) + 0.40(BL\_General) + -0.75(BL\_PHQ\text{-}9) + -0.51(BL\_BMI) + \varepsilon$$

Table 29. Effect of IBH Intervention on Twelve Month Duke General Health Score, Mercy Intervention Group Compared to Primary Comparison Group

Variable	Duke General Health (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	**4.01**	**1.64**	**0.02**
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.08	0.08	0.33
Male[a]	-2.24	2.31	0.34
Female (ref)	--	--	--
BL_General	0.40	0.07	<0.001
BL_PHQ-9	-0.75	0.22	0.001
BL_BMI	-0.51	0.12	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).
[a] Included in the model a priori despite not having met the stepwise inclusion criteria

As previously noted, models were created separately to examine the component domains of the composite Duke General Health score: Physical Health, Mental Health, and Social Health. These analyses aimed to understand the statistically significant improvement in quality of life in the intervention group. Each of the three component scores began with the same possible model for selection as the General Health score, substituting the corresponding baseline Duke Health Profile domain score for the baseline General Health score. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes.

For the Duke Physical Health score, the covariates selected using the backward selection approach were: alcohol consumption, baseline Physical Health score, baseline GAD7, and baseline BMI. Age and sex were maintained based on a priori selection. Estimates by covariate for the final model of Duke Physical Health score are presented in **Table 30**.

On average, there is a 6.69-point increase in Duke Physical Health score at 12 months for participants in the intervention group compared to those in the primary comparison group, holding all other variables in the model constant. This result is statistically significant with a p-value of 0.004.

$$Y_{(DUKE\ Physical)} = 87.54 + 6.69(Intervention) + -0.23(Age) + 2.84(Male) + 5.13(No\ Alcohol\ Use) + 0.40(BL\_Physical) + -0.98(BL\_GAD7) + -0.95(BL\_BMI) + \varepsilon$$

Table 30. Effect of IBH Intervention on Twelve Month Duke Physical Health Score, Mercy Intervention Group Compared to Primary Comparison Group

Variable	Duke Physical Health (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	**6.69**	**2.29**	**0.004**
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.23	0.12	0.05
Male[a]	2.84	3.35	0.40
Female (ref)	--	--	--
No Alcohol Use	5.13	3.01	0.09
Alcohol Use (ref)	--	--	--
BL_Physical	0.40	0.06	<0.001
BL_GAD-7	-0.98	0.27	<0.001
BL_BMI	-0.95	0.18	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).
[a] Included in the model a priori despite not having met the stepwise inclusion criteria.

For the Duke Mental Health score, the covariates selected using the backward selection approach were: primary language, baseline Mental Health score, baseline PHQ-9 score, baseline Spirituality Index, and baseline BMI. Age and sex were maintained based on a priori selection. Estimates by covariate for the final model of Duke Mental Health score are presented in **Table 31**.

No statistical difference in Duke Mental Health score at 12 months was identified for participants in the intervention group compared to those in the primary comparison group, holding all other variables in the model constant (p=0.18).

$$Y_{(DUKE\ Mental)} = 81.43 + 2.86(Intervention) + -0.10(Age) + -0.74(Male) + -7.34(English) + 0.16(BL\_Mental) + -1.07(BL\_PHQ\text{-}9) + 0.26(BL\_SpiritualityIndex) + -0.37(BL\_BMI) + \varepsilon$$

Table 31. Effect of IBH Intervention on Twelve Month Duke Mental Health Score, Mercy Intervention Group Compared to Primary Comparison Group

Variable	Duke Mental Health (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	2.86	2.12	0.18
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.10	0.11	0.35
Male[a]	-0.74	3.00	0.81
Female (ref)	--	--	--
English	-7.34	3.50	0.04
Spanish (ref)	--	--	--
BL_Mental	0.16	0.07	0.03
BL_PHQ-9	-1.07	0.30	<0.001
BL_SpiritualityIndex	0.26	0.11	0.02
BL_BMI	-0.37	0.16	0.02

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

For the Duke Social Health score, the covariates selected using the backward selection approach were: baseline Social Health score, baseline PHQ-9 score, and baseline BMI. Age and sex were maintained based on a priori selection. Estimates by covariate for the final model of Duke Social Health score are presented in **Table 32**.

No difference in Duke Social Health score was identified for participants in the intervention group compared to those in the primary comparison group, holding all other variables in the model constant (p=0.15).

$$Y_{(DUKE\ Social)} = 63.37 + 2.87(Intervention) + 0.09(Age) + -6.59(Male) + 0.32(BL\_Social) + -0.83(BL\_PHQ\text{-}9) + -0.22(BL\_BMI) + \varepsilon$$

Table 32. Effect of IBH Intervention on Twelve Month Duke Social Health Score, Full Mercy Sample (Primary Comparison Group)

Variable	Duke Social Health (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	2.87	1.97	0.15
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.09	0.10	0.35
Male[a]	-6.59	2.78	0.02
Female (ref)	--	--	--
BL_Social	0.32	0.06	<0.001
BL_PHQ-9	-0.83	0.21	<0.001
BL_BMI	-0.22	0.15	0.13

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistical significance (p-value < 0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for Duke General Health score (not shown). The model considered included an interaction term of study group and baseline Spirituality Index score quintile groupings (five groups of equal size based on the distribution of the Spirituality Index score).

Additional Analyses

Longitudinal analyses were used to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. There was a significant time by intervention group interaction for Duke General Health score over the 12-month study (p=0.003), indicating that the Duke General Health trajectories from baseline to 6 months, and then to 12 months were different between the two study arms (see **Table 33**). For each 6 months in the study, the intervention group experienced a relative increase of 5.35 points in Duke General Health score compared to the primary comparison group. Adjusting for the covariates that were selected in the primary model—age, sex, marital status, baseline PHQ-9 score, and number of comorbidities at baseline—did not alter these results.

Table 33. Effect of IBH Intervention on Trajectory of Duke General Health Score Across Twelve-Month Study, Full Mercy Sample (Primary Comparison Group)

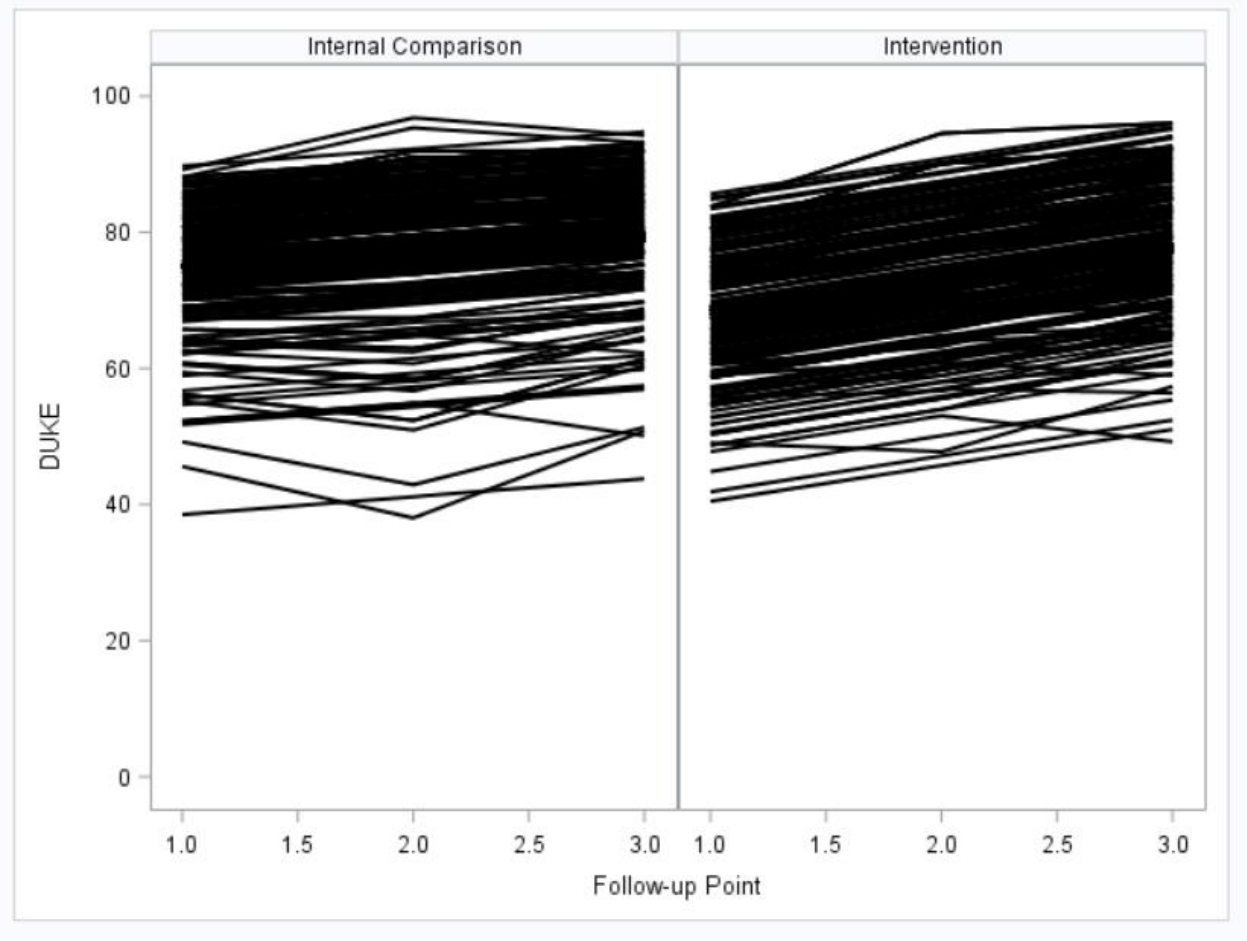
Variable	Duke General Health (n=332)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	**5.35**	**1.78**	**0.003**
Time*Primary Comparison (ref)	--	--	--
Time	5.10	1.25	<0.001
Intervention	-6.85	1.65	<0.001
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To visualize the longitudinal effect of the intervention on Duke General Health score, a two-panel spaghetti plot using PROC SGPANEL was used. **Figure 7** displays the primary comparison group trajectory

in the left panel and the intervention group trajectory in the right panel. The x-axis of the graph shows the study follow-up points with 1.0 representing baseline, 2.0 is the 6-month point, and 3.0 is the 12-month end-point. The trajectory figure visually displays the differences identified in the longitudinal statistical model, identifying lower mean Duke General Health scores at baseline within the intervention compared to the primary comparison group. The increase seen in the trajectories from baseline to 12 months is more salient for those in the intervention than in the primary comparison group indicating a greater increase over time for the intervention participants. There also appears to be more variability across time in the primary comparison with some participants showing a decrease in Duke General Health score at 6 months and then an increase from 6 to 12 months. The trend of trajectories in the intervention group appears to increase between the three time points.

Figure 7. Individual Trajectories of Duke General Health Score Across Twelve Month Study Period for IBH Intervention and Primary Internal Comparison Groups



To understand whether and how the different intervention components had effects on the relationship between the intervention and Duke General Health score, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received. When including these four visit variables in the model

for Duke General Health score, the number of behavioral health visits was selected based on an established selection criterion of a p-value <0.15. The results of the model with this additional variable are presented in **Table 34**. Accounting for the number of behavioral health visits, the average Duke General Health score at 12-months is significantly increased by 5.00 points for participants in the intervention group compared to participants in the primary comparison group, holding all other variables in the selected model constant (p=0.002).

Table 34. Effect of IBH Intervention on Twelve Month Duke General Health Score Including Number of Visits, Full Mercy Sample (Primary Comparison Group)

Variable Selected	Duke General Health (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	**5.00**	**1.63**	**0.002**
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.08	0.08	0.29
Male[a]	-1.87	2.36	0.43
Female (ref)	--	--	--
No Alcohol Use	3.40	2.11	0.11
Alcohol Use (ref)	--	--	--
BL_General	0.37	0.07	<0.001
BL_PHQ-9	-0.62	0.22	0.01
BL_BMI	-0.52	0.12	<0.001
Number of behavioral health visits	-0.79	0.23	0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

According to Mercy clinic staff, participants in both the intervention and primary comparison group experienced difficulty completing the Duke assessments at each time point due to Spanish translations of the instrument which used language unfamiliar to Mercy patients. This may have resulted in an instrumentation bias; however, this bias would presumably be similar among these groups as they are from the same clinic population. *(Note: The Duke Spanish language surveys used in the Sí Texas study had been validated in the literature and HRiA conducted focus groups in the study area to ensure that the survey language was regionally appropriate).*

Anxiety Symptoms

Question 4. **Do patients who participate in the Sí Three intervention experience improvements in anxiety symptoms, as measured by GAD-7, after 12 months compared to patients who do not participate? In addition, do these improvements differ by type of behavioral-health service received (medical/behavioral or faith-based services)? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on anxiety, data were analyzed from the GAD-7 assessment tool. The additional exploratory question aimed at understanding whether there was

effect modification by the type of behavioral health services received was not performed as participants exclusively selected the faith-based services, therefore there was no second group for stratified analyses. This is a deviation from the SEP; however, the decision was based on the data collected representing Mercy's clinical practice. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for GAD-7 score. The sample sizes for the presented analyses of GAD-7 score are as follows: bivariate analyses (n=290), primary linear regression analyses (n=283), and longitudinal analyses (n=332).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean GAD-7 score data in each study period for the overall sample as well as the intervention and primary comparison groups. The overall sample had a mean GAD-7 score of 5.2 at baseline. This decreased to 3.8 for participants who returned at 6-month follow-up and again to 2.5 for those who returned at 12-month follow-up. The intervention group began the study with a higher mean GAD-7 score of 6.2 at baseline while the primary comparison group had a lower mean GAD-7 score of 4.3 at baseline, indicating less anxiety symptoms. Aligning with the overall sample trend, for participants who completed a follow-up assessment, the intervention group mean GAD-7 score decreased at both 6 and 12-month follow-up to 4.6 and 2.6 respectively. The primary comparison group also followed this trend with the mean GAD-7 score for those who completed a follow-up decreasing overtime to 2.7 at 6 months and 2.4 at 12 months. As previously noted in **Table 9**, the intervention and primary comparison groups were not statistically equivalent on GAD-7 score at baseline.

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The decreases from baseline to 12-month follow-up in GAD-7 score within both the intervention and primary comparison groups were statistically significant.

Bivariate analyses were also performed between the intervention and primary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for GAD-7 score when comparing the intervention and primary comparison group at 12 months, the null hypothesis cannot be rejected. The mean GAD-7 scores were not significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, GAD-7 score. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for GAD-7 score were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline GAD-7 score, baseline PHQ-9 score, baseline Duke General Health score, the number of comorbidities at baseline, and baseline Spirituality Index score. The inclusion of baseline GAD-7 score controlled for the statistical imbalance between intervention and primary comparison groups at baseline. Additionally, to further understand whether and how physical

and mental health are associated in this study population, baseline BMI was included for possible selection. This model was specified as follows:

$$Y_{(GAD7)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_GAD\text{-}7 + \beta_{10} BL\_PHQ\text{-}9 + \beta_{11} BL\_General + \beta_{12} BL\_Comorbidities + \beta_{13} BL\_Spirituality + \beta_{14} BL\_BMI + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, a multiple imputation approach was considered but not performed due to the completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model for GAD-7 score included those covariates with p-value of 0.15 or less: primary language, baseline GAD-7 score, baseline Duke General Health score, and baseline BMI. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(GAD\text{-}7)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 BL\_GAD\text{-}7 + \beta_6 BL\_General + \beta_7 BL\_BMI + \varepsilon$$

Because the baseline quality of life measure was selected for inclusion in the final model of anxiety, and quality of life and anxiety are known to be related, an additional test was conducted to quantify any multicollinearity between the Duke General Health score with GAD-7 score. The variance inflation factor (VIF) of Duke General Health score in the GAD-7 score model was 1.9, below the commonly accepted cutoff of 5 indicating minimal influence on the variance from the correlation of these variables (Belsley et al., 1980; O'Brien, 2007; Lasser, et al. 2017).

Findings

Estimates by covariate for the final model of GAD-7 score are presented in **Table 35**.

On average, there was a 0.79-point decrease in GAD-7 score at 12 months for participants in the intervention group compared to those in the primary comparison group, holding all other variables in the model constant. This result is statistically significant with a p-value of 0.03; the effect size (using Cohen's d) is 0.22. Below is the selected model with each covariate's effect estimate included:

$$Y_{(GAD7} = 1.13 + -0.79(Intervention) + 0.001(Age) + -0.29(Male) + 1.36(English) + 0.30(BL\_GAD\text{-}7) + -0.04(BL\_General) + 0.08(BL\_BMI) + \varepsilon$$

Table 35. Effect of IBH Intervention on Twelve Month GAD-7 Score, Full Mercy Sample (Primary Comparison Group)

Variable	GAD-7 (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	**-0.79**	**0.37**	**0.03**
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.001	0.02	0.96
Male[a]	-0.29	0.52	0.58
Female (ref)	--	--	--
English	1.36	0.61	0.03
Spanish (ref)	--	--	--
BL_GAD-7	0.30	0.05	<0.001
BL_General	-0.04	0.01	0.01
BL_BMI	0.08	0.03	0.005

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for GAD-7 score (not shown). The models estimated included interaction terms between intervention group and baseline anxiety and baseline Spirituality Index quintiles.

Additional Analyses

Longitudinal analyses were used to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate a linear mixed model, the PROC MIXED procedure in SAS was used. There was a significant time by intervention group interaction on trajectory of GAD-7 score over the 12-month study (p=0.002), indicating that the GAD-7 trajectories from baseline to 6 months, and then to 12 months were different between the two study arms (**Table 36**). For each 6-month period in the study, the intervention group experienced a relative decrease of 1.58 points in GAD-7 score compared to the primary comparison group. Adjusting for the covariates that were selected in the primary model—age, sex, language, baseline Duke General Health score, and baseline BMI—did not alter these results.

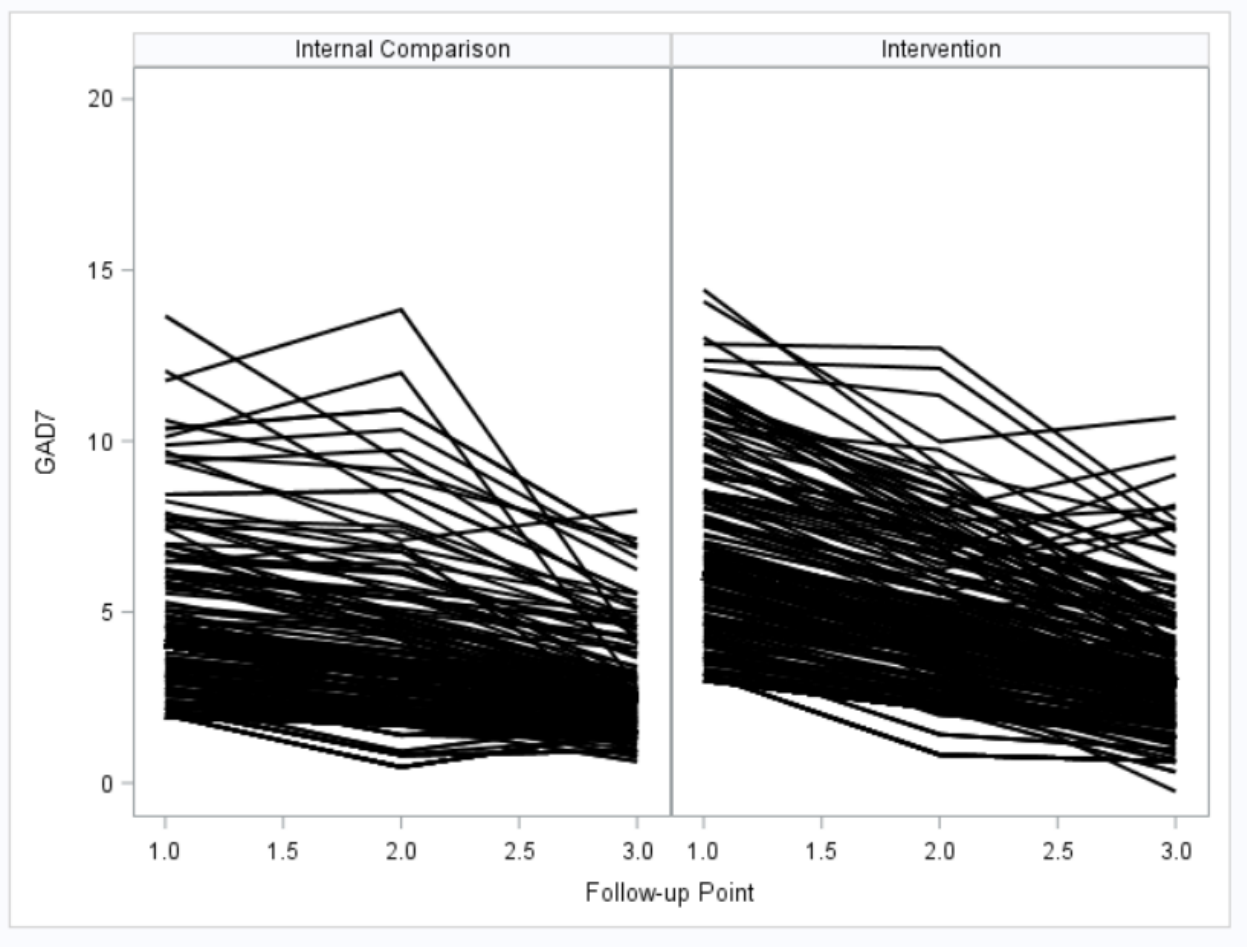
Table 36. Effect of IBH Intervention on Trajectory of GAD-7 Score Across Twelve Month Study Follow Up, Full Mercy Sample (Primary Comparison Group)

Variable	GAD-7 (n=332)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	**-1.58**	**0.51**	**0.002**
Time*Primary Comparison (ref)	--	--	--
Time	-1.78	0.36	<0.001
Intervention	1.98	0.51	<0.001
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

The longitudinal effect of the intervention on GAD-7 score was visualized by a two-panel spaghetti plot using PROC SGPANEL. **Figure 8** displays the primary comparison group trajectory in the left panel and the intervention group trajectory appears in the right panel. The x-axis of the graph shows the study follow-up points with 1.0 representing baseline, 2.0 is the 6-month point, and 3.0 is the 12-month end-point. The trajectory figure visually displays the differences identified in the longitudinal statistical model, identifying a higher GAD-7 score in the intervention group than in the primary comparison group. The decrease seen in the trajectories from baseline to 12 months is more salient for those in the intervention than in the primary comparison group suggesting a greater decrease over time for the intervention participants. There also appears to be more variability across time in the primary comparison with some participants showing an increase in GAD-7 score at 6 months followed by a decrease to 12 months. In comparison, the trend of trajectories in the intervention appears to decrease at both time points for nearly all participants.

Figure 8. Individual Trajectories of GAD-7 Score Across Twelve Month Study Period for IBH Intervention and Primary Internal Comparison Groups



To understand whether and how the different intervention components affected the relationship between the intervention and mean GAD-7 score, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received. When including these four visit variables in the model for GAD-7 score, the number of behavioral health visits was selected based on the selection criteria of p-value <0.15. The results of the model with this additional variable are presented in **Table 37**. Accounting for the number of behavioral health visits, the effect on GAD-7 score is statistically significant (p=0.003). On average, for participants in the intervention group, there is a 1.10 decrease in GAD-7 score at 12 months holding all other variables in the selected model constant compared to participants in the primary comparison group.

Table 37. Effect of IBH Intervention on Twelve Month GAD-7 Score Including Number of Visits, Full Mercy Sample (Primary Comparison Group)

Variable	GAD-7 (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	**-1.10**	**0.36**	**0.003**
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.003	0.02	0.85
Male[a]	-0.07	0.51	0.89
Female (ref)	--	--	--
English	1.30	0.58	0.03
Spanish (ref)	--	--	--
BL_GAD-7	0.28	0.05	<0.001
BL_General	-0.02	0.01	0.12
BL_BMI	0.08	0.03	0.003
Number of behavioral health visits	0.23	0.05	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

No limitations specific to this measure to note.

CAGE-AID

Question 5. **Do patients who participate in the Sí Three intervention experience improvements in addiction symptoms, as measured by CAGE-AID, after 12 months compared to patients who do not participate? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on addictive behavior, data were collected using the CAGE-AID tool. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for the CAGE-AID score. The sample sizes for the presented analyses of CAGE-AID score are as follows: bivariate analyses (n=290), primary linear regression analyses (n=283), and longitudinal analyses (n=332).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean CAGE-AID score data in each study period for the overall sample as well as the intervention and primary comparison groups. The overall sample had a mean CAGE-AID score of 0.2 at baseline. This decreased to 0.1 for participants who returned at 6-month follow-up and remained the same for those who returned at 12-month follow-up. Mean scores for the intervention and primary comparison groups were equivalent to the study means across the three data collection points. As previously noted, the intervention and primary comparison groups were statistically equivalent on CAGE-AID score at baseline (**Table 9**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The observed changes from baseline to 12-month follow-up within both the intervention and primary comparison groups for CAGE-AID score were not statistically significant.

Bivariate analyses also were performed between the intervention and primary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for CAGE-AID score when comparing the intervention and primary comparison group at 12 months, the null hypothesis cannot be rejected. The mean CAGE-AID scores were not significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, CAGE-AID score. Age and sex were selected a priori for inclusion due to known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for CAGE-AID score were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline CAGE-AID, baseline PHQ-9 score, baseline Duke General Health score, baseline GAD-7 score, the number of comorbidities at baseline, and baseline Spirituality Index score. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline BMI was included for possible selection.

$$Y_{(CAGE\text{-}AID)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5\ MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_CAGE\text{-}AID + \beta_{10} BL\_PHQ\text{-}9 + \beta_{11} BL\_General + \beta_{13} BL\_GAD\text{-}7 + \beta_{13} BL\_Comorbidities + \beta_{14} BL\_Spirituality + \beta_{15} BL\_BMI + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model for CAGE-AID score included those covariates with p-value of 0.15 or less; alcohol consumption, baseline CAGE-AID score, and number of comorbidities at baseline. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(CAGEAID)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Alcohol + \beta_5 BL\_CAGE\text{-}AID + \beta_6 BL\_comorbidites + \varepsilon$$

Findings

Estimates by covariate for the final model of CAGE-AID score are presented in **Table 38**.

Mean CAGE-AID score at 12 months did not differ significantly between the intervention and primary comparison group (p=0.33). Below is the selected model with each covariate's effect estimate included:

$Y_{(CAGEAID)}$ = -0.15 + -0.06(Intervention) + -0.001(Age) + 0.30(Male) + 0.12(No Alcohol Use) + 0.36(BL_CAGE-AID) + 0.05(BL_Comorbidities) + ε

Table 38. Effect of IBH Intervention on Twelve Month CAGE-AID Score, Full Mercy Sample (Primary Comparison Group)

Variable	CAGE-AID (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	-0.06	0.06	0.33
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.001	0.003	0.68
Male[a]	0.30	0.09	0.001
Female (ref)	--	--	--
No Alcohol	0.12	0.08	0.14
Alcohol (ref)	--	--	--
BL_CAGE-AID	0.36	0.05	<0.001
BL_Comorbidities	0.05	0.02	0.05

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for CAGE-AID score (not shown). The models included interaction terms between intervention group and baseline severe addiction and baseline Spirituality Index quintiles.

Additional Analyses

Longitudinal analyses were used to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate a linear mixed model, the PROC MIXED procedure in SAS was used. There was no significant time by intervention group interaction on trajectory of CAGE-AID score over the 12-month study (p=0.56), indicating that the CAGE-AID trajectories from baseline to 6 months, and then to 12 months were not statistically significantly different between the two study arms (see **Table 39**).

For each 6-month period in the study, the intervention group experienced a relative decrease of 0.04 points in CAGE-AID score compared to the primary comparison group. Adjusting for the covariates that were selected in the primary model—age, sex, alcohol use, and number of comorbidities at baseline—did not alter these results.

Table 39. Effect of IBH Intervention on Trajectory of CAGE-AID Score Across Twelve Month Study Follow Up, Full Mercy Sample (Primary Comparison Group)

Variable	CAGE-AID (n=332)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	-0.04	0.07	0.56
Time*Primary Comparison (ref)	--	--	--
Time	-0.05	0.05	0.34
Intervention	0.01	0.06	0.83
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To understand whether and how the different intervention components had effects on the relationship between the intervention and CAGE-AID score, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received. When including these four visit variables in the model for CAGE-AID score, the number of primary care visits was selected based on the selection criteria of a p-value <0.15. The results of the model with this additional variable are presented in **Table 40**. When the model is adjusted for the number of primary care visits the intervention effect on CAGE-AID score is not statistically significant (p=0.67).

Table 40. Effect of IBH Intervention on Twelve Month CAGE-AID Score Including Number of Visits, Full Mercy Sample (Primary Comparison Group)

Variable	CAGE-AID (n=283)		
	Estimate (β)	Standard Error	P-value
Intervention	-0.03	0.06	0. 67
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.0002	0.003	0.95
Male[a]	0.28	0.09	0.002
Female (ref)	--	--	--
No Alcohol	0.12	0.08	0.14
Alcohol (ref)	--	--	--
BL_CAGE-AID	0.35	0.05	<0.001
BL_Comorbidities	0.05	0.02	0.02
Number of primary care visits	-0.02	0.01	0.03

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

This measure is limited by the combination of the categorical scale and the high frequency of responses of 0 for all four items of the CAGE-AID at both baseline and 12 months.

Blood Pressure

Question 6. **Do patients who participate in the Sí Three intervention experience improvements in blood pressure, when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on blood pressure, data were collected for both systolic (SBP) and diastolic (DBP) blood pressure. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for systolic or diastolic blood pressure. The sample sizes for the presented analyses of systolic and diastolic blood pressure are as follows: bivariate analyses (n=291), primary linear regression analyses (n=283), and longitudinal analyses (n=338).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean systolic and diastolic blood pressure data in each study period for the overall sample as well as the intervention and primary comparison groups. The overall sample had a mean blood pressure of 124.6/74.2 mmHg at baseline. For those who returned for a follow-up assessment, the mean decreased to 121.9/72.9 mmHg at 6-month follow-up and increased again at 12-month follow-up (124.0/74.2 mmHg). The intervention group began the study with a higher mean blood pressure, 125.3/74.9 mmHg at baseline while the primary comparison group had a lower mean blood pressure of 123.9/73.5 mmHg at baseline. As with the overall sample trend, each group's mean blood pressure decreased at 6 months and increased again at 12-month follow-up. In the intervention group, for those who returned for a follow-up assessment, the mean blood pressure at 6 months was 121.7/73.2 mmHg and 124.6/74.4 mmHg at 12 months; in the primary comparison group, the 6-month mean blood pressure was 122.2/72.6 mmHg and 124.0/74.0 mmHg at the 12-month follow-up. As previously noted, the two groups were equivalent on both blood pressure measures at baseline (**Table 9**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The decrease from baseline to 12-month follow-up for systolic blood pressure in the intervention group was statistically significant, but no other differences in systolic or diastolic blood pressure were found to be statistically significant. The changes in diastolic blood pressure were not statistically significant in either the intervention or the primary comparison.

Bivariate analyses also were performed between the intervention and primary comparison groups comparing mean blood pressure at 12-month follow-up (

Table 67). Based on p-values greater than 0.05 for both systolic and diastolic blood pressure, when comparing the intervention and primary comparison group at 12 months and without controlling for any

additional covariates, the null hypotheses cannot be rejected. The mean blood pressure measures at 12-months were not significantly different between the two study groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcomes of systolic and diastolic blood pressure. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for both systolic and diastolic blood pressure were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline systolic blood pressure, baseline diastolic blood pressure, the number of comorbidities at baseline, and baseline Spirituality Index score. Additionally, to further understand whether and how physical and behavioral health are associated in this study population, baseline PHQ-9 score was included for possible selection in the initial full models. These models were specified as follows:

$$Y_{(SBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_SBP + \beta_{10} BL\_DBP + \beta_{11} BL\_Comorbidities + \beta_{12} BL\_Spirituality + \beta_{13} BL\_PHQ\text{-}9 + \varepsilon$$

$$Y_{(DBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_DBP + \beta_{10} BL\_SBP + \beta_{11} BL\_Comorbidities + \beta_{12} BL\_Spirituality + \beta_{13} BL\_PHQ\text{-}9 + \varepsilon$$

Two variations of each model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, a multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of systolic blood pressure included those covariates with p-value of 0.15 or less: sex, marital status, baseline systolic blood pressure, and baseline diastolic blood pressure. Age was maintained based on a priori selection. Continuous age was forced in as a predictor due to the known biological influence of age on health outcomes. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification for was:

$$Y_{(SBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Married + \beta_5 BL\_SBP + \beta_6 BL\_DBP + \varepsilon$$

Because diastolic blood pressure was selected for inclusion into the final model of systolic blood pressure, and systolic and diastolic blood pressure are known to be related, an additional test was conducted to quantify any multicollinearity between systolic and diastolic blood pressure values. The variance inflation factor (VIF) of diastolic blood pressure in the systolic blood pressure model was 1.8 which is below the accepted cutoff of 5 representing a minimal influence on the variance from the correlation of these two variables (Belsley et al., 1980; O'Brien, 2007; Lasser, et al. 2017).

The final model for diastolic blood pressure included those covariates with p-value of 0.15 or less: baseline diastolic blood pressure. Age and sex were maintained based on a priori selection. Age was

modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(DBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 BL\_DBP + \varepsilon$$

Findings

Estimates for the final models of systolic and diastolic blood pressure are presented in **Table 41**.

Mean systolic blood pressure at 12 months did not differ significantly between the intervention and primary comparison group (p=0.63). Below is the selected model with each covariate's effect estimate included:

$$Y_{(SBP)} = 62.29 + -0.71(Intervention) + 0.28(Age) + 4.59(Male) + -2.68(Married) + 0.28(BL\_SBP) + 0.20(BL\_DBP) + \varepsilon$$

Mean diastolic blood pressure at 12 months did not differ significantly between the intervention and primary comparison group (p=0.56). Below is the selected model with each covariate's effect estimate included:

$$Y_{(DBP)} = 44.13 + -0.60(Intervention) + 0.01(Age) + 0.13(Sex) + 0.40(BL\_DBP) + \varepsilon$$

Table 41. Effect of IBH Intervention on Twelve Month Systolic and Diastolic Blood Pressure Value, Full Mercy Sample (Primary Comparison Group)

Variable	Systolic Blood Pressure (n=283)		
	Estimate (β)	Standard Error	p-value
Intervention	-0.71	1.48	0.63
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.28	0.08	<0.001
Male	4.59	2.16	0.03
Female (ref)	--	--	--
Married	-2.68	1.48	0.07
Unmarried (ref)	--	--	--
BL_SBP	0.28	0.06	<0.001
BL_DBP	0.20	0.10	0.05
Variable Selected	Diastolic Blood Pressure (n=283)		
	Estimate (β)	Standard Error	p-value
Intervention	-0.60	1.02	0.56
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.01	0.05	0.91
Male[a]	0.13	1.49	0.93
Female (ref)	--	--	--
BL_DBP	0.40	0.05	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for systolic or diastolic blood pressure (not shown). The models estimated included interaction terms between intervention group and baseline hypertension and baseline Spirituality Index quintiles.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. There was no significant time by intervention group interaction on trajectory of systolic blood pressure over the 12-month study (p=0.43), indicating that the trajectories from baseline to 6 months, and then to 12 months were not different between the two study arms for systolic blood pressure (**Table 42**). Adjusting for the covariates that were selected in the primary model—age, sex, marital status, and baseline diastolic blood pressure—did not alter these results.

For diastolic blood pressure there was no significant time by intervention group interaction, indicating that the trajectories from baseline to 6 months, and then to 12 months were not different between the two study arms for diastolic blood pressure (p=0.17; **Table 42**). Adjusting for the covariates that were selected in the primary model—age and sex—did not alter these results.

Table 42. Effect of IBH Intervention on Trajectory of Systolic and Diastolic Blood Pressure Value Across Twelve Month Study, Full Mercy Sample (Primary Comparison Group)

Variable	Systolic Blood Pressure (n=338)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	-1.41	1.79	0.43
Time*Primary Comparison (ref)	--	--	--
Time	-0.83	1.26	0.51
Intervention	0.99	1.63	0.54
Primary Comparison (ref)	--	--	--
Variable	Diastolic Blood Pressure (n=338)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	-1.55	1.14	0.17
Time*Primary Comparison (ref)	--	--	--
Time	0.38	0.80	0.63
Intervention	1.33	0.92	0.15
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To understand whether and how the different intervention components affected the relationship between the intervention and blood pressure, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and

included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received. When including these four visit variables in the model for blood pressure, the number of primary care visits was selected based on the selection criteria of a p-value <0.15 in the model for systolic blood pressure. In the model for diastolic blood pressure, the number of primary care and exercise coaching visits were all selected. The results of the models with the additional variable are presented in **Table 43**.

Mean systolic blood pressure at 12-months did not differ significantly by intervention group after accounting for the primary care visits and holding all other variables constant (p=0.98). Similarly, mean diastolic blood pressure at 12-months did not differ significantly by intervention group after accounting for primary care, behavioral health, and exercise coaching visits, and holding all other variables in the selected model constant(p=0.50).

Table 43. Effect of IBH Intervention on Twelve Month Systolic and Diastolic Blood Pressure Value Including Number of Visits, Full Mercy Sample (Primary Comparison Group)

Variable	Systolic Blood Pressure (n=283)		
	Estimate (β)	Standard Error	p-value
Intervention	0.04	1.52	0.98
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.31	0.08	<0.001
Male	4.09	2.16	0.06
Female (ref)	--	--	--
Married	-2.79	1.48	0.06
Unmarried (ref)	--	--	--
BL_SBP	0.28	0.06	<0.001
BL_DBP	0.22	0.10	0.03
Number of primary care visits	-0.47	0.24	0.05
Variable	Diastolic Blood Pressure (n=283)		
	Estimate (β)	Standard Error	p-value
Intervention	-0.73	1.08	0.50
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.03	0.05	0.60
Male[a]	0.02	1.49	0.99
Female (ref)	--	--	--
BL_DBP	0.43	0.05	<0.001
Number of primary care visits	-0.33	0.16	0.04
Number of exercise coaching visits	0.08	0.04	0.03

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

There are no limitations specific to these measures to note.

HbA1c Level

Question 7. **Do patients with a history or diagnosis of diabetes who participate in the Sí Three intervention experience improvements in HbA1c measures after 12 months when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question of intervention impact on diabetes management among diabetic patients, data were collected on patient HbA1c levels. As previously stated, it is Mercy's clinical practice to recommend HbA1C test and subsequently collect among patients who are: (1) known/self-reported to be diabetic, (2) have an elevated blood glucose at time of clinic visit or are suspected to be diabetic through other signs and symptoms. Therefore, the sample size is reduced for these analyses compared to other impact measure analyses. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for HbA1c level. The sample sizes for the presented analyses of HbA1c level are as follows: bivariate analyses (n=165), primary linear regression analyses (n=140), and longitudinal analyses (n=182).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean HbA1c level data in each study period for the overall sample as well as the intervention and primary comparison groups. The overall study sample had a mean HbA1c of 7.3% at baseline. For those who returned for a follow-up assessment, this decreased to 6.9% at 6-month follow-up and increased again at 12-month follow-up (7.0%). The intervention group began the study with a slightly higher mean HbA1c of 7.4% at baseline while the primary comparison group had a slightly lower mean HbA1c of 7.1% at baseline. For participants who returned for a follow-up visit, the intervention group mean HbA1c decreased at 6-month follow-up to 7.0% and remained the same at 12 months. For those participants in the primary comparison group who returned for a follow-up visit, the mean HbA1c decreased at 6 months to 6.8% and increased to 6.9% at 12 months. As previously noted, the difference between the two groups was not significantly significant (**Table 9**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The decrease from baseline to 12-month follow-up within both study groups for HbA1c was statistically significant.

Bivariate analyses were also performed between the intervention and primary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for HbA1c when comparing the intervention and primary comparison group at 12 months, the null hypothesis cannot be rejected. The mean HbA1c measure at 12-months was not significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, HbA1c level. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for HbA1c level were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline HbA1c level, the number of comorbidities at baseline, and baseline Spirituality Index score. Additionally, to further understand whether and how physical and behavioral health are associated in this study population, baseline PHQ-9 score was included for possible selection. This model was specified as follows:

$$Y_{(HbA1c)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_HbA1c + \beta_{10} BL\_Comorbidities + \beta_{11} BL\_Spirituality + \beta_{12} BL\_PHQ\text{-}9 + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of HbA1c level included those covariates with p-value of 0.15 or less: alcohol consumption and baseline HbA1c level. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(HbA1c)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Alcohol + \beta_5 BL\_HbA1c + \varepsilon$$

Findings

Estimates for the final model of HbA1c level are presented in **Table 44**.

Mean HbA1c level at 12 months did not differ significantly between the intervention and primary comparison group (p=0.60). Below is the selected model with each covariate's effect estimate included:

$$Y_{(HbA1c)} = 1.88 + -0.09(Intervention) + 0.003(Age) + 0.23(Male) + 0.64(No\ Alcohol\ Use) + 0.62(BL\_HbA1c) + \varepsilon$$

Table 44. Effect of IBH Intervention on Twelve Month HbA1c Level, Full Mercy Sample (Primary Comparison Group), Known or Suspected Diabetics Only

Variable Selected	HbA1c (n=140)		
	Estimate (β)	Standard Error	P-value
Intervention	-0.09	0.18	0.60
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.003	0.01	0.76
Male[a]	0.23	0.28	0.41
Female (ref)	--	--	--
No Alcohol	0.64	0.29	0.03

Alcohol (ref)	--	--	--
BL_HbA1c	0.62	0.05	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for HbA1c level (not shown). The models estimated included interaction terms between intervention group and baseline diabetes and baseline Spirituality Index quintiles.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. There was no significant time by intervention group interaction for HbA1c level over the 12-month study (p=0.42), indicating that the HbA1c trajectories from baseline to 6 months, and then to 12 months were not significantly different between the two study arms (**Table 45**). Adjusting for the covariates that were selected in the primary model—age, sex, and alcohol consumption—did not alter these results (not shown).

Table 45. Effect of IBH Intervention on Trajectory of HbA1c Value Across Twelve Month Study, Full Mercy Sample (Primary Comparison Group), Known or Suspected Diabetics Only

Variable	HbA1c (n=182)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	-0.16	0.20	0.42
Time*Primary Comparison (ref)	--	--	--
Time	-0.17	0.15	0.25
Intervention	0.23	0.23	0.31
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To understand whether and how the different intervention components had effects on the relationship between the intervention and HbA1c, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received. When including these four visit variables in the model for HbA1c, none of the additional variables were selected based on the selection criteria of a p-value <0.15; additional results are not presented as the results are identical to the primary model results for HbA1c.

Limitations

As noted in the SEP, the limited availability of HbA1c data could result in insufficient sample size to detect a statistical difference.

Waist Circumference

Question 8. **Do patients who participate in the Sí Three intervention experience improvements in waist circumference after 12 months when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on waist circumference, data were collected on waist circumference, and data analysis was conducted separately for males and females. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for waist circumference. The sample sizes for the presented analyses of waist circumference are as follows: bivariate analyses (n=283: 244 female, 39 male), primary linear regression analyses (n=261: 224 female, 37 male), and longitudinal analyses (n=304: 274 female, 40 male).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 64** presents the mean waist circumference data by sex in each study period for the overall sample as well as the intervention and primary comparison groups.

For males, the overall sample had a mean waist circumference of 42.2 inches at baseline. This remained the same for participants who returned at 6-month follow-up and increased slightly to 42.3 inches for those who returned at 12-month follow-up. Males in the intervention group began the study with a lower mean waist circumference of 41.5 inches at baseline while the primary comparison group had a higher mean waist circumference of 43.0 inches at baseline. For male participants who completed a follow-up assessment, the intervention group mean waist circumference increased at both 6 and 12-month follow-up to 41.9 and 42.1 inches, respectively. Male participants in the primary comparison group followed a different trend with the mean waist circumference for those who completed a follow-up decreasing overtime to 42.4 inches at 6 months and remaining the same at 12 months. As previously noted, this difference between the two groups was not statistically significant (**Table 9**).

For females, the overall sample had a mean waist circumference of 43.6 inches at baseline. This decreased to 43.3 inches for participants who returned at 6-month follow-up and again to 42.2 inches for those who returned at 12-month follow-up. Females in the intervention group and primary comparison groups had similar mean waist circumference of 43.7 and 43.5 inches respectively at baseline. For female participants who completed a follow-up assessment, the intervention group mean waist circumference decreased at both 6 and 12-month follow-up to 43.5 inches and 42.6, respectively. Female participants in the primary comparison group followed a similar trend with the mean waist circumference for those who completed a follow-up also decreasing overtime to 43.0 inches at 6 months and 41.8 inches at 12 months. As previously noted, this difference between the two groups was not statistically significant (**Table 9**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). For male participants, the increases observed from baseline to 12-month follow-up within the intervention and primary comparison groups for waist circumference were not statistically significant. For female participants, the decreases from baseline to 12-month follow-up within both the intervention and primary comparison groups for waist circumference were statistically significant.

Bivariate analyses were also performed between the intervention and primary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for waist circumference when comparing the intervention and primary comparison group at 12 months, the null hypotheses cannot be rejected. The mean waist circumference was not significantly different between the two groups for either sex when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome of waist circumference. Separate models were created for males and females and therefore sex was not included as a possible covariate in any waist circumference models. Age was selected a priori for inclusion due to the known biological influence of age on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for waist circumference were: age, primary language, marital status, smoking, alcohol consumption, employment, baseline waist circumference, the number of comorbidities at baseline, and baseline Spirituality Index score. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline PHQ-9 score was included in the initial full model for possible selection. The models were specified as follows:

$$Y_{(Waist\ Male)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Language + \beta_4 MaritalStatus + \beta_5 Smoke + \beta_6 Alcohol + \beta_7 Employment + \beta_8 BL\_Waist + \beta_9 BL\_Comorbidities + \beta_{10} BL\_Spirituality + \beta_{11} BL\_PHQ\text{-}9 + \varepsilon$$

$$Y_{(Waist\ Female)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Language + \beta_4 MaritalStatus + \beta_5 Smoke + \beta_6 Alcohol + \beta_7 Employment + \beta_8 BL\_Waist + \beta_9 BL\_Comorbidities + \beta_{10} BL\_Spirituality + \beta_{11} BL\_PHQ\text{-}9 + \varepsilon$$

Two variations of each model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of waist circumference in males included those covariates with p-value of 0.15 or less: language, marital status, and baseline waist circumference. Age was maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(Waist\ Male)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Language + \beta_4 MaritalStatus + \beta_5 BL\_Waist + \varepsilon$$

The final model of waist circumference in females included those covariates with p-value of 0.15 or less: marital status, employment, and baseline waist circumference. Age was maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(Waist\ Female)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 MaritalStatus + \beta_4 Employment + \beta_5 BL\_Waist + \varepsilon$$

Findings

Estimates for the final model of waist circumference, for males and females, are presented in **Table 46**.

Mean waist circumference for either sex at 12 months did not differ significantly between the intervention and primary comparison group (p=0.66 for males and p= 0.94 for females). Below are the selected models with each covariate's effect estimate included:

$$Y_{(\text{Waist Male})} = -0.06 + -0.31(\text{Intervention}) + 0.005(\text{Age}) + 3.12(\text{English}) + -1.08(\text{Married}) + 1.03(\text{BL\_Waist}) + \varepsilon$$

$$Y_{(\text{Waist Female})} = 2.83 + 0.04(\text{Intervention}) + -0.005(\text{Age}) + -1.21(\text{Married}) + -0.94(\text{Employed}) + 0.92(\text{BL\_Waist}) + \varepsilon$$

Table 46. Effect of IBH Intervention on Twelve Month Waist Circumference, By Sex, Full Mercy Sample (Primary Comparison Group)

Variable	Male Waist Circumference (n=37)		
	Estimate (β)	Standard Error	P-value
Intervention	-0.31	0.71	0.66
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	0.005	0.04	0.91
English	3.12	1.15	0.01
Spanish (ref)	--	--	--
Married	-1.08	0.73	0.15
Unmarried (ref)	--	--	--
BL_Waist	1.03	0.09	<0.001
Variable	Female Waist Circumference (n=224)		
	Estimate (β)	Standard Error	P-value
Intervention	0.04	0.58	0.94
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.005	0.03	0.85
Married	-1.21	0.62	0.05
Unmarried (ref)	--	--	--
Employed	-0.94	0.63	0.14
Unemployed(ref)	--	--	--
BL_Waist	0.92	0.05	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for waist circumference among males or females (not shown). The models estimated included interaction terms between intervention group and baseline obesity and baseline Spirituality Index quintiles.

Additional Analyses

Longitudinal analyses were used to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate linear mixed models, the PROC MIXED procedure in SAS was used. No significant time by intervention group interaction for waist circumference was identified for males (p=0.52) or females (p=0.72), indicating that the trajectories from baseline to 6 months, and then to 12 months were not different between the two study arms for waist circumference among males or females separately (**Table 47**). Adjusting for the covariates that were selected in the primary model—age, language, and marital status —did not alter these results.

Table 47. Effect of IBH Intervention on Trajectory of Waist Circumference, By Sex, Full Mercy Sample (Primary Comparison Group)

Variable	Male Waist Circumference (n=40)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	0.44	0.68	0.52
Time*Primary Comparison (ref)	--	--	--
Time	0.70	0.48	0.16
Intervention	-1.57	1.44	0.28
Primary Comparison (ref)	--	--	--
Variable	Female Waist Circumference (n=274)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	0.19	0.53	0.72
Time*Primary Comparison (ref)	--	--	--
Time	-1.81	0.37	<0.001
Intervention	0.37	0.59	0.52
Primary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistical significance (p-value < 0.05)

To understand whether and how the different intervention components had effects on the relationship between the intervention and waist circumference, an additional linear regression model was examined. This model utilized the same backwards selection methods as the primary linear regression model and included the same set of prospective predictors for possible selection with the addition of four continuous variables representing the number of primary care, behavioral health, health education, and exercise coaching visits participants received.

When including these four visit variables in the model for waist circumference for males, the variable for number of behavioral health visits was selected based on the selection criteria of a p-value <0.15. The results of the model with this additional variable are presented in
Table 48. Accounting for the number of behavioral health visits in the model, the intervention is not significantly associated with a difference in waist circumference among males at 12 months, holding all other variables in the selected model constant (p=0.44).

When including these four visit variables in the model for waist circumference for females, the variable for number exercise coaching visits was selected based on the selection criteria of a p-value <0.15. The results of the model with this additional variable are presented in
Table 48. Accounting for the number of exercise coaching visits in the model, the intervention is not significantly associated with a difference in waist circumference among females at 12 months, holding all other variables in the selected model constant(p=0.62).

Table 48. Effect of IBH Intervention on Twelve Month Waist Circumference Including Number of Visits, Full Mercy Sample (Primary Comparison Group)

Variable	Male Waist Circumference (n=37)		
	Estimate (β)	Standard Error	P-value
Intervention	0.77	0.98	0.44
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.01	0.04	0.76
English	2.94	1.13	0.01
Spanish (ref)	--	--	--
Married	-1.19	0.72	0.11
Unmarried (ref)	--	--	--
BL_Waist	0.97	0.10	<0.001
Number of behavioral health visits	-0.50	0.32	0.13
Variable	Female Waist Circumference (n=224)		
	Estimate (β)	Standard Error	P-value
Intervention	0.30	0.61	0.62
Primary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.01	0.03	0.85
Married	-1.15	0.62	0.06
Unmarried (ref)	--	--	--
Employed	-1.07	0.64	0.09
Unemployed(ref)	--	--	--
BL_Waist	0.92	0.05	<0.001
Number of exercise coaching visits	-0.03	0.02	0.14

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

A limitation to consider for this measure is the smaller sample size for males. Because of the smaller sample size, a statistical difference may be more difficult to detect. Also, compared to the other impact measures, there was a slightly greater amount of missing data on this measure; however, multiple imputation approaches were considered, but determined not necessary as the missing data were not substantial.

Secondary Comparison Group Analysis and Results

Final impact study results using the secondary comparison group are presented by research question. This section serves as a sensitivity analyses assessing the external validity of the primary comparison analyses. Included in this section are details the statistical methods used, noting any deviations from what was planned in the SEP based on field conditions and analytic judgment at the time of analysis, and findings for the secondary assessment of data collected for the Mercy study.

Descriptive statistics for complete data are examined in this report for the intervention and secondary comparison group. These statistics include patients' sociodemographic characteristics and other key covariates. These covariates were examined to assist in identifying potential factors that may result in nonequivalence between the two groups. Chi-square tests, and Fisher's Exact Tests when necessary based on cell counts, were used for comparison of categorical data to examine baseline equivalence. Two sample t-tests were used for continuous data that were normally distributed, and the Wilcoxon Signed Rank test was used for non-normally distributed data. Because a nonequivalent comparison group design is employed in the study, an intent-to-treat analysis will be conducted with adjustment of potential nonequivalence of covariates and baseline outcome measure. The decision was made not to perform secondary power calculations as the final sample size was just shy of the target and prior research indicated that these tests are not necessarily helpful in the interpretation of observed results (Goodman and Berlin, 1994).

All descriptive, baseline equivalence, bivariate, multivariate, and longitudinal analyses reported in this final report were performed with SAS version 9.4 (Cary, NC). PROC GLM was utilized for the primary linear regression models. For impact measures that were assessed to be non-normally distributed, analyses were conducted using both PROC GLM and PROC GENMOD in order to assess any possible bias deriving from the non-normality. For linear regression models, using normal linear regression methods (e.g., PROC GLM) produced results consistent with those produced with methods accounting for the non-normality of these data (e.g., PROC GENMOD). Differences were considered statistically significant at $p<0.05$. Effect sizes were calculated for both confirmatory outcomes regardless of statistical significance of model results and for any exploratory outcome with a statistically significant result. Results are presented in the "Findings" section under research questions when applicable. The statistic utilized for these calculations was Cohen's d using the following equation:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s} = \frac{\mu_1 - \mu_2}{s}$$

Unit of Analysis and Overview of Analyses Performed

The unit of analysis was the individual patient. An "end-point" analysis was our primary analytic approach. This "end-point" analysis approach is a conventional approach to analyze clinical trial data collected from individuals with both baseline data and end-point data of primary interest (Liebschutz, et al., 2017). We employed generalized regression analysis following a modeling sequence from bivariate models to multiple regression models adjusting for baseline levels of outcome measures and covariates assessed to be relevant based on review of the scientific literature or found to be unbalanced between the two groups at baseline. The parameter of interest was the dichotomous variable that differentiates the treatment status (i.e., intervention vs. secondary comparison). Between-group comparison of baseline and single follow-up outcomes were assessed by end-point analyses that accounted for the

baseline level of impact measures. Additionally, because multiple follow-up impact measures form individual trajectories, we conducted longitudinal analyses assessing whether the impact measure trajectories differ by intervention status (Fitzmaurice et al., 2004). A time measure was developed and applied to denote baseline, 6 and 12-month follow-up measures.

In addition to adjusting for key covariates, potential collinearity and its impact on the standard error estimates for the covariates in the model was evaluated by examining variance inflation factor when necessary. As stated in the SEP, in areas where multiple comparisons are necessary, adjustment of the p-value was used to account for multiple comparisons, such as the Bonferroni correction. This step was ultimately unnecessary for the executed analyses since there was no need to account for multiple comparisons.

To evaluate the intervention effect, a multiple linear regression model approach was used following a sequence of interrelated models. The analysis sequence began by developing a bivariate model regressing the follow-up outcome measure on intervention status (intervention vs. secondary comparison) followed by the estimation of an adjusted model accounting for the baseline measure of interest and further adjustment for key covariates. Parametric two sample t-tests were used for bivariate analysis of one of the confirmatory impact measures (BMI) as well as some of the exploratory study outcomes (blood pressure, and waist circumference). One of the confirmatory variables (PHQ-9)) and some exploratory outcomes (GAD-7, HbA1c, Duke General Health) were found to be non-normally distributed. In these bivariate analyses, nonparametric Wilcoxon Rank Sum tests were conducted due to the increased sensitivity to detect a difference in non-normally distributed data. The nonparametric results are presented throughout this report; however, additional parametric t-tests were performed for these measures to align with linear regression methods for the final analyses. Though the parametric results are not presented, both the nonparametric and parametric bivariate analyses produced consistent results.

Following bivariate comparisons, multivariate and longitudinal analyses were performed separately to answer each research question. As previously mentioned, multiple imputation methods were not necessary due to the near completeness of the submitted data. Propensity score matching was explored but was not included in the outcome analyses for the reasons described previously in the methods section of this report. The primary adjusted multivariate analysis models the outcome of interest on intervention status with relevant covariates included. The longitudinal analysis evaluates whether the impact measure trajectories differ by intervention status across the 12-month study. Effect modification of the intervention-outcome relationship also was examined. Possible effect modification of baseline health condition was explored for the corresponding impact measure (e.g. baseline depression as an effect modifier for impact on PHQ-9 score at 12 months).

The SEP indicated a set of planned covariates for adjustment in the models. Of those listed, age (continuous and categorical), sex, employment, number of comorbidities, and time were included in one or more of the analyses. Categorical age was operationally defined by the following categories: 18-24-year-olds, 25-34-year-olds, 35-44-year-olds, 45-54-year-olds, 55-64-year-olds, and those who are 65 years or older. Employment was included as a dichotomized variable with categories of “employed”, including employed, migrant farm workers, and self-employed participants, and “not employed”, including unemployed and student participants. As anticipated, the study population was fairly homogeneous on ethnicity and thus this was not included in the final models. Additional data on characteristics of the study populations including primary language, smoking, alcohol consumption, and

marital status were included for possible selection in one or more of the analyses. Marital status was considered a dichotomous variable with categories or "married", including only those who indicated they were married, and "not married", which includes all other categories for the marital status variable.

A backward elimination modeling selection procedure was used for an end-point analysis approach where covariates with p-value larger than 0.15 were excluded from the final model for parsimony. In some cases, age and sex were selected for inclusion in statistical models a priori due to the known biological influence of these characteristics on health outcomes; this is noted where relevant under each research question. For some research questions, predictor variables were included that could be correlated with the outcome of interest. Where relevant, the variation inflation factor (VIF) is reported in the model selection process. Using PROC CORR, the range of correlation between the predictors included in the model and the outcomes of interest is -0.51, the Pearson coefficient for baseline PHQ-9 score and 12-month Duke General Health score, to 0.97, the Pearson coefficient for baseline BMI and 12-month BMI.

Results for each of the outcome variables are presented separately by research question. At the end of this section, **Table 64** presents the mean scores for each of the outcome variables.

Depressive Symptoms

Question 1. **Do patients who participate in the Sí Three intervention experience improvements in depressive symptoms, as measured by PHQ-9, after 12 months compared to patients who do not participate? This question is <u>confirmatory</u>.**

<u>Overview of Analysis</u>
To answer this confirmatory question about intervention impact on depressive symptoms, data were analyzed from the PHQ-9 assessment tool. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for PHQ-9 score. The sample sizes for the presented analyses of PHQ-9 score are as follows: bivariate analyses (n=397), primary linear regression analyses (n=392), and longitudinal analyses (n=460).

<u>Descriptive Statistics and Bivariate Comparisons</u>
At the end of this section, **Table 65** presents the mean PHQ-9 score data in each study period for the overall sample as well as the intervention and secondary comparison groups. The overall sample had a mean PHQ-9 score of 3.8 at baseline. This decreased to 3.1 for participants who returned at 6-month follow-up and again to 2.0 for those who returned at 12-month follow-up. The intervention group began the study with a higher mean PHQ-9 score of 6.7 at baseline while the secondary comparison group had a lower mean PHQ-9 score of 2.1 at baseline. Aligning with the overall sample trend, for participants who completed a follow-up assessment, the intervention group mean PHQ-9 score decreased at both 6 and 12-month follow-up to 4.9 and 2.9 respectively. The secondary comparison group also followed this trend with the mean PHQ-9 score for those who completed a follow-up decreasing overtime to 1.9 at 6 months and 1.5 at 12 months. As previously noted, this difference between the two groups was statistically significant (see **Table 11**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional

covariates (**Table 66**). The decreases in PHQ-9 from baseline to 12-month follow-up within both the intervention and secondary comparison groups were statistically significant.

Bivariate analyses also were performed between the intervention and secondary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value less than 0.05 for PHQ-9 score when comparing the intervention and secondary comparison group at 12 months, the null hypothesis can be rejected. The mean PHQ-9 score was significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, PHQ-9 score. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for PHQ-9 score were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline PHQ-9 score, baseline Duke General Health score, baseline GAD-7 score, and the number of comorbidities at baseline. The inclusion of baseline PHQ-9 score controlled for the statistical imbalance between intervention and secondary comparison groups at baseline. Additionally, to further understand whether and how physical and behavioral health are associated in this study population, baseline BMI was included for possible selection.

$$Y_{(PHQ\text{-}9PHQ9)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_PHQ\text{-}9 + \beta_{10} BL\_General + \beta_{11} BL\_GAD\text{-}7 + \beta_{12} BL\_Comorbidities + \beta_{13} BL\_BMI + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, a multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of PHQ-9 score included those covariates with p-value of 0.15 or less: marital status, smoking, baseline PHQ-9 score, baseline Duke General Health score, and baseline BMI. Age and sex were maintained based on a priori selection. Age was modeled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(PHQ\text{-}9PHQ9)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 MaritalStatus + \beta_5 Smoke + \beta_6 BL\_PHQ\text{-}9PHQ9 + \beta_7 BL\_General + \beta_8 BL\_BMI + \varepsilon$$

Because baseline quality of life measure was selected for inclusion into the final model of depressive symptoms, and quality of life and depression are known to be related, an additional test was conducted to quantify any multicollinearity between the Duke General Health and PHQ-9 scores. The variance inflation factor (VIF) of Duke General Health score in the PHQ-9 score model was 2.2, below the commonly accepted cutoff of 10 indicating minimal influence on the variance from the correlation of these variables (O'Brien, 2007).

Findings

Estimates by covariate for the final model of PHQ-9 score are presented in **Table 49**.
Mean PHQ-9 score at 12 months did not differ significantly between the intervention and secondary comparison group (p=0.76); the effect size (using Cohen's d) is 0.03. Below is the selected model with each covariate's effect estimate included:

$$Y_{(PHQ\text{-}9PHQ9)} = 1.68 + -0.11(\text{Intervention}) + 0.01(\text{Age}) + -0.37(\text{Male}) + 0.46(\text{Married}) + 1.34(\text{Current Smoker}) + 1.45(\text{Former Smoker}) + 0.28(\text{BL\_PHQ-9}) + -0.04(\text{BL\_General}) + 0.06(\text{BL\_BMI}) + \varepsilon$$

Table 49. Effect of IBH Intervention on Twelve Month PHQ-9 Score, Full Mercy Sample (Secondary Comparison Group)

Variable	PHQ-9 (n=392)		
	Estimate (β)	Standard Error	p-value
Intervention	-0.11	0.35	0.76
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	0.01	0.01	0.49
Male[a]	-0.37	0.37	0.32
Female (ref)	--	--	--
Married	0.46	0.30	0.12
Unmarried (ref)	--	--	--
Current smoker	1.34	0.62	0.03
Former smoker	1.45	0.66	0.03
Never smoker (ref)	--	--	--
BL_PHQ-9	0.28	0.05	<0.001
BL_General	-0.04	0.01	0.001
BL_BMI	0.06	0.02	0.003

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for PHQ-9 score (not shown). The model estimated included an interaction term of study group and baseline depression.

Additional Analyses

Longitudinal analyses were used to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. For PHQ-9 score, only adjusting for intervention status and time, there was a significant time/group interaction with a p-value of <0.001, indicating that the trajectories from baseline to 6 months, and then to 12 months were different between the two study arms for PHQ-9 score (see **Table 50**). For each 6-month period in the study, the intervention group experienced a relative decrease of 2.84 points in PHQ-9 score compared to the secondary comparison group. Adjusting for the covariates that were selected in the primary model—age, sex, marital status, smoking, baseline Duke General Health score, and baseline BMI —did not alter these results.

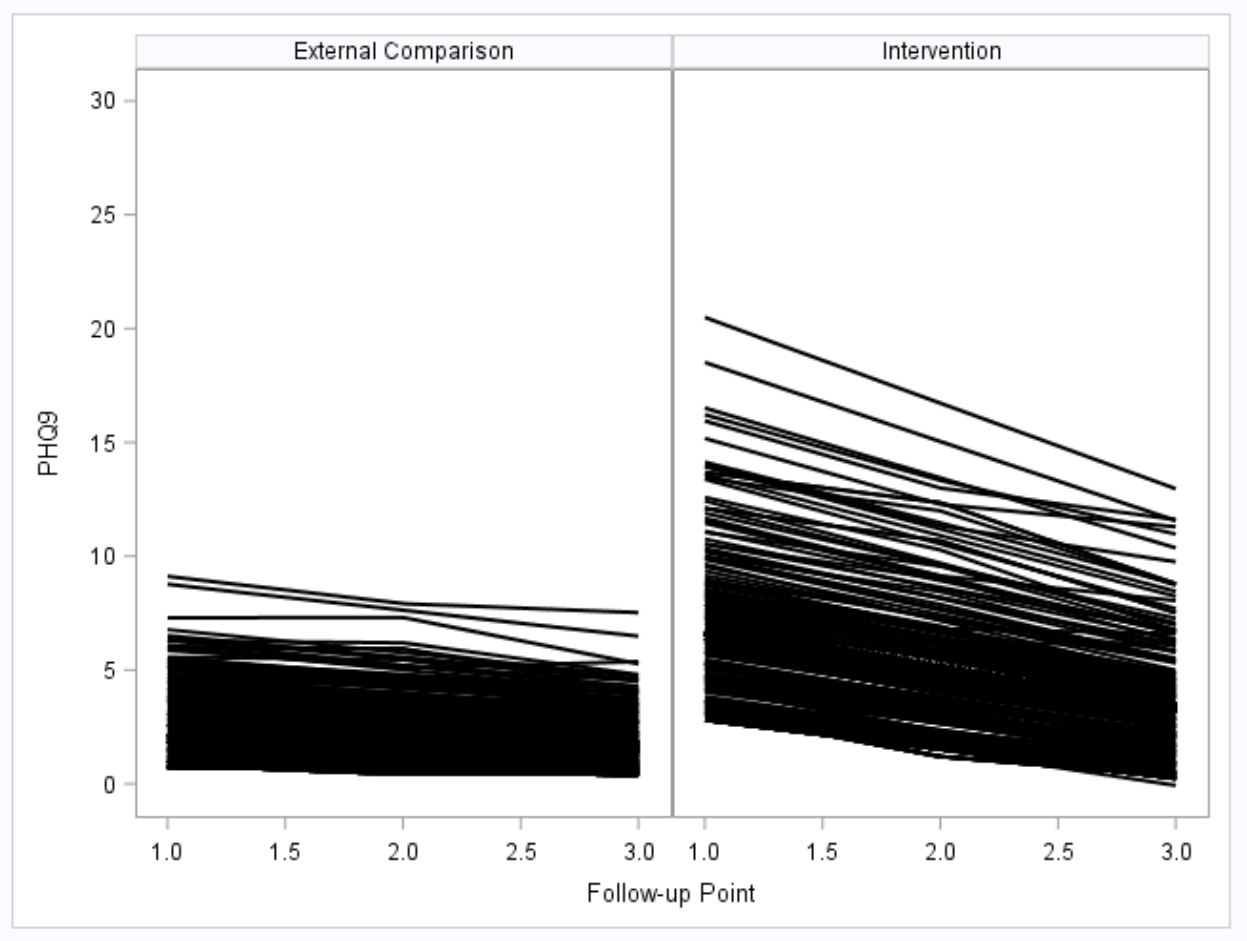
Table 50. Effect of IBH Intervention on Trajectory of PHQ-9 Score Across Twelve Month Study, Full Mercy Sample (Secondary Comparison Group)

Variable	PHQ-9 (n=460)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	**-2.84**	**0.39**	**<0.001**
Time*Secondary Comparison (ref)	--	--	--
Time	-0.67	0.23	0.004
Intervention	4.51	0.36	<0.001
Secondary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To visualize the longitudinal effect of the intervention on PHQ-9 score, a two-panel spaghetti plot was produced using PROC SGPANEL. In **Figure 9**, the secondary comparison group trajectory appears in the left panel and the intervention group trajectory appears in the right panel. The x-axis of the graph shows the study follow-up points with 1.0 representing baseline, 2.0 is the 6-month point, and 3.0 is the 12-month end-point. Looking at the trajectories, the two groups clearly differ from one another. The intervention group's PHQ-9 score measurements start higher and are more variable than in the secondary comparison group. The trend in the secondary comparison group trajectories appears much flatter than the sharp decrease seen in the intervention PHQ-9 scores over time.

Figure 9. Individual Trajectories of PHQ-9 Score Over Twelve Month Study Period for Intervention and Secondary External Comparison Group



Limitations

NCDV staff verbally administered the assessment to participants while most Mercy participants self-administered the assessment. This may have resulted in instrumentation bias.

Body Mass Index

Question 2. **Do patients who participate in the Sí Three intervention experience improvements in BMI after 12 months when compared to patients that do not participate in the intervention? This question is confirmatory.**

Overview of Analysis

To answer this confirmatory question about intervention impact on patient BMI, data on weight, height, and BMI were collected and analyzed. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for BMI. The sample sizes for the presented analyses of BMI are as follows: bivariate analyses (n=396), primary linear regression analyses (n=391), and longitudinal analyses (n=461).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 65** presents the mean BMI data in each study period for the overall sample as well as the intervention and secondary comparison groups. The overall sample had a mean BMI of 34.6 kg/m^2 at baseline. This increased to 34.7 kg/m^2 for those who returned at 6-month follow-up with a further increase at 12 months for those who completed a follow-up (34.9 kg/m^2). The intervention group began the study with a lower mean BMI of 33.2 kg/m^2 at baseline while the secondary comparison group had a higher mean BMI of 35.4 kg/m^2 at baseline. Aligning with the overall sample trend, for those who completed an assessment at follow-up, the intervention group mean BMI increased to 33.3 kg/m^2 at 6-month follow-up and again to 34.0 kg/m^2 at 12 months. In the secondary comparison group, the mean BMI increased from baseline to 6 months to 35.6 kg/m^2 and decreased at 12 months for those who completed a follow-up assessment (35.3 kg/m^2). As previously noted, this difference between the two groups was not statistically significant (**Table 11**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The increases from baseline to 12-month follow-up within both the intervention and secondary comparison groups for BMI were not statistically significant.

Bivariate analyses were also performed between the intervention and secondary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value less than 0.05 for BMI when comparing the intervention and secondary comparison group at 12 months, the null hypothesis can be rejected. The mean BMI measure at 12 months was significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, BMI. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for BMI were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline BMI, and the number of comorbidities at baseline. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline PHQ-9 score was included for possible selection in the initial full model. This model was specified as follows:

$$Y_{(BMI)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_BMI + \beta_{10} BL\_Comorbidities + \beta_{11} BL\_PHQ\text{-}9 + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model for BMI included those covariates with p-value of 0.15 or less: smoking, baseline BMI, the number of comorbidities at baseline, and baseline PHQ-9. Continuous age and sex were forced in as

predictors due to the known biological influence of age on health outcomes. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(BMI)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Smoke + \beta_5 BL\_BMI + \beta_6 BL\_comorbidities + \beta_7 BL\_PHQ\text{-}9 + \varepsilon$$

Findings

Estimates by covariate for the final model of BMI are presented in **Table 51**.

Mean BMI at 12 months did not differ significantly between the intervention and secondary comparison group (p=0.21); the effect size (using Cohen's d) is 0.04. Below is the selected model with each covariate's effect estimate included:

$$Y_{(BMI)} = 1.38 + 0.31(Intervention) + -0.02(Age) + 0.05(Male) + -1.22(Current\ Smoker) + 0.81(Former\ Smoker) + 0.98(BL\_BMI) + -0.18(BL\_Comorbidities) + 0.06(BL\_PHQ9) + \varepsilon$$

Table 51. Effect of IBH Intervention on Twelve Month BMI, Full Mercy Sample (Secondary Comparison Group)

Variable	BMI (n=391)		
	Estimate (β)	Standard Error	p-value
Intervention	0.31	0.24	0.21
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.02	0.01	0.09
Male	0.06	0.25	0.82
Female (ref)	--	--	--
Current Smoker	-1.22	0.41	0.003
Former Smoker	0.81	0.44	0.07
Never Smoker (ref)	--	--	--
BL_BMI	0.98	0.01	<0.001
BL_Comorbidities	-0.18	-0.11	0.10
BL_PHQ-9	0.06	0.03	0.02

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for BMI. The model considered included an interaction term of study group and baseline obesity.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. For BMI, only adjusting for intervention status and time, there was no significant time/group interaction with a p-value of 0.14, indicating that the trajectories from baseline to 6 months, and then to 12 months were not different between the two study arms for BMI

(see **Table 52**). Adjusting for the covariates that were selected in the primary model—age, sex, smoking, baseline PHQ-9 score, and number of comorbidities at baseline—did not alter these results.

Table 52. Effect of IBH Intervention in Trajectory of BMI Across Twelve Month Study Follow-Up, Full Mercy Sample (Secondary Comparison Group)

Variable	BMI (n=461)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	0.30	0.20	0.14
Time*Secondary Comparison (ref)	--	--	--
Time	-0.19	0.12	0.12
Intervention	-2.22	0.65	0.001
Secondary Comparison (ref)	--	--	--

Note: "ref" indicates the reference category used to calculate the estimate for a covariate

Limitations

There are no limitations specific to this measure to note.

Functioning and Quality of Life

Question 3. **Do patients who participate in the Sí Three intervention experience improvements in quality of life, as measured by the Duke Health Profile, after 12 months when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on functioning and quality of life, data were analyzed from the Duke Health Profile, specifically the General Health score. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for the Duke Health Profile. Analyses were also conducted on the Duke Health Profile components that comprise the General Health score: Physical Health, Mental Health, and Social Health scores. The sample sizes for the presented analyses of Duke General Health score are as follows: bivariate analyses (n=397), primary linear regression analyses (n=392), and longitudinal analyses (n=460).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 65** presents the mean Duke General Health score data in each study period for the overall sample as well as the intervention and secondary comparison groups. The overall sample had a mean Duke General Health score of 76.6 at baseline. This increased to 79.7 for participants who returned at 6-month follow-up and again to 83.3 for those who returned at 12-month follow-up. The intervention group began the study with a lower mean Duke General Health score of 67.7 at baseline while the secondary comparison group had a higher mean Duke General Health score of 81.6 at baseline. Aligning with the overall sample trend, for participants who completed a follow-up assessment, the intervention group mean Duke General Health score increased at both 6 and 12-month follow-up to 72.9 and 78.5 respectively. The secondary comparison group also followed this trend with the mean Duke General Health score for those who completed a follow-up increasing overtime to 84.1 at 6 months and 85.9 at 12 months. As previously noted, this difference between the two groups was statistically significant (**Table 11**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The increases observed from baseline to 12-month follow-up within both the intervention and secondary comparison groups for Duke General Health score were statistically significant.

Bivariate analyses were also performed between the intervention and secondary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value less than 0.05 for Duke General Health score when comparing the intervention and secondary comparison group at 12 months, the null hypothesis can be rejected. The mean Duke General Health score was significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, Duke General Health score. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for Duke General Health score were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline Duke General Health score, baseline PHQ-9 score, baseline GAD-7 score, the number of comorbidities at baseline, and baseline Spirituality Index score. The inclusion of baseline Duke General Health score controlled for the statistical imbalance between intervention and primary comparison groups at baseline. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline BMI was included for possible selection. This model was specified as follows:

$$Y_{(DUKE\ General)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_General + \beta_{10} BL\_PHQ\text{-}9 + \beta_{11} BL\_GAD\text{-}7 + \beta_{12} BL\_Comorbidities + \beta_{13} BL\_BMI + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model for Duke General Health score included those covariates with p-value of 0.15 or less: age, baseline Duke General Health score, baseline PHQ-9 score, and baseline BMI. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(DUKE\ General)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 BL\_General + \beta_5 BL\_PHQ9 + \beta_6 BL\_BMI + \varepsilon$$

Because the baseline depression measure was selected for inclusion in the final model of quality of life, an additional test was conducted to quantify any multicollinearity between the Duke General Health score with PHQ-9 score. The variance inflation factor (VIF) of PHQ-9 score in the Duke General Health

score model was 2.3, below the commonly accepted cutoff of 10 indicating minimal influence on the variance from the correlation of these variables (O'Brien, 2007).

Findings

Estimates by covariate for the final model of Duke General Health score are presented in **Table 53**.

Mean Duke General Health score at 12 months did not differ significantly between the intervention and secondary comparison group (p=0.83). Below is the selected model with each covariate's effect estimate included:

$$Y_{(\text{Duke General})} = 70.42 + 0.30(\text{Intervention}) + -0.09(\text{Age}) + -1.69(\text{Male}) + 0.42(\text{BL\_General}) + -0.63(\text{BL\_PHQ9}) + -0.35(\text{BL\_BMI}) + \varepsilon$$

Table 53. Effect of IBH Intervention on Twelve Month Duke General Health Score, Full Mercy Sample (Secondary Comparison Group)

Variable	Duke General Health (n=392)		
	Estimate (β)	Standard Error	p-value
Intervention	0.30	1.45	0.83
Secondary Comparison (ref)	--	--	--
Age (continuous)	-0.09	-.06	0.12
Male[a]	-1.69	1.54	0.27
Female (ref)	--	--	--
BL_General	0.42	0.05	<0.001
BL_PHQ-9	-0.63	0.21	0.003
BL_BMI	-0.35	0.08	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

As previously noted, models were created to examine the individual components of the composite Duke General Health score. These analyses aimed to further understand the statistically significant improvement in quality of life in the intervention group. The three component scores began with the same possible model for selection as the General Health score, substituting the corresponding baseline Duke Health Profile score for the baseline General Health score. The models for Physical Health and Mental Health scores also were not statistically significant when comparing the intervention to the secondary comparison group (not shown).

When examining Social Health score as the outcome of interest, the results were statistically significant (p=<0.001). On average, for participants in the intervention group, there is a 6.41-point decrease in Duke Social Health at 12 months holding all other variables in the model constant compared to those in the secondary comparison group. These results are presented in **Table 54**.

$$Y_{(\text{DUKE Social})} = 76.00 + -6.41(\text{Intervention}) + 0.04(\text{Age}) + -5.29\ (\text{Male}) + 0.20(\text{BL\_Social}) + -0.93\ (\text{BL\_PHQ-9PHQ9}) + \varepsilon$$

Table 54. Effect of IBH Intervention on Twelve Month Duke Social Health Score, Full Mercy Sample (Secondary Comparison Group)

Variable	Duke Social Health (n=392)		
	Estimate (β)	Standard Error	p-value
Intervention	**-6.41**	**1.76**	**<0.001**
Secondary Comparison (ref)	--	--	--
Age (continuous)	0.04	0.07	0.59
Male[a]	-5.29	1.80	0.004
Female (ref)	--	--	--
BL_Social	0.20	0.04	<0.001
BL_PHQ-9	-0.93	0.20	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no effect modifications for Duke General Health score considered as there is no corresponding condition in which differences in effect may exist.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. There was a significant time by intervention group interaction for Duke General Health score over the 12-month study (p<0.001), indicating that Duke General Health score trajectories from baseline to 6 months, and then to 12 months were significantly different between the intervention group and the secondary comparison group (**Table 55**). For each 6-month period in the study, the intervention group experienced a relative increase of 6.24 points in Duke General Health score compared to the secondary comparison group. Adjusting for the covariates that were selected in the primary model—age, sex, baseline PHQ-9 score, and baseline BMI—did not alter these results (not shown).

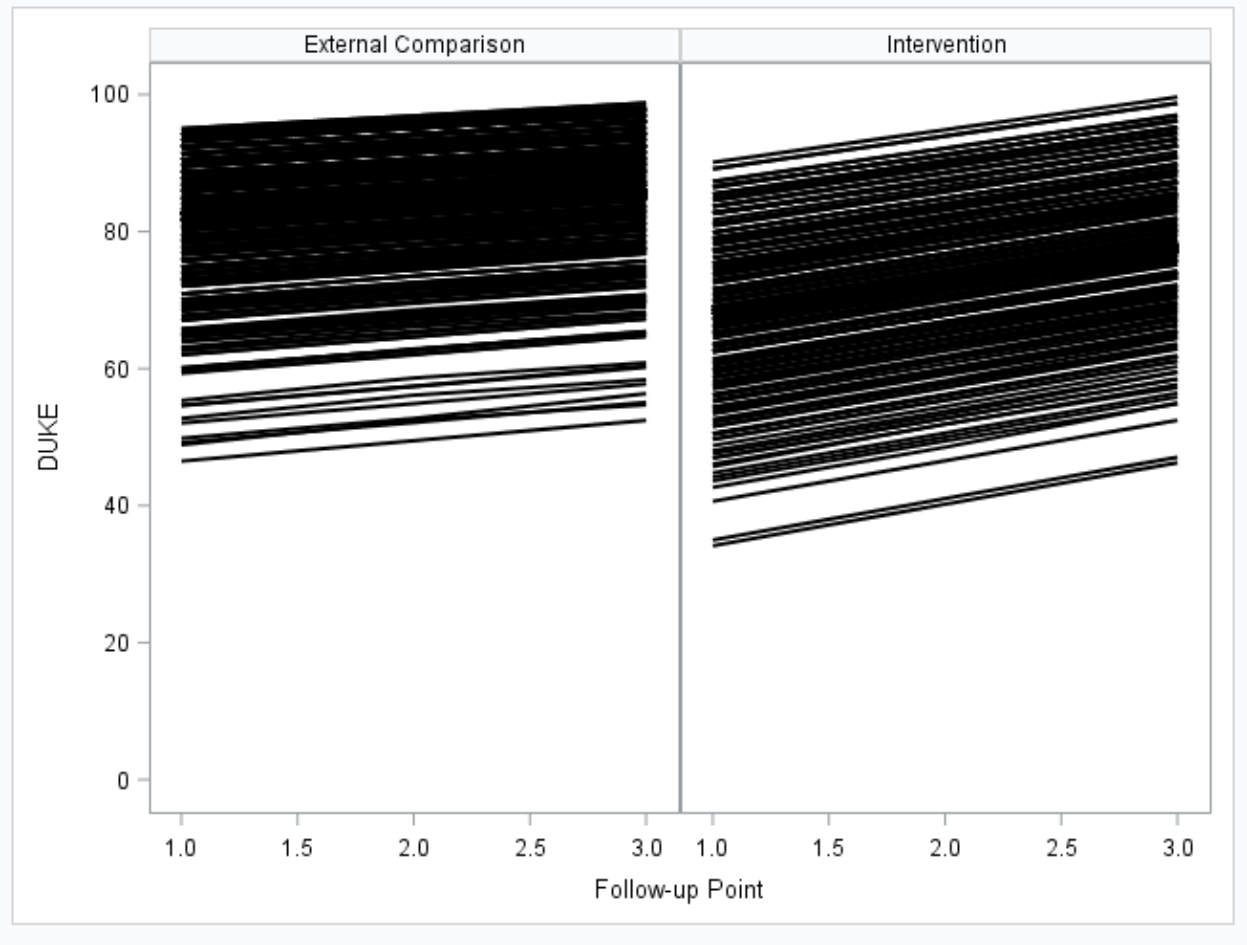
Table 55. Effect of IBH Intervention on Trajectory of Duke General Health Score Across Twelve Month Study, Full Mercy Sample (Secondary Comparison Group)

Variable	Duke General Health (n=460)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	**6.24**	**1.47**	**<0.001**
Time*Secondary Comparison (ref)	--	--	--
Time	4.32	0.87	<0.001
Intervention	-13.85	1.36	<0.001
Secondary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

To visualize the longitudinal effect of the intervention on Duke General Health score, a two-panel spaghetti plot was produced using PROC SGPANEL.
Figure 10 displays the secondary comparison group trajectory in the left panel and the intervention group trajectory in the right panel. The x-axis of the graph shows the study follow-up points with 1.0 representing baseline, 2.0 is the 6-month point, and 3.0 is the 12-month end-point. The trajectory figure visually displays the differences identified in the longitudinal statistical model, identifying a lower Duke General Health score measurement for the intervention group compared to the secondary comparison group. The increase seen in the trajectories from baseline to 12 months is steeper for those in the intervention than in the secondary comparison group reflecting the greater increase over time for the intervention participants.

Figure 10. Individual Trajectories of DUKE General Health Score Across Twelve Month Study Period for IBH Intervention and Secondary External Comparison Groups



Limitations
According to Mercy clinic staff, participants experienced difficulty completing the Duke assessments at each time point due to Spanish translations of the instrument which used language unfamiliar to Mercy patients. NCDV staff reported that the secondary comparison group participants experienced challenges understanding one item on the assessment. Also, NCDV staff verbally administered the assessment while most Mercy participants self-administered the assessment. This may have resulted in an

instrumentation bias. *(Note: The Duke Spanish language surveys used in the Sí Texas study had been validated in the literature and HRiA conducted focus groups in the study area to ensure that the survey language was regionally appropriate).*

Anxiety Symptoms

Question 4. **Do patients who participate in the Sí Three intervention experience improvements in anxiety symptoms, as measured by GAD-7, after 12 months compared to patients who do not participate?**

Overview of Analysis

To answer this exploratory question about intervention impact on anxiety, data were analyzed from the GAD-7 assessment tool. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for GAD-7 score. The sample sizes for the presented analyses of GAD-7 score are as follows: bivariate analyses (n=397), primary linear regression analyses (n=392), and longitudinal analyses (n=460).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 65** presents the mean GAD-7 score data in each study period for the overall sample as well as the intervention and secondary comparison groups. The overall sample had a mean GAD-7 score of 3.2 at baseline. This decreased to 2.7 for participants who returned at 6-month follow-up and again to 1.7 for those who returned at 12-month follow-up. The intervention group began the study with a higher mean GAD-7 score of 6.2 at baseline while the secondary comparison group had a lower mean GAD-7 score of 1.5 at baseline. Aligning with the overall sample trend, for participants who completed a follow-up assessment, the intervention group mean GAD-7 score decreased at both 6 and 12-month follow-up to 4.6 and 2.6 respectively. In the secondary comparison group, for those who completed a follow-up, the mean GAD-7 score was constant through 6 months and decreased to 1.2 at 12 months. As previously noted, this difference between the two groups was statistically significant (**Table 11**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The decreases in mean GAD-7 score from baseline to 12-month follow-up within both the intervention and secondary comparison groups were statistically significant.

Bivariate analyses were also performed between the intervention and secondary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value less than 0.05 for GAD-7 score when comparing the intervention and secondary comparison group at 12 months, the null hypothesis can be rejected. The mean GAD-7 score at 12-months was significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, GAD-7 score. Age and sex were selected a priori for

inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for GAD-7 score were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline GAD-7 score, baseline PHQ-9 score, baseline Duke General Health score, and the number of comorbidities at baseline. The inclusion of the baseline GAD-7 score controlled for the statistical imbalance between intervention and secondary comparison groups at baseline. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline BMI was included for possible selection. This model was specified as follows:

$$Y_{(GAD\text{-}7)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_GAD\text{-}7 + \beta_{10} BL\_PHQ\text{-}9 + \beta_{11} BL\_General + \beta_{12} BL\_Comorbidities + \beta_{13} BL\_BMI + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, a multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of GAD-7 score included those covariates with p-value of 0.15 or less: smoking, baseline GAD-7 score, baseline Duke General Health score, and baseline BMI. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(GAD\text{-}7)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Smoke + \beta_5 BL\_GAD\text{-}7 + \beta_6 BL\_General + \beta_7 BL\_BMI + \varepsilon$$

Because the baseline quality of life measure was selected for inclusion in the final model of anxiety, and quality of life and anxiety are known to be related, an additional test was conducted to quantify any multicollinearity between the General Health score with GAD-7 score. The variance inflation factor (VIF) of the Duke General Health score in the GAD-7 score model was 1.9, below the commonly accepted cutoff of 10 indicating minimal influence on the variance from the correlation of these variables (O'Brien, 2007).

Findings

Estimates by covariate for the final model of GAD-7 score are presented in **Table 56**.

Mean GAD-7 score at 12 months did not differ significantly between the intervention and secondary comparison group (p=0.93). Below is the selected model with each covariate's effect estimate included:

$$Y_{(GAD\text{-}7)} = 1.69 + 0.03(Intervention) + - <0.001(Age) + -0.31(Male) + 0.53(Current\ Smoker) + 1.27(Former\ Smoker) + 0.24(BL\_GAD\text{-}7) + -0.03(BL\_General) + 0.04(BL\_BMI) + \varepsilon$$

Table 56. Effect of IBH Intervention on Twelve Month GAD-7 Score, Full Mercy Sample (Secondary Comparison Group)

Variable Selected	GAD-7 (n=392)		
	Estimate (β)	Standard Error	P-value
Intervention	0.03	0.30	0.93
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	- <0.001	0.01	0.99
Male[a]	-0.31	0.32	0.33
Female (ref)	--	--	--
Current Smoker	0.53	0.53	0.32
Former Smoker	1.27	0.57	0.03
Never Smoker (ref)	--	--	--
BL_GAD-7	0.24	0.04	<0.001
BL_General	-0.03	0.01	0.002
BL_BMI	0.04	0.02	0.01

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for GAD-7 score (not shown). The model considered included an interaction term of study group and baseline anxiety.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. There was a significant time by intervention group interaction for GAD-7 score over the 12 month study (p<0.001), indicating that the GAD-7 trajectories from baseline to 6 months, and then to 12 months were different between the intervention group compared to the secondary comparison group (**Table 57**).

For each 6-month period in the study, the intervention group experienced a relative decrease of 3.09 points in GAD-7 score compared to the secondary comparison group. Adjusting for the covariates that were selected in the primary model—age, sex, smoking, baseline Duke General Health score, and baseline BMI—did not alter these results.

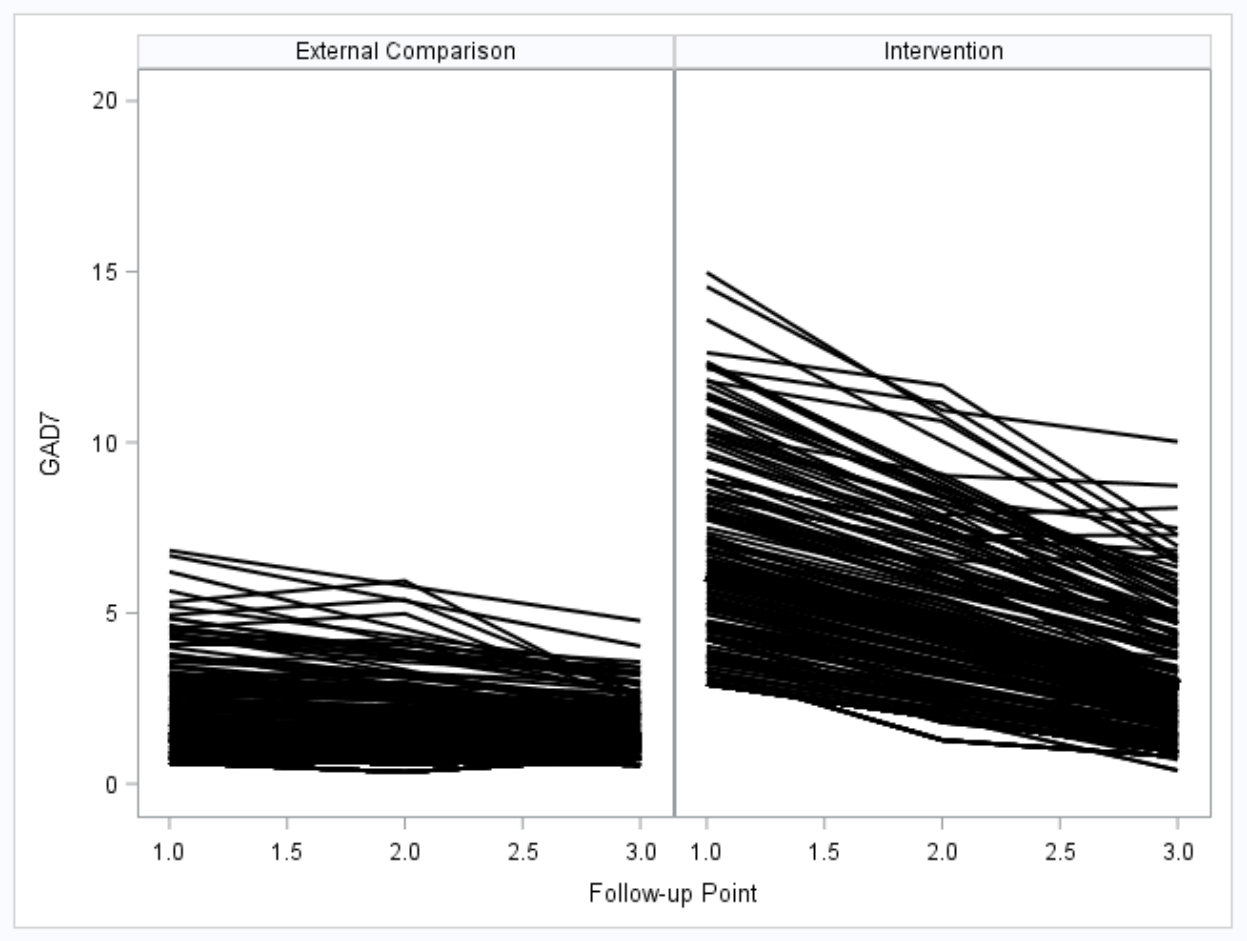
Table 57. Effect of IBH Intervention on Trajectory of GAD-7 Score Across Twelve Month Study, Full Mercy Sample (Secondary Comparison Group)

Variable	Duke General Health (n=460)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	**-3.09**	**0.37**	**<0.001**
Time*Secondary Comparison (ref)	--	--	--
Time	-0.29	0.22	0.19
Intervention	4.68	0.33	<0.001
Secondary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

To visualize the longitudinal effect of the intervention on GAD-7 score, a two-panel spaghetti plot was produced using PROC SGPANEL. **Figure 11** displays the secondary comparison group trajectory in the left panel and the intervention group trajectory in the right panel. The x-axis of the graph shows the study follow-up points with 1.0 representing baseline, 2.0 is the 6-month point, and 3.0 is the 12-month end-point. The trajectory figure visually displays the differences identified in the longitudinal statistical model, identifying a higher mean baseline GAD-7 score at baseline and greater variability among the intervention group. There is not much decrease in the GAD-7 scores of the secondary comparison group with many of the trajectories being quite flat. The decrease seen in the trajectories from baseline to 12 months is clearly steeper for those in the intervention than in the secondary comparison group indicating a greater decrease over time for the intervention participants.

Figure 11. Individual Trajectories of GAD-7 Score Across Twelve Month Study Period for IBH Intervention and Secondary External Comparison Groups



Limitations

No limitations specific to this measure to note.

Blood Pressure

Question 6. **Do patients who participate in the Sí Three intervention experience improvements in blood pressure, when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on blood pressure, data on systolic and diastolic blood pressure were recorded and analyzed separately. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for systolic or diastolic blood pressure. The sample sizes for the presented analyses of systolic and diastolic blood pressure are as follows: bivariate analyses (n=399), primary linear regression analyses (n=394), and longitudinal analyses (n=460).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 65** presents the mean systolic and diastolic blood pressure data in each study period for the overall sample as well as the intervention and secondary comparison groups. The overall sample had a mean blood pressure of 128.8/79.0 mmHg at baseline. For those who returned for a follow-up assessment, this decreased to 126.7/77.4 mmHg at 6-month follow-up and decreased again at 12-month follow-up (125.8/77.4 mmHg). The intervention group began the study with a lower mean blood pressure, 125.3/74.9 mmHg at baseline while the secondary comparison group had a higher mean blood pressure of 130.8/81.3 mmHg at baseline. In the intervention group, for those who returned for a follow-up assessment, the mean blood pressure decreased at 6 months to 121.7/73.2 mmHg and increased again at 12-month follow-up (124.6/74.4 mmHg). In the secondary comparison group, the 6-month mean blood pressure decreased to 122.2/72.6 mmHg and then increased to 124.0/74.0 mmHg at the 12-month follow-up. As previously noted, these differences between the two groups were statistically significant (**Table 11**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The decrease from baseline to 12-month follow-up for systolic blood pressure in the intervention group was statistically significant, but the decrease in the secondary comparison group was not statistically significant. The changes in diastolic blood pressure were not statistically significant in either the intervention or the secondary comparison.

Bivariate analyses also were performed between the intervention and secondary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for systolic blood pressure, when comparing the intervention and secondary comparison group at 12 months and without controlling for any additional covariates, the null hypothesis cannot be rejected. The difference in mean systolic blood pressure measure at 12-months is not significantly different between the two groups when not adjusting for any additional covariates. Based on a p-value less than 0.05 for diastolic blood pressure, when comparing the intervention and secondary comparison group at 12 months and without controlling for any additional covariates, the null hypothesis can be rejected. The mean diastolic blood pressure measure is significantly different between the two study groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcomes of systolic and diastolic blood pressure. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for both systolic and diastolic blood pressure were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline systolic blood pressure, baseline diastolic blood pressure, and the number of comorbidities at baseline. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline PHQ-9 score was included for possible selection. These models were specified as follows:

$$Y_{(SBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_SBP + \beta_{10} BL\_DBP + \beta_{11} BL\_Comorbidities + \beta_{12} BL\_PHQ\text{-}9 + \varepsilon$$

$$Y_{(DBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_DBP + \beta_{10} BL\_SBP + \beta_{11} BL\_Comorbidities + \beta_{12} BL\_PHQ\text{-}9 + \varepsilon$$

Two variations of each model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, a multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of systolic blood pressure included those covariates with p-value of 0.15 or less: sex, marital status, alcohol consumption, and baseline systolic blood pressure. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(SBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Married + \beta_5 Alcohol + \beta_6 BL\_SBP + \varepsilon$$

The final model of diastolic blood pressure included those covariates with p-value of 0.15 or less: marital status and baseline diastolic blood pressure. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(DBP)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 MaritalStatus + \beta_5 BL\_DBP + \varepsilon$$

Findings

Estimates for the final models of systolic and diastolic blood pressure are presented in **Table 58**.

Mean systolic blood pressure at 12 months did not differ significantly between the intervention and secondary comparison group (p=0.75). Below is the selected model with each covariate's effect estimate included:

$$Y_{(SBP)} = 59.41 + 0.46(Intervention) + 0.15(Age) + 5.97(Male) + -2.50(Married) + 3.10(No\ Alcohol\ Use) + 0.42(BL\_SBP) + \varepsilon$$

On average, for participants in the intervention group, there is a 2.99 mmHg decrease in diastolic blood pressure at 12 months holding all other variables in the model constant compared to those in the secondary comparison group. This result is statistically significant with a p value of 0.001; the effect size (using Cohen's d) is 0.33. Below is the selected model with each covariate's effect estimate included:

$$Y_{(DBP)} = 55.48 + -2.99(Intervention) + -0.04(Age) + -0.10(Sex) + -1.62(Married) + 0.33(BL\_DBP) + \varepsilon$$

Table 58. Effect of IBH Intervention on Twelve Month Systolic and Diastolic Blood Pressure Value, Full Mercy Sample (Secondary Comparison Group)

Variable	Systolic Blood Pressure (n=394)		
	Estimate (β)	Standard Error	p-value
Intervention	0.46	1.44	0.75
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	0.15	0.06	0.02
Male	5.97	1.77	0.001
Female (ref)	--	--	--
Married	-2.50	1.40	0.07
Unmarried (ref)	--	--	--
No Alcohol Use	3.10	1.81	0.09
Alcohol Use (ref)	--	--	--
BL_SBP	0.42	0.04	<0.001
Variable Selected	Diastolic Blood Pressure (n=394)		
	Estimate (β)	Standard Error	p-value
Intervention	**-2.99**	**0.93**	**0.001**
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.04	0.04	0.29
Male[a]	-0.10	1.06	0.93
Female (ref)	--	--	--
Married	-1.62	0.89	0.07
Unmarried (ref)	--	--	--
BL_DBP	0.33	0.04	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for systolic of diastolic blood pressure (not shown). The models estimated included interaction terms between intervention group and baseline hypertension.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. No significant time by intervention group interaction was identified for systolic blood pressure (p=0.27) or diastolic blood pressure (p=0.34) over the 12-month study, indicating that the trajectories from baseline to 6 months, and then to 12 months did not differ between the two study arms for systolic or diastolic blood pressure (see **Table 59**). Adjusting for the covariates that were selected in the primary model—age, sex, marital status, and alcohol consumption —did not alter these results.

Table 59. Effect of IBH Intervention on Trajectory of Systolic and Diastolic Blood Pressure Value Across Twelve Month Study, Full Mercy Sample (Secondary Comparison Group)

Variable	Systolic Blood Pressure (n=460)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	1.87	1.68	0.27
Time*Secondary Comparison (ref)	--	--	--
Time	-4.17	1.01	<0.001
Intervention	-6.82	1.68	<0.001
Secondary Comparison (ref)	--	--	--
Variable	Diastolic Blood Pressure (n=460)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	1.14	1.04	0.28
Time*Secondary Comparison (ref)	--	--	--
Time	-2.22	0.63	<0.001
Intervention	-6.81	0.85	<0.001
Secondary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

There are no limitations to report on this measure.

HbA1c Level

Question 7. **Do patients with a history or diagnosis of diabetes who participate in the Sí Three intervention experience improvements in HbA1c measures after 12 months when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on diabetes management among known or suspected diabetic patients, data were collected on patient HbA1c levels. As previously stated, it is Mercy's clinical practice to only recommend HbA1C test and subsequently collect among patients who are: (1) known/self-reported to be diabetic, (2) have an elevated blood glucose at time of clinic visit or are suspected to be diabetic through other signs and symptoms. Therefore, the sample size is reduced for these analyses compared to other impact measure analyses. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for HbA1c level. The sample sizes for the presented analyses of HbA1c level are as follows: bivariate analyses (n=342), primary linear regression analyses (n=327), and longitudinal analyses (n=376).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 65** presents the mean HbA1c level data in each study period for the overall sample as well as the intervention and secondary comparison groups. The overall study sample

had a mean HbA1c of 7.3% at baseline. For those who returned for a follow-up assessment, this decreased to 7.1% at 6-month follow-up and increased again at 12-month follow-up (7.2%). The intervention group began the study with a slightly higher mean HbA1c of 7.4% at baseline while the secondary comparison group HbA1c at baseline was 7.3%. For participants who returned for a follow-up visit, the intervention group mean HbA1c decreased at 6-month follow-up to 7.0% and remained the same at 12 months. For those participants in the secondary comparison group who returned for a follow-up visit, the mean HbA1c decreased at 6 months to 7.2% and increased to 7.3% at 12 months. As previously noted, this difference between the two groups was not statistically significant (**Table 11**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). The observed decrease in mean HbA1c from baseline to 12-month follow-up within both the intervention and secondary comparison groups was statistically significant.

Bivariate analyses were also performed between the intervention and secondary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for HbA1c when comparing the intervention and secondary comparison group at 12 months, the null hypothesis cannot be rejected. The mean HbA1c measure was not significantly different between the two groups when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome, HbA1c level. Age and sex were selected a priori for inclusion due to the known biological influence of age and sex on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for HbA1c level were: age, sex, primary language, marital status, smoking, alcohol consumption, employment, baseline HbA1c level, and the number of comorbidities at baseline. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline PHQ-9 score was included for possible selection. This model was specified as follows:

$$Y_{(HbA1c)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 Language + \beta_5 MaritalStatus + \beta_6 Smoke + \beta_7 Alcohol + \beta_8 Employment + \beta_9 BL\_HbA1c + \beta_{10} BL\_Comorbidities + \beta_{11} BL\_PHQ\text{-}9 + \varepsilon$$

Two variations of the model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of HbA1c level included those covariates with p-value of 0.15 or less: baseline HbA1c level and number of comorbidities at baseline. Age and sex were maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(HbA1c)} = \beta_0 + \beta_1 StudyArm + \beta_2 Age + \beta_3 Sex + \beta_4 BL\_HbA1c + \beta_5 BL\_Comorbidities + \varepsilon$$

Findings

Estimates for the final model of HbA1c level are presented in **Table 60**.

On average, for participants in the intervention group, there is a 0.51 decrease in HbA1c at 12 months holding all other variables in the model constant compared to those in the secondary comparison group. This result is statistically significant with a p value of 0.01; the effect size (using Cohen's d) is 0.27. Below is the selected model with each covariate's effect estimate included:

$$Y_{(HbA1c)} = 1.68 + -0.51(\text{Intervention}) + - <0.001(\text{Age}) + 0.06(\text{Male}) + 0.73(\text{BL\_HbA1c}) + 0.16(\text{BL\_Comorbidities}) + \varepsilon$$

Table 60. Effect of IBH Intervention on Twelve Month HbA1c Level, Full Mercy Sample (Secondary Comparison Group)

Variable	HbA1c (n=327)		
	Estimate (β)	Standard Error	P-value
Intervention	**-0.51**	**0.21**	**0.01**
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	- <0.001	0.01	0.99
Male[a]	0.06	0.18	0.75
Female (ref)	--	--	--
BL_HbA1c	0.73	0.04	<0.001
Number of comorbidities at baseline	0.16	0.08	0.04

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for HbA1c level (not shown). The models estimated included interaction terms between intervention group and baseline diabetes.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. The time by intervention group interaction term was not statistically significant for HbA1c (p=0.07), indicating that the HbA1c level trajectories from baseline to 6 months, and then to 12 months were not different between the two study arms (see **Table 61**). Adjusting for the covariates that were selected in the primary model—age, sex, and number of comorbidities at baseline —did not alter the estimates.

Table 61. Effect of IBH Intervention on Trajectory of HbA1c Level Across Twelve Month Study, Full Mercy Sample (Secondary Comparison Group)

Variable	HbA1c (n=376)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	-0.32	0.18	0.07
Time*Secondary Comparison (ref)	--	--	--
Time	-0.01	0.09	0.89
Intervention	-0.04	0.21	0.86
Secondary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

Limitations

As noted in the SEP, the limited availability of HbA1c data could result in an insufficient sample size to detect a statistical difference. Given a significant result was detected, this is not a concern for the linear regression, but could have led to a weaker result and also may have affected the longitudinal analysis results.

Waist Circumference

Question 8. **Do patients who participate in the Sí Three intervention experience improvements in waist circumference after 12 months when compared to patients that do not participate in the intervention? This question is exploratory.**

Overview of Analysis

To answer this exploratory question about intervention impact on waist circumference, data were collected on waist circumference, and analyses were conducted separately for males and females. While systematic checks for outliers were performed and questions sent to study site staff for verification on a quarterly basis, there were no unique data cleaning processes needed for waist circumference. The sample sizes for the presented analyses of waist circumference are as follows: bivariate analyses (n=395: 312 female, 83 male), primary linear regression analyses (n=377: 295 female, 82 male), and longitudinal analyses (n=444: 350 female, 94 male).

Descriptive Statistics and Bivariate Comparisons

At the end of this section, **Table 65** presents the mean waist circumference data by sex in each study period for the overall sample as well as the intervention and secondary comparison groups.

For males, the overall sample had a mean waist circumference of 40.8 inches at baseline. This increased to 41.7 for participants who returned at 6-month follow-up and increased slightly to 41.8 inches for those who returned at 12-month follow-up. Males in the intervention group began the study with a higher mean waist circumference of 41.5 inches at baseline while the secondary comparison group had a lower mean waist circumference of 40.6 inches at baseline. For male participants who completed a follow-up assessment, the intervention group mean waist circumference increased at both 6 and 12-month follow-up to 41.9 inches and 42.1 respectively. Male participants in the secondary comparison

group followed the same trend with the mean waist circumference for those who completed a follow-up increasing overtime to 41.6 inches at 6 months 41.7 inches at 12 months. As previously noted, this difference between the two groups was not statistically significant (**Table 11**).

For females, the overall sample had a mean waist circumference of 41.5 inches at baseline. This increased to 41.8 inches for participants who returned at 6-month follow-up and decreased to 41.5 inches for those who returned at 12-month follow-up. Females in the intervention group began the study with a higher mean waist circumference of 43.7 inches at baseline while the secondary comparison group had a lower mean waist circumference of 40.1 inches at baseline. For female participants who completed a follow-up assessment, the intervention group mean waist circumference decreased at both 6 and 12-month follow-up to 43.5 inches and 42.6 respectively. Female participants in the secondary comparison group followed a different trend with the mean waist circumference for those who completed a follow-up increasing overtime to 40.6 inches at 6 months and 40.8 inches at 12 months. As previously noted, this difference between the two groups was statistically significant (**Table 11**).

Bivariate analyses were performed within each study group, testing the statistical significance of the change in impact measures from baseline to 12-month follow-up without controlling for additional covariates (**Table 66**). For male participants, the increases from baseline to 12-month follow-up within both the intervention and secondary comparison groups for waist circumference were not statistically significant. For female participants, the decreases from baseline to 12-month follow-up within both the intervention and secondary comparison groups for waist circumference were statistically significant.

Bivariate analyses also were performed between the intervention and secondary comparison groups comparing mean impact measures at 12-month follow-up (

Table 67). Based on a p-value greater than 0.05 for waist circumference when comparing the intervention and secondary comparison group at 12 months, the null hypotheses cannot be rejected for male waist circumference. The mean waist circumference for males was not significantly different between the two groups when not adjusting for any additional covariates. Based on a p-value less than 0.05 for waist circumference when comparing the intervention and secondary comparison group at 12 months, the null hypotheses for female waist circumference can be rejected. The mean waist circumference was significantly different between the two groups, for females, when not adjusting for any additional covariates.

Model Selection Process

A backward elimination model selection approach was used to identify a parsimonious model with covariates that contributed to the outcome of waist circumference. Models were stratified by sex, and age was selected a priori for inclusion due to the known biological influence of age on health outcomes. Other covariates were removed from the model if their p-value was found to be greater than 0.15. The initial covariates that were input into the models for waist circumference were: age, primary language, marital status, smoking, alcohol consumption, employment, baseline waist circumference, and the number of comorbidities at baseline. The inclusion of baseline waist circumference for females controlled for the statistical imbalance between intervention and secondary comparison groups at baseline. Additionally, to further understand whether and how physical and mental health are associated in this study population, baseline PHQ-9 score was included for possible selection. These models were specified as follows:

$$Y_{(\text{Waist Male})} = \beta_0 + \beta_1\text{StudyArm} + \beta_2\text{Age} + \beta_3\text{Language} + \beta_4\text{MaritalStatus} + \beta_5\text{Smoke} + \beta_6\text{Alcohol} + \beta_7\text{Employment} + \beta_8\text{BL\_Waist} + \beta_9\text{BL\_Comorbidities} + \beta_{10}\text{BL\_PHQ-9} + \varepsilon$$

$$Y_{(\text{Waist Female})} = \beta_0 + \beta_1\text{StudyArm} + \beta_2\text{Age} + \beta_3\text{Language} + \beta_4\text{MaritalStatus} + \beta_5\text{Smoke} + \beta_6\text{Alcohol} + \beta_7\text{Employment} + \beta_8\text{BL\_Waist} + \beta_9\text{BL\_Comorbidities} + \beta_{10}\text{BL\_PHQ-9} + \varepsilon$$

Two variations of each model were run to assess the best fit model: one including age as a continuous predictor and another utilizing categorical age. As previously stated, multiple imputation approach was considered but not performed due to the near completeness of the evaluated data, thus the analysis was completed on those study participants who were not missing data on any of the included variables.

The final model of waist circumference in males included those covariates with p-value of 0.15 or less: baseline waist circumference. Age was maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(\text{Waist Male})} = \beta_0 + \beta_1\text{StudyArm} + \beta_2\text{Age} + \beta_3\text{BL\_waist} + \varepsilon$$

The final model of waist circumference in females included those covariates with p-value of 0.15 or less: smoking and baseline waist circumference. Age was maintained based on a priori selection. Age was modelled as a continuous variable for parsimony based on similar adjusted R-square results across the two models. The final model specification was:

$$Y_{(\text{Waist Female})} = \beta_0 + \beta_1\text{StudyArm} + \beta_2\text{Age} + \beta_3\text{Smoking} + \beta_4\text{BL\_waist} + \varepsilon$$

Findings

Estimates for the final model of waist circumference, for males and females, are presented in **Table 62**.

Mean waist circumference for males at 12 months did not differ significantly between the intervention and secondary comparison group (p=0.79). Below is the selected model with each covariate's effect estimate included:

$$Y_{(\text{Waist Male})} = 1.34 + 0.23(\text{Intervention}) + -0.03(\text{Age}) + 1.01(\text{BL\_waist}) + \varepsilon$$

On average, for participants in the intervention group, there is a 2.31-inch decrease in waist circumference for females at 12 months holding all other variables in the model constant compared to those in the secondary comparison group. This result is statistically significant with a p value of <0.001; the effect size (using Cohen's d) is 0.37. Below is the selected model with each covariate's effect estimate included:

$$Y_{(\text{Waist Female})} = 1.10 + -2.31(\text{Intervention}) + 0.004(\text{Age}) + -1.69(\text{Current Smoker}) + 2.23(\text{Former Smoker}) + 0.98(\text{BL\_waist}) + \varepsilon$$

Table 62. Effect of IBH Intervention on Twelve Month Waist Circumference, By Sex, Full Mercy Sample (Secondary Comparison Group)

Variable	Male Waist Circumference (n=82)		
	Estimate (β)	Standard Error	p-value
Intervention	0.23	0.86	0.79
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	-0.03	0.03	0.45
BL_Waist	1.01	0.05	<0.001
Variable Selected	Female Waist Circumference (n=295)		
	Estimate (β)	Standard Error	p-value
Intervention	**-2.31**	**0.43**	**<0.001**
Secondary Comparison (ref)	--	--	--
Age (continuous)[a]	0.004	0.02	0.81
Current Smoker	-1.69	1.01	0.10
Former Smoker	2.23	0.93	0.02
Never Smoker (ref)	--	--	--
BL_Waist	0.98	0.03	<0.001

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05). [a] Included in the model a priori despite not having met the stepwise inclusion criteria.

There were no statistically significant effect modifications for waist circumference among males or females (not shown). The model considered included an interaction term of study group and baseline obesity.

Additional Analyses

Longitudinal analyses were conducted to examine time as an independent variable, assessing whether the outcome trajectories differ by intervention status. To estimate the linear mixed model, the PROC MIXED procedure in SAS was used. No significant time by intervention group interaction for waist circumference among males was identified (p=0.50;
Table 63), indicating that the trajectories from baseline to 6 months, and then to 12 months were not different between the two study arms for waist circumference among males.

There was a significant time by intervention group interaction for waist circumference among females (p<0.001), indicating that the trajectories from baseline to 6 months, and then to 12 months were different between the two study arms for waist circumference among females (see
Table 63). For each 6 months in the study, the intervention group experienced a relative 2.15-inch decrease in waist circumference compared to the secondary comparison group. Adjusting for the covariates that were selected in the primary model—age and smoking —did not alter these results (not shown).

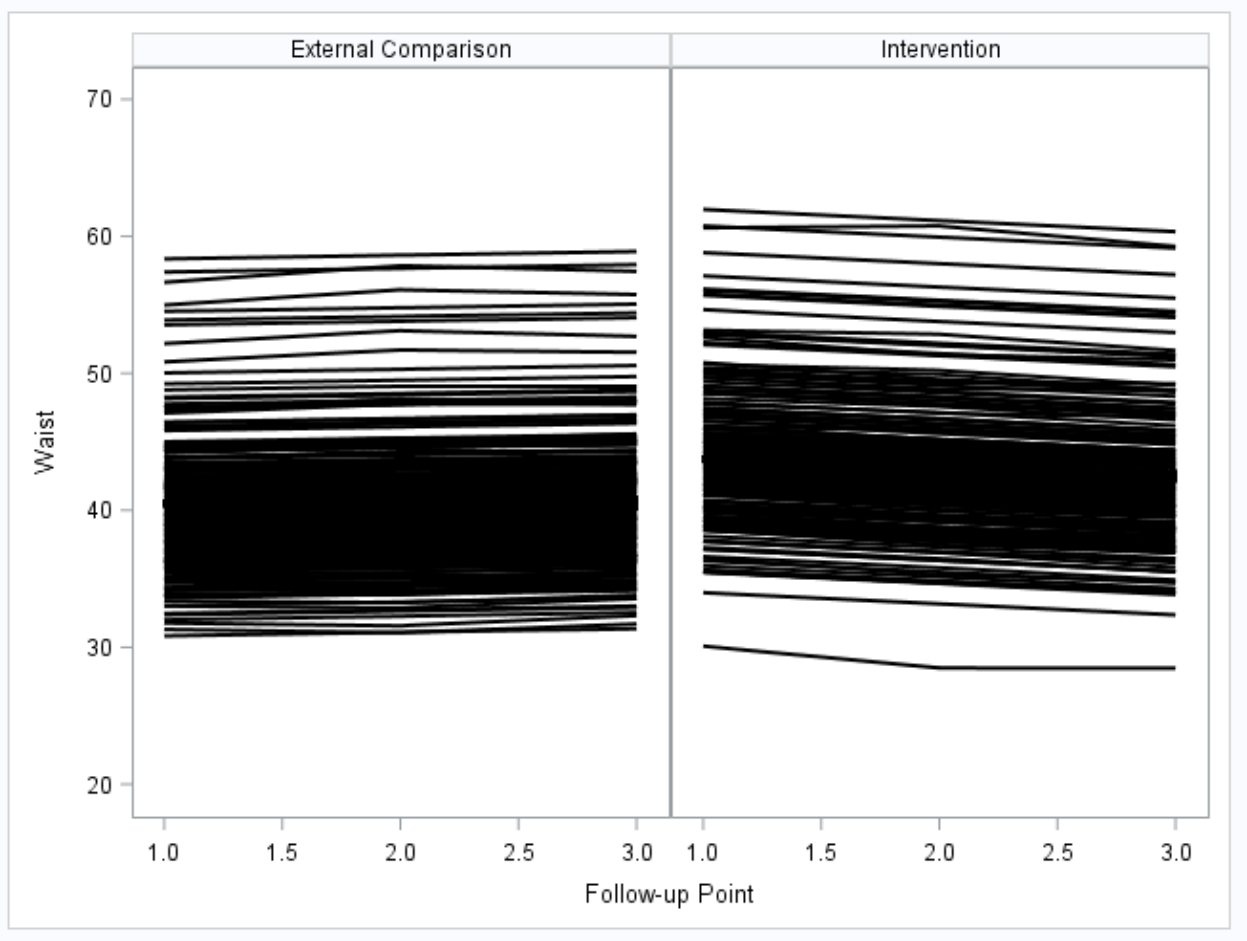
Table 63. Effect of IBH Intervention on Trajectory of Waist Circumference, By Sex, Full Mercy Sample (Secondary Comparison Group)

Variable	Male Waist Circumference (n=94)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	0.52	0.76	0.50
Time*Secondary Comparison (ref)	--	--	--
Time	0.58	0.36	0.11
Intervention	0.85	1.46	0.56
Secondary Comparison (ref)	--	--	--
Variable	Female Waist Circumference (n=350)		
	Estimate (β)	Standard Error	p-value
Time*Intervention	**-2.15**	**0.37**	**<0.001**
Time*Secondary Comparison (ref)	--	--	--
Time	0.53	0.23	0.02
Intervention	3.84	0.56	<0.001
Secondary Comparison (ref)	--	--	--

Notes: "ref" indicates the reference category used to calculate the estimate for a covariate. Bold denotes statistically significant difference between intervention and comparison groups (p-value<0.05).

To visualize the longitudinal effect of the intervention on female waist circumference score, a two-panel spaghetti plot was produced using PROC SGPANEL.. **Figure 12** displays the secondary comparison group trajectory in the left panel and the intervention group trajectory in the right panel. The x-axis of the graph shows the study follow-up points with 1.0 representing baseline, 2.0 is the 6-month point, and 3.0 is the 12-month end-point. The trajectory figure visually displays the differences identified in the longitudinal statistical model, whereby the female waist circumference trajectory of the secondary comparison group is increasing whereas the intervention group has a decreasing trend.

Figure 12. Individual Trajectories of Female Waist Circumference Across Twelve Month Study Period for IBH Intervention and Secondary External Comparison Groups



Limitations

A limitation to consider for this measure is the smaller sample size for males. Because of the smaller sample size, there may not have been sufficient power to detect a statistically significant difference. Also, compared to the other impact measures, there was a slightly greater amount of missing data on this measure. Multiple imputation approaches were considered but assessed to not be necessary as the missing data were not substantial.

Table 64. Health Impact Measures by Study Arm and Follow-up Period for Intervention and Primary Comparison Group

	Full Sample			Intervention Group			Primary Comparison Group		
	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo
	n=410	n= 312	n=293	n=207	n= 169	n=142	n=203	n= 143	n=151
Measure		Mean (SD)			Mean (SD)			Mean (SD)	
Blood pressure									
Systolic	124.6 (17.5)	121.9 (15.5)	124.0 (14.6)	125.3 (18.4)	121.7 (16.6)	124.7 (14.0)	123.9 (16.5)	122.2 (14.2)	123.3 (15.3)
Diastolic	74.2 (9.8)	72.9 (9.1)	74.2 (9.4)	74.9 (10.1)	73.2 (9.1)	74.4 (9.7)	73.5 (9.5)	72.6 (9.1)	74.0 (9.0)
Missing	1	7	2	--	7	--	1	--	2
HbA1c	n=190	n=172	n=165	n=99	n=95	n=85	n=91	n=77	n=80
HbA1c	7.3 (1.9)	6.9 (1.4)	7.0 (1.4)	7.4 (2.0)	7.0 (1.4)	7.0 (1.5)	7.1 (1.6)	6.8 (1.3)	6.9 (1.4)
Missing	--	--	--	--	--	--	--	--	--
BMI									
BMI	32.9 (6.6)	33.0 (6.9)	33.3 (6.5)	33.2 (7.1)	33.3 (7.5)	34.0 (7.4)	32.5 (6.0)	32.7 (6.2)	32.7 (5.5)
Missing	--	9	6	--	6	3	--	3	3
Waist Circumference									
Males	42.2 (5.1)	42.2 (4.7)	42.3 (4.4)	41.5 (4.2)	41.9 (4.9)	42.1 (4.5)	43.0 (6.0)	42.4 (4.6)	42.4 (4.3)
Females	43.6 (5.6)	43.3 (5.9)	42.2 (6.6)	43.7 (6.0)	43.5 (5.9)	42.6 (7.4)	43.5 (5.2)	43.0 (5.8)	41.8 (5.6)
Missing	24	18	10	21	12	4	3	6	6
CAGE-AID									
CAGE-AID Score	0.2 (0.6)	0.1 (0.5)	0.1 (0.6)	0.2 (0.6)	0.1 (0.5)	0.1 (0.5)	0.2 (0.6)	0.1 (0.6)	0.1 (0.6)
Missing	--	23	3	--	5	2	--	18	1
PHQ-9									
PHQ-9 Score	5.5 (5.5)	4.2 (5.1)	2.7 (4.1)	6.7 (6.0)	4.9 (5.5)	2.9 (4.6)	4.4 (4.6)	3.3 (4.5)	2.5 (3.6)
Missing	--	23	3	--	5	2	--	18	1
GAD-7									
GAD-7 Score	5.2 (5.3)	3.8 (4.8)	2.5 (3.6)	6.2 (5.6)	4.6 (5.3)	2.6 (3.7)	4.3 (4.7)	2.7 (3.6)	2.4 (3.5)
Missing	--	23	3	--	5	2	--	18	1
DUKE Health									
General Health	71.1 (17.2)	75.4 (17.2)	79.3 (16.8)	67.7 (17.5)	72.9 (17.8)	78.5 (16.9)	74.5 (16.2)	78.6 (15.9)	80.0 (16.6)
Mental Health	76.0 (21.0)	80.7 (21.1)	84.5 (20.0)	72.1 (21.9)	77.4 (21.7)	83.1 (20.8)	80.0 (19.3)	85.1 (19.5)	85.7 (19.2)

	Full Sample			Intervention Group			Primary Comparison Group		
	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo
	n=410	n= 312	n=293	n=207	n= 169	n=142	n=203	n= 143	n=151
Measure		Mean (SD)			Mean (SD)			Mean (SD)	
Social Health	70.7 (19.0)	74.0 (18.9)	78.4 (18.2)	68.1 (19.7)	72.6 (19.6)	78.0 (18.3)	73.4 (17.9)	75.9 (17.8)	78.8 (18.2)
Physical Health	66.6 (24.3)	71.4 (24.0)	74.7 (23.7)	62.9 (24.1)	68.8 (24.8)	74.6 (24.6)	70.3 (24.0)	74.8 (22.7)	74.9 (22.9)
Missing	--	23	3	--	5	2	--	18	1

Table 65. Health Impact Measures by Study Arm and Follow-up Period for Intervention and Secondary Comparison Group

	Full Sample			Intervention			Secondary Comparison		
	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo
	n=573	n=420	n=399	n=207	N=169	n=142	n=366	n=251	n=257
Measure		Mean (SD)			Mean (SD)			Mean (SD)	
Blood pressure									
Systolic	128.8 (18.9)	126.7 (17.5)	125.8 (16.1)	125.3 (18.4)	121.7 (16.6)	124.7 (14.0)	130.8 (18.9)	129.9 (17.3)	126.5 (17.2)
Diastolic	79.0 (10.6)	77.4 (9.8)	77.4 (9.4)	74.9 (10.1)	73.2 (9.1)	74.4 (9.7)	81.3 (10.2)	80.1 (9.4)	79.1 (8.8)
Missing	--	7	--	--	7	--	--	--	--
HbA1c	n=465	n=346	n=342	n=99	n=95	n=85	n=366	n=251	n=257
HbA1c	7.3 (2.0)	7.1 (1.8)	7.2 (1.9)	7.4 (2.0)	7.0 (1.4)	7.0 (1.5)	7.3 (2.0)	7.2 (1.9)	7.3 (2.0)
Missing	--	--	--	--	--	--	--	--	--
BMI									
BMI	34.6 (7.5)	34.7 (7.4)	34.9 (7.5)	33.2 (7.1)	33.3 (7.5)	34.0 (7.4)	35.4 (7.6)	35.6 (7.3)	35.3 (7.6)
Missing	--	6	3	--	6	3	--	--	--
Waist Circumference									
Males	40.8 (6.2)	41.7 (7.0)	41.8 (7.1)	41.5 (4.2)	41.9 (4.9)	42.1 (4.5)	40.6 (6.6)	41.6 (7.5)	41.7 (7.8)
Females	41.5 (5.8)	41.8 (5.6)	41.5 (6.4)	43.7 (6.0)	43.5 (5.9)	42.6 (7.4)	40.1 (5.4)	40.6 (5.1)	40.8 (5.5)
Missing	21	12	4	21	12	4	--	--	--
PHQ-9									
PHQ-9 Score	3.8 (4.7)	3.1 (4.4)	2.0 (3.4)	6.7 (6.0)	4.9 (5.5)	2.9 (4.6)	2.1 (2.5)	1.9 (2.9)	1.5 (2.4)
Missing	--	5	2	--	5	2	--	--	--
GAD-7									

Sí Texas Subgrantee: Mercy Ministries of Laredo
Program Title: Sí Three: Integration of 3-D Health Services

	Full Sample			Intervention			Secondary Comparison		
	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo	Baseline	6-Mo	12-Mo
	n=573	n=420	n=399	n=207	N=169	n=142	n=366	n=251	n=257
Measure		Mean (SD)			Mean (SD)			Mean (SD)	
GAD-7 Score	3.2 (4.4)	2.7 (4.3)	1.7 (2.9)	6.2 (5.6)	4.6 (5.3)	2.6 (3.7)	1.5 (2.2)	1.5 (2.8)	1.2 (2.1)
Missing	--	5	2	--	5	2	--	--	--
General Health	76.6 (17.5)	79.7 (17.4)	83.3 (15.6)	67.7 (17.5)	72.9 (17.8)	78.5 (16.9)	81.6 (15.4)	84.1 (15.6)	85.9 (14.2)
Mental Health	79.1 (20.9)	81.9 (21.1)	85.7 (18.6)	72.1 (21.9)	77.4 (21.7)	83.1 (20.8)	83.0 (19.2)	84.8 (20.3)	87.1 (17.1)
Physical Health	69.5 (23.7)	73.7 (23.8)	76.9 (23.5)	62.9 (24.1)	68.8 (24.8)	74.6 (24.6)	73.1 (22.8)	76.9 (22.6)	78.1 (22.9)
Social Health	81.2 (20.9)	83.6 (18.9)	87.4 (16.8)	68.1 (19.7)	72.6 (19.6)	78.0 (18.3)	88.6 (17.6)	90.8 (14.5)	92.2 (13.6)
Missing	--	5	2	--	5	2	--	--	--

Table 66. Within Group Bivariate Analyses Comparing Baseline to 12 Months

	12-Month Mean (SD)	Baseline Mean (SD)	12-month (–) Baseline Mean Difference (SD)	p-value
INTERVENTION GROUP (n=142)				
BMI[a]	33.9 (7.5)	33.7 (7.2)	0.2 (1.6)	0.42
BP – Systolic	**124.6 (14.0)**	**128.2 (19.8)**	**-3.6 (17.5)**	**0.02**
BP – Diastolic	74.7 (10.1)	76.1 (10.5)	-1.4 (10.9)	0.07
Waist Circumference – Males	42.1 (4.5)	41.0 (3.9)	1.1 (2.5)	0.08
Waist Circumference – Females	**42.6 (7.4)**	**44.2 (6.1)**	**-1.8 (4.4)**	**<0.001**
Nonparametric Tests[a]	12-Month Median (SD)	Baseline Median (SD)		p-value
PHQ-9	**1.0 (4.6)**	**6.0 (6.0)**		**<0.001**
General Health	**83.3 (16.8)**	**66.7 (18.2)**		**<0.001**
GAD-7	**1.0 (3.7)**	**5.0 (5.6)**		**<0.001**
HbA1c[c]	**6.4 (1.5)**	**6.7 (1.9)**		**<0.001**
CAGE-AID	0 (0.5)	0 (0.6)		0.20
PRIMARY COMPARISON GROUP (n=151)				
BMI[a]	32.7 (5.5)	32.5 (5.7)	0.2 (1.5)	0.16
BP – Systolic	123.4 (15.3)	124.2 (16.1)	-0.8 (15.3)	0.52
BP – Diastolic	74.1 (9.0)	73.8 (9.5)	0.4 (9.7)	0.64
Waist Circumference – Males	42.4 (4.3)	41.9 (3.7)	0.6 (1.6)	0.12
Waist Circumference – Females	**41.8 (5.6)**	**43.7 (5.2)**	**-1.9 (4.2)**	**<0.001**
Nonparametric Tests[a]	12-Month Median (SD)	Baseline Median (SD)		p-value
PHQ-9	**1.0 (3.6)**	**3.0 (4.2)**		**<0.001**
General Health	**83.3 (16.7)**	**76.7 (16.5)**		**<0.001**
GAD-7	**1.0 (3.6)**	**2.0 (4.4)**		**<0.001**
HbA1c[c]	**6.5 (1.4)**	**6.7 (1.5)**		**<0.001**
CAGE-AID	0 (0.6)	0 (0.6)		0.38
SECONDARY COMPARISON GROUP (n=257)				
BMI[a]	35.5 (7.5)	35.3 (7.5)	-0.01 (0.06)	0.09
BP – Systolic	**131.3 (18.6)**	**126.5 (17.5)**	**-4.7 (16.8)**	**<0.001**
BP – Diastolic	**81.5 (10.2)**	**79.2 (8.9)**	**-2.3 (10.6)**	**<0.001**
Waist Circumference – Males	41.0 (6.9)	41.5 (7.7)	0.6	0.15
Waist Circumference – Females	**40.3 (5.5)**	**40.8 (5.6)**	**0.48**	**0.01**
Nonparametric Tests[a]	12-Month Median (SD)	Baseline Median (SD)		p-value
PHQ-9	**1.0 (2.6)**	**0.0 (2.4)**		**<0.001**
General Health	**86.7 (16.0)**	**90.0 (14.2)**		**<0.001**
GAD-7	0.0 (2.3)	0.0 (2.1)		0.10
HbA1c	7.0 (1.9)	6.7 (2.0)		0.92

Note: Bold denotes statistical significance (p-value < 0.05)

[a] The Wilcoxon Signed Rank test was used to examine non-normally distributed data

[b] A log transformation was used and then exponentiated

[c] The total sample for this measure was 85 participants in the intervention and 80 participants in the primary comparison

Table 67. Between Group Bivariate Analyses: Intervention vs. Primary Comparison Groups at 12 Months

	Full Sample n=293 Mean (SD)	Intervention Group n=142 Mean (SD)	Primary Comparison Group n=151 Mean (SD)	p-value
BMI[a]	33.3 (6.5)	33.8 (7.3)	32.7 (5.5)	0.23
BP – Systolic	124.0 (14.6)	124.6 (14.0)	123.4 (15.3)	0.47
BP – Diastolic	74.2 (9.4)	74.4 (9.7)	74.1 (9.0)	0.76
Waist Circumference – Males	42.3 (4.4)	42.1 (4.5)	42.4 (4.3)	0.82
Waist Circumference – Females	42.2 (6.6)	42.6 (7.4)	41.8 (5.6)	0.38
Nonparametric Tests[b]	Median (SD)	Median (SD)	Median (SD)	p-value
PHQ-9	1.0 (4.1)	1.0 (4.6)	1.0 (3.6)	0.85
General Health	83.3 (16.8)	83.3 (16.9)	83.3 (16.6)	0.52
GAD-7	1.0 (3.6)	1.0 (3.7)	1.0 (3.5)	0.81
HbA1c [c]	6.4 (1.4)	6.4 (1.5)	6.5 (1.4)	0.90
CAGE-AID	0.0 (0.6)	0.0 (0.5)	0.0 (0.6)	0.58

Note: Bold denotes statistical significance (p-value < 0.05)

[a] A log transformation was used and then exponentiated

[b] The Wilcoxon Signed Rank test was used to examine non-normally distributed data

[c] The total sample size for this measure was 165 participants, 85 for the intervention and 80 for the primary comparison groups

Table 68. Between Group Bivariate Analyses: Intervention Group vs. Secondary Comparison Group at 12 Months

	Full Sample n=399 Mean (SD)	Intervention Group n=142 Mean (SD)	Secondary Comparison Group n=257 Mean (SD)	p-value
BMI[a]	34.9 (7.5)	**33.8 (7.3)**	**35.4 (7.6)**	**0.03**
Systolic	125.8 (16.1)	124.6 (14.0)	126.5 (17.5)	0.26
Diastolic	**77.4 (9.4)**	**74.4 (9.7)**	**79.2 (8.9)**	**<0.001**
Waist Circumference – Males	41.8 (7.1)	42.1 (4.5)	41.5 (7.7)	0.76
Waist Circumference – Females	**41.5 (6.4)**	**42.6 (7.4)**	**40.8 (5.6)**	**0.02**
Nonparametric Tests[b]		Median (SD)	Median (SD)	p-value
PHQ-9	**0.0 (3.4)**	**1.0 (4.6)**	**0.0 (2.4)**	**0.001**
General Health	**86.7 (15.6)**	**83.3 (16.9)**	**90.0 (14.1)**	**<0.001**
GAD-7	**0.0 (2.9)**	**1.0 (3.7)**	**0.0 (2.1)**	**<0.001**
HbA1c	6.7 (1.9)	6.4 (1.5)	6.7 (2.1)	0.80

Note: Bold denotes statistical significance (p-value < 0.05)

[a] A log transformation was used and then exponentiated

[b] The Wilcoxon Signed Rank test was used to examine non-normally distributed data; these results aligned with t test results.

CONCLUSION – SUMMARY OF FINDINGS, LESSONS LEARNED, AND NEXT STEPS

This final report provides an overview of findings for the evaluation of the Mercy Ministries Sí Three program. Mercy implemented an IBH model in their Laredo, TX clinic to serve uninsured residents of Webb County. Mercy implemented a QED study to compare intervention participants receiving the delivery of IBH services with a primary comparison group within their own clinic who did not receive IBH services and a secondary comparison group who received usual clinic care at Nuestra Clinica del Valle Edcouch and Alton clinics. Study results indicate that the Sí Three program improved behavioral health among intervention participants. Consistent improvements were noted in behavioral health outcomes between the intervention participants and two comparison groups with intervention participants primarily receiving faith-based counseling services.

This evaluation study achieves a preliminary level of evidence. This evaluation study used a QED design which was designed to mitigate major threats to internal validity, particularly selection bias, through including primary and secondary comparison groups. The primary comparison group allowed for the examination of observed improvements in the intervention group as they relate to patients who receive services at the same clinic but chose not to receive the intervention (patients with similar demographics and disease characteristics in the same setting). A secondary comparison group allowed for the examination of observed improvements in the intervention group as they relate to patients who used a different clinic (factors related to being part of a different population in the same region). The study also meets the criteria for effective evidence because it demonstrates positive, significant findings for several exploratory outcomes. There were no negative intervention effects on confirmatory or exploratory outcomes. Both the Duke General Health score and GAD-7 exploratory outcomes achieved small effect sizes (Cohen's $d > 0.2$) for the primary analysis comparing intervention participants with the primary comparison group. Three exploratory outcomes achieved small effect sizes for the secondary analysis (intervention compared to secondary comparison group).

The study showed that, when controlling for baseline measures and other covariates, the intervention participants had significantly greater improvements when compared with the primary comparison group participants in the depression confirmatory outcome over time (reduced depression as measured through PHQ-9, $\beta=-1.76$, $p=0.001$) and additional outcomes identified in the logic model (increased Duke General Health score at 12 months $\beta = 4.01$, $p=0.02$, Cohen's 0.24; increased Duke Physical Health Score at 12 months $\beta=6.69$, $p=0.004$; increased Duke General Health Score over time $\beta=5.35$, $p=0.03$; decreased GAD-7 at 12 months $\beta=-0.79$, $p=0.03$, Cohen's d 0.22; and decreased GAD-7 over time $\beta=-1.58$, $p=0.002$).

Further, the study showed that, when controlling for baseline measures and other covariates, the intervention participants had significantly greater improvements when compared with the secondary comparison group participants in the depression confirmatory outcome over time (reduced depression as measured through PHQ-9, $\beta=-2.78$, $p=0.001$) and additional outcomes identified in the logic model (increased Duke General Health Score over time $\beta=5.96$, $p=0.001$; decreased GAD-7 over time $\beta=-3.05$, $p=0.001$; decreased diastolic blood pressure at 12 months $\beta=-2.99$, $p= 0.001$, Cohen's d 0.33; decreased HbA1c at 12 months $\beta=-0.51$, $p=0.01$, Cohen's d 0.27; decreased HbA1c over time $\beta=-0.35$, $p=0.05$; decreased female waist circumference at 12 months $\beta=-2.31$, $p=0.001$, 0.37; and decreased female waist circumference over time $\beta=-2.13$, $p<0.001$).

Mercy is a primary healthcare clinic located in Webb County, Texas, which provides healthcare and health education to some of the poorest neighborhoods and *colonias* in the U.S. Ninety-five percent of the population is Hispanic/Latino of Mexican descent and 42% are at 200% of FPL. The study sample reflects the population of this region. This is one of the first studies examining the impact of an integrated care model featuring faith-based behavioral health counseling with a primarily Hispanic population. Given these characteristics, external validity could be a limitation. Use of a secondary comparison group with different demographic and morbidity characteristics, although also primarily Hispanic, may address some of this limitation. Study findings, however, can inform other organizations interested in faith-based integrated care.

Given the internal validity of this study, the fidelity to which the evaluation and program were implemented, the significant results, the absence of negative intervention effects, the existence of small effect sizes for multiple outcomes (exploratory), and the unique and important contribution to the field, this study achieves a preliminary level of evidence to improve our understanding of the impact of a faith-based integrated care model. Implications of these findings are explored further below.

Summary of Implementation Findings

The implementation evaluation examined fidelity to Mercy's program model by conducting focus groups and interviews and examining patient visit data. A slightly delayed timeline in data collection was the main deviation from the SEP; mid-point interviews were conducted 8 months post-enrollment rather than 6 months, and final interviews and focus groups were conducted 4 months after study conclusion rather than immediately after.

Evaluation of the implementation of Mercy's program shows that the program was implemented in alignment with the program logic model and that the program was implemented with strong fidelity. Mercy met the enrollment target for the study and 85% of their overall 12-month retention target (final sample was 293 total participants compared to a target of 328 participants.)

All participants enrolled in the intervention met study eligibility criteria, and all who remained in the study for the 12 months received the intervention as designed including physical and behavioral health referrals and services. Intervention group participants received nutrition education, behavioral health counseling with an option for faith-based counseling, coordinated care and other services, while the primary comparison group participants received usual clinic care that did not include the option of faith-based counseling, care navigation, or exercise coaching physical activity and nutritional education classes. Intervention group participants received more than 1500 referrals for integrated services compared to the 44 for the primary comparison group participants. These same intervention group participants completed 70% of their referral visits. More specifically, intervention group participants received 184 referrals for behavioral health services. All of these participants qualified for and chose faith-based behavioral health counseling services and 98% of these referrals were completed.

Of the five core principles in the AIMS IBH checklist (patient-centered care, population-based care, measurement-based treatment to target, evidence-based care, and accountable care), Mercy applies four of them to most or all of their patients. This represents an improvement in each of these principles from pre-intervention implementation.

Facilitators to program implementation included communication among staff, staff experience with using an electronic medical record, moving staff offices to facilitate communication among medical and behavioral health staff, hiring staff who had specific roles that supported IBH, leadership and staff buy-in to the program, and clinic workflow adjustments to ensure patient needs for services could be met. For patients, additional factors that facilitated their participation included the low cost of services, clinic staff flexibility to meet their needs, strong rapport between patients and staff, and support for patient transportation services.

Strong program implementation and the facilitators discussed above led to high satisfaction with the program among staff and patients. Patients reported that the services they received at Mercy Ministries were "complete", meaning these services addressed patient physical, behavioral, and spiritual needs in one setting. Further, patients reported high satisfaction because the services increased their health knowledge and led to real and perceived improvements in health, such as weight loss and improved perceived quality of life.

Implementation barriers included layout of the clinic space which resulted in multiple adjustments to location of staff offices as well as general space limitations and difficulties in finding qualified applicants for specific staff roles. For patients, the cost of services and transportation to services were challenges in addition to being facilitators. An additional barrier was the socio-political environment which heightened anxiety of the region in general.

Summary of Impact Findings

The QED impact study and its related analyses were conducted as proposed in the SEP with the exception of the enrollment and follow-up periods. Enrollment was extended by four months to ensure an adequate sample in the primary comparison group. This resulted in proportional extension in the follow-up assessment timelines.

The QED impact study demonstrated that the faith-based integrated care model had a significant association with physical and behavioral health improvements among study participants. After 12 months in the program, intervention participants were more likely than primary comparison group participants to see significant improvements in depression over time, quality of life and quality of life over time, and anxiety and anxiety over time, when controlling for age, sex, and baseline characteristics. After 12 months in the program intervention participants were more likely than secondary comparison group participants to see significant improvements in depression over time, quality of life over time, anxiety over time, diastolic blood pressure, HbA1c, and female waist circumference and female waist circumference over time, when controlling for age, sex, and baseline characteristics. Given the strength of the study design, there is considerable evidence that the intervention contributed to the positive changes in health outcomes among participants. However, no significant changes were seen in obesity, waist circumference among males, or systolic blood pressure in the intervention group compared to either the primary or secondary comparison group.

Although there are some consistent findings among differences between the intervention and both comparison groups, the findings are not completely consistent. The discrepancy in impact findings for the two comparison groups is likely driven by differences in characteristics among all three study groups as well as the intervention. Compared to the primary and secondary comparison groups, intervention

group participants were more likely to have higher scores on depression and anxiety measures and lower scores on quality of life measures at baseline.

Intervention and primary comparison group participants had similar demographics and baseline physical health measures, except for employment status. Intervention group participants were less likely to be employed than primary comparison group participants. According to Mercy staff, this reflects a difference in formal employment between the two groups where participants in either group would only self-assess as being employed if they had a full-time job. The absence of statistically significant differences on physical outcome measures may be due to several factors. First, the intervention and primary comparison groups had a similar number of median primary care visits over the 12-months program period (7 and 6 respectively). Participants in both groups were seeing a primary care provider typically every other month, which likely heightened participant attention to physical health. For example, both the intervention and primary comparison group female participants had similar reductions in waist circumference between baseline and 12-month follow-up.

Second, it is possible that there was intervention contamination in the primary comparison group because of all the heightened activity at the Mercy clinic to care for one's physical and behavioral health. Although there appears to be no contamination due to primary comparison group participants participating in health coaching or physical activity, it is possible that these participants improved their physical health through their own efforts or accessed these services through other offerings in the Laredo area.

In contrast, the secondary comparison group differed from the intervention group on many demographic measures and had poorer physical health measure scores at baseline. Statistical analyses procedures were used to control for these differences in analyses models. Patients in the Alton and Edcouch clinics received no additional services, with the exception of nutritional counseling offered later in the study period.

Lessons Learned

This evaluation contributes to our understanding of the impact of the integration of behavioral health services, including optional faith-based counseling services, in a primary care service context. Prior evidence for this approach includes RCTs from Druss et al. which found positive effects of integrated behavioral health and Worthington and colleagues meta-analysis of RCTs found that religious/spiritual counseling resulted in greater improvements in psychological and spiritual outcomes as compared to alternate secular therapies (Worthington, Hook, Davis, & McDaniel, 2011). A more recent review of 30 studies using religious and spiritual interventions found that these approaches improve quality of life, pain outcomes, physical activity, and health behaviors (Gonçalves, Lucchetti, Menezes, and Vallada 2017). The results of the Sí Three evaluation build on this work by examining the impact of the integrated model with optional faith-based behavioral health services in a predominantly Hispanic, low-income population.

While the intervention and evaluation were implemented with strong fidelity, many lessons emerged that could inform other organizations interested in implementing an integrated care model with optional faith-based behavioral health services.

Operational Facilitators

As detailed in findings from the implementation evaluation, there were a number of critical elements from an operational perspective that facilitated Mercy's success. First, leadership support and staff commitment to the program were instrumental to early implementation success and quality improvement efforts around implementation. Specifically, Mercy's president provided strong leadership for the program and supported staff throughout program implementation. Clinic staff approached implementation by communicating among themselves and with participants, focusing on developing a strong rapport and consistent engagement with patients. Internal communication occurred through staff huddles to address implementation issues and meet patient needs. As a team, the clinic staff had the experience and ability to provide a range of behavioral health counseling and physical health improvement strategies to meet patient needs.

Second, the intervention extended existing services and was designed to meet observed patient needs. According to Mercy staff, the Laredo community as a whole is very faith oriented, which is exemplified by using "god" language in everyday conversations. Spiritual care always has been a component of Mercy's care, so the addition of integrated behavioral health services with faith-based counseling was a natural addition to existing services. In addition, Mercy's Sí Three program provided tailored intervention strategies to meet the varied physical and behavioral health needs of the patient population. The care coordinator was instrumental in facilitating participant engagement in the tailored intervention through coaching so that participants would appreciate the importance of staying connected with care. In addition, the care coordinator was persistent in contacting patients so that they would not be lost to follow-up.

Third, clinic staff were committed to ensuring that the clinic flow supported intervention services including a major relocation of offices and non-clinical staff. Because the Mercy clinic was not initially designed for integrated care, clinic staff, mostly in the early implementation of the intervention, found that the clinic flow was not optimal for the Sí Three program. Clinic staff reworked clinic flow several times so that it made sense to patients and staff and facilitated patients seeing all providers before leaving.

These operational facilitators led to high provider and patient satisfaction with the Sí Three program at the Mercy clinic. In turn, this satisfaction and strong rapport with patients led to high participation rates in intervention activities by intervention participants. These results support the theory of change for the Sí Three program—high satisfaction with and participation in intervention services would result in improved physical and behavioral health measures among participants.

Sustainability Planning

Despite the effectiveness IBH can have on patient health, a number of persistent challenges continue to create barriers to IBH implementation. At the forefront of these concerns is deciding how to best support consumers with complex, co-morbid needs to address patient health and be financially sustainable. This program was sponsored by a grant from Methodist Healthcare Ministries through the Social Innovation Fund and matching funds from the Lamar Bruni Vergara Trust, Guadalupe and Lilia Martinez Foundation, Mercy Caritas, the Hogg Foundation, and the Meadows Foundation. In applying for the multi-year grant, program planning focused on a model that would be most effective to improve health within this primarily Hispanic, low income population. Mercy has already expanded IBH services to the entire clinic and is actively addressing challenges of financial sustainability of the model.

Evaluation Lessons

Mercy's success with the evaluation study can be attributed to several factors. Prior to implementing the intervention, Mercy conducted mock patient drills that included staff role play as patients and completing all survey instruments to calculate time required for patient enrollment. The program manager observed the clinic flow among providers and ensured that all staff were familiar with data collection and study protocols. These drills also allowed Mercy to identify challenges with clinic flow and needs for additional staff training. In addition, Mercy staff had experience with using the EMR for data input, record keeping, and analysis. This allowed for rapid modification of the EMR to accommodate additional data collection fields for clinical and evaluation study needs.

Although the EMR could quickly be adapted, the external evaluator experienced data submission and analysis challenges because the EMR was created primarily for clinical rather than research data collection needs. Specifically, the output for each patient encounter for each individual patient was on separate spreadsheet lines, rather than a continuous spreadsheet line for an individual patient. This resulted in considerable manual data cleaning and recoding by the external evaluator which could have potentially introduced human error.

Study Limitations and Implications for Future Research

It is important to note the limitations of this study. Mercy evaluation findings show that intervention participants were more likely than primary comparison group participants to experience significant improvements in their depression, quality of life, and anxiety but there were no statistically significant improvements observed in blood pressure, obesity, or diabetes. Further, Mercy evaluation findings show that intervention participants were more likely than secondary comparison group participants to see significant improvements in their depression, quality of life, anxiety, diabetes, diastolic blood pressure, and waist circumference among females, but there were no statistically significant improvements in systolic blood pressure or waist circumference among males, perhaps due to small sample size. Given that Mercy did not reach retention targets among the intervention and primary comparison groups, it is possible that there was insufficient power to detect significant differences in all physical measures between these groups. Moreover, the Mercy intervention population had the highest depression and anxiety scores and lowest quality of life scores among the three groups and the differences between the intervention and comparison groups on these measures at baseline were significant. The improvements observed in the intervention group on these measures may be due to participating in a research study in general in addition to the effectiveness of the intervention.

As previously discussed, the Sí Three program was evaluated using a QED evaluation design with primary and secondary comparison groups to minimize threats to internal validity. Using a primary comparison group allowed for the examination of observed improvements in the intervention group as they relate to patients who choose to obtain services at the same clinic but choose not to receive the intervention (patients with similar demographics and disease characteristics in the same setting).

The secondary comparison group was included to examine observed improvements in the intervention group as they relate to patients who use a different clinic (factors related to being part of a larger population). These participants were significantly different from the intervention group on nearly all outcome measures at baseline. Although propensity score matching was not used because of an

insufficient number of variables to match on, the complete case analyses sufficiently controlled for these baseline differences.

This study examined the effectiveness of the intervention as a whole and was not designed to evaluate the effectiveness of each specific component of the intervention. Mercy created this approach to meet the needs of the clinic patients, who are primarily Hispanic and low income. In the future, researchers might want to consider examining the extent to which other specific populations would benefit from a highly tailored integrated behavioral health model. In addition, this study demonstrated that the primarily Hispanic, low income population that Mercy serves preferred the spiritual behavioral health treatment approach. Researchers also may wish to examine if faith-based or spiritual behavioral health counseling is more effective than secular behavioral health counseling among other populations.

Next Steps

Mercy now offers the Sí Three program to all adult patients. Financial resources to maintain the program for all patients poses the greatest challenge for sustainability.

OTHER ASPECTS OF STUDY LOGISTICS AND FEASIBILITY

Human Subjects Protection

Mercy Ministries received Institutional Review Board approval from Mercy Health Springfield for a duration of 12 months beginning January 15, 2016. In accordance with Mercy Health Springfield procedures, Mercy submitted Continuing Review/Progress reports and received approval for continuation of the research in January and October 2016, July 2017, and July 2018. The Springfield location was merged with Mercy Saint Louis and the protocol remains at the Saint Louis facility. No deviations in research protocol have occurred to date.

Timeline

Mercy experienced challenges during their enrollment due to the high number of patients who were interested in enrolling in the intervention. To address shortfalls in primary comparison group enrollment, Mercy extended their recruitment timeline. Enrollment was completed in July 2016. This timeline represents a change from the SEP and is reflected in **Appendix A: Revised Project Timeline.** No other major changes to the timeline occurred during the study.

Evaluator/Subgrantee Role and Involvement

No major changes were made to the evaluator and subgrantee personnel listed in the subgrantee evaluation plan during the program period. The Principal Investigator of record for the study under the IRB protocol is Sister Maria Luisa Vera of Mercy Ministries of Laredo.

Budget

No changes were made to the evaluation budget.

REFERENCES

Advancing Integrated Mental Health Solutions Center (AIMS), University of Washington (2011). Retrieved. http://aims.uw.edu/resource-library/checklist-collaborative-care-principles-and-components

American Diabetes Association. (2014). Standards of medical care in diabetes--2014. *Diabetes Care*, *37 Suppl 1*, S14–80. http://doi.org/10.2337/dc14-S014

American Heart Association. (2015). Understanding Blood Pressure Readings. Retrieved July 22, 2015, from http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp

American Psychiatric Association. (1994). *Quick Reference to the Diagnostic Criteria from DSM-IV*. Washington, DC: Press, Inc.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Brown, R. L., & Rounds, L. A. (1995). Conjoint screening questionnaires for alcohol and other drug abuse: criterion validity in a primary care practice. *Wisconsin Medical Journal*, *94*(3), 135–40. Retrieved

CDC. (2011). Health Related Quality of Life. Retrieved July 22, 2015, from http://www.cdc.gov/hrqol/concept.htm

Cohen, D. J., Balasubramanian, B. A., Davis, M., Hall, J., Gunn, R., Stange, K. C., … Miller, B. F. (2015). Understanding Care Integration from the Ground Up: Five Organizing Constructs that Shape Integrated Practices. *Journal Of The American Board Of Family Medicine: JABFM*, *28 Suppl 1*, S7–S20. https://doi.org/10.3122/jabfm.2015.S1.150050

Daaleman, T. P., Frey, B. B., Wallace, D., & Studenski, S. A. (2002). Spirituality Index of Well-Being Scale: development and testing of a new measure. *The Journal of Family Practice*, *51*(11), 952. Retrieved from http://utsa.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMw1V3JTsMwELUKSIgLYg2r5BMXlCqJ3cZG4lCxqIhdLRK3yrEd1ANlS_-f8dakID6AU5WmaRv7zZqZNwiRrJ3EP3QCK6lgHUEVzXRadFQpwDQJkmQSTJLsqDCawHLYtFqh2usHr80_3_jB-xgC_qnzry0ZonEITY4uLrTNHkhTQ0x6oWFqVmZeGcYNVwQtzKjx41

Druss, B. G., Rohrbaugh, R. M., Levinson, C. M., & Rosenheck, R. A. (2001). Integrated medical care for patients with serious psychiatric illness: a randomized trial. *Archives of General Psychiatry*, *58*(9), 861–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11545670

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken, N.J.: Wiley-Interscience.

Goodman, S.N., & Berlin, J.A. (1994). The Use of Predicted Confidence Intervals When Planning Experiments and the Misuse of Power When Interpreting Results. *Ann Intern Med, 121, 200-206.* Retrieved from: https://dirnagl.files.wordpress.com/2014/07/goodman-the-use-of-predicted-confidence-intervals-when-planning-experiments-and-the-misuse-of-power-when-interpreting-results-ann-intmed-1994.pdf

Gonçalves, J.B., Lucchetti, G., Menezes, P.R., & Vallada, H. (2017). Complementary religious and spiritual interventions in physical health and quality of life: A systematic review of randomized controlled clinical trials. *Plos One, 12(*10), e0186539. Doi:10.1371/journal.pone.0186539.

Grady, C. (2005). Payment of clinical research subjects. *The Journal of Clinical Investigation*, *115*(7), 1681–7. http://doi.org/10.1172/JCI25694

Guide to Community Preventive Services. (2010). Improving mental health and addressing mental illness: collaborative care for the management of depressive disorders. Retrieved June 12, 2015, from http://www.thecommunityguide.org/mentalhealth/collab-care.html

Heath, B., Wise Romero, R., & Reynolds, K. (2013). *A Standard Framework for Levels of Integrated Healthcare*. Retrieved from http://www.integration.samhsa.gov/integrated-care-models/A_Standard_Framework_for_Levels_of_Integrated_Healthcare.pdf

Koenig, H. G. (2012). Religion, Spirituality, and Health: The Research and Clinical Implications. Retrieved June 26, 2015, from http://uq5sd9vt7m.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info:sid/summon.serialssolutions.com&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.atitle=Religion%2C+spirituality%2C+and+health%3A

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatric Annals*, *32*(9), 509–515. http://doi.org/10.3928/0048-5713-20020901-06

Lasser KE, Quintiliani LM, Truong V, Xuan Z, Murillo J, Jean C, Pbert L. Effect of Patient Navigation and Financial Incentives on Smoking Cessation Among Primary Care Patients at an Urban Safety-Net Hospital: A Randomized Clinical Trial. *JAMA Intern Med.* 2017 Dec 01; 177(12):1798-1807

Liebschutz JM**,** Xuan Z, Shanahan CW, LaRochelle M, Keosaian J, Beers D, Guara G, O''Connor K, Alford DP, Parker V, Weiss RD, Samet JH, Crosson J, Cushman PA, Lasser KE. Improving Adherence to Long-term Opioid Therapy Guidelines to Reduce Opioid Misuse in Primary Care: A Cluster-Randomized Clinical Trial. *JAMA Intern Med*. 2017 Sep 01; 177(9):1265-1272.

National Guideline Clearinghouse. (2014). Obesity: identification, assessment and management of overweight and obesity in children, young people and adults. Rockville MD: Agency for Healthcare Research and Quality (AHRQ). Retrieved from http://www.guideline.gov/content.aspx?id=48872&search=waist+circumference

National Guidelines Clearinghouse. (2011). Hypertension. Clinical management of primary hypertension in adults. Rockville MD: Agency for Healthcare Research and Quality (AHRQ). Retrieved from http://www.guideline.gov/content.aspx?id=34824

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity. 41. 673-690. 10.1007/s11135-006-9018-6.

Parkerson, G. R., Broadhead, W. E., & Tse, C. K. (1990). The Duke Health Profile. A 17-item measure of health and dysfunction. *Medical Care*, *28*(11), 1056–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2250492

Rubin DB and Thomas N. Matching using estimated propensity scores: relating theory to practice, 1996, BIOMETRICS , 52, 249-264.

Sadock, B., & Sadock, V. (2007). *Kaplan & Sadock's Synopsis of Psychiatry: Behavioral Sciences/Clinical Psychiatry* (10th editi). Philadelphia, PA: Lippincott, Williams & Wilkins.

Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Ann Int Med 2010;152. Epub 24 March.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–7. http://doi.org/10.1001/archinte.166.10.1092

Stuart EA. Matching methods for causal inference: A review and look forward. Stat Sci. 2010 February 1; 25(1):1-21.

U.S. Census Bureau. (2013). 2013 American Community Survey 5-Year Estimates. Retrieved from http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_5YR_S0501&prodType=table

University of Wisconsin Population Health Institute. (2015). County Health Rankings 2015. Retrieved June 26, 2015, from http://www.countyhealthrankings.org/sites/default/files/state/downloads/2015 County Health Rankings Texas Data - v1_0.xls

World Health Organization definition of integrated care. (2008). Retrieved from http://www.who.int/healthsystems/service_delivery_techbrief1.pdf

Worthington, E. L., Hook, J. N., Davis, D. E., & McDaniel, M. A. (2011). Religion and spirituality. *Journal of Clinical Psychology*, *67*(2), 204–14. http://doi.org/10.1002/jclp.20760

APPENDICES

Sí Texas Subgrantee: Mercy Ministries of Laredo
Program Title: Sí Three: Integration of 3-D Health Services

Appendix A: Revised Project Timeline

	2015								2016												2017												2018									
	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10
Planning & Program Administration																																										
Program awarded	X																																								
SEP development & approval		X	X	X	X	X	X	X	X	X	X	X	X	X																											
Protocol development		X	X	X	X	X	X	X																																	
Instrument development		X	X	X	X	X	X																																		
IRB approval process				X	X	X	X	X																																	
Staff training		X	X	X	X	X	X	X																																	
Program start									X																																
Program implementation																																										
Enrollment									X	X	X	X	X	X	X																											
Data Collection									X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X															
Baseline *(0-7 month)*									X	X	X	X	X	X	X																											
Intermediate *(6-9 months)*														X	X	X	X	X	X	X	X																				
Final *(12 month)*																				X	X	X	X	X	X	X	X														
Data analysis* & reporting																																										
HRiA (quarterly reporting)												X			X			X			X			X			X			X				X							
Data cleaning & analysis[1,2]												X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X						X	X	X	X	X			
Report writing & editing[1,2]																						X	X	X	X					X	X					X	X	X	X	X	X	X
Report to CNCS[1,2]																									X	X																X
Reports to partners/stakeholders[1,2]																											X								X	X	X	X	X	X	X	X
Reports to general public/scientific com. [1,2]																											X								X	X	X	X	X	X	X	X

**HRiA has been contracted by MHM as the Si Texas program evaluator. All data analyses and reporting will be done on a collaborative basis with the subgrantee; [1] Annual; [2] Final*

Appendix B: Program Logic Model

Figure 1. Logic Model

			Outcomes		
Inputs/Resources	**Activities**	**Outputs**	**Short-term**	**Intermediate**	**Long-term**
Program personnel: • Principal investigator • Program Manager • Navigators • Care coordinator • LPC/pastoral counselor • Behavioral Coach • Contracted Evaluators • Data Entry • Exercise Coach • Nurse Educator **Program partners:** • LPC (part-time) • Laredo Health Dept • Border Region Behavioral Health Center • SCAN • Faith-based counselors • Other organizations **Program funder:** • Methodist Healthcare Ministries	**Clinic level:** • Clinic-capacity building activities: Phone conferences, case conferences, face-to-face interactions, EPIC training, ongoing feedback and mentoring • Administer staff satisfaction surveys • Communications with patients about and coordination of internal and external components of patient's behavioral and physical health • Enter all patient data in EMR • Track, monitor, and remind patients of appointments **Patient level:** • Diagnosis of diabetes, obesity and hypertension by the navigator. • ***Administration of surveys to assess behavioral health and spirituality.*** • ***Administer patient satisfaction surveys*** • ***Re-assess patients quarterly on physical and behavioral health measures***	**Clinic level**: • ***Recruit 205 participants into each arm of the study*** • ***Increased understanding of integration*** • ***Provider and staff buy-in to model*** • Use of standard measurement protocols • Ongoing quality improvement among clinic staff **Patient level**: • Development of a patient care plan (including behavioral health treatment plans) • ***Referral to one or more in-house services (exercise coach, educator, dietician, medical and/or faith-based behavioral health counselor) and/or community resources and chronic disease management programs aligned with patient needs***	**Clinic level**: • Scheduling of follow-up appointments with in-house or community resources • Entering data for and tracking and monitoring patient use of services • Improved communication across providers • Improved adherence to program model **Patient level**: • Improved patient knowledge of and skills for self-management • Patients take an active role and can explain their treatment plans	**Clinic level**: • Improved workflow alignment across providers and services • Improved clinic efficiency • Improved rate of successful referrals • Improved provider satisfaction **Patient level**: • Patients participate in and are satisfied with in-house or community resources for physical health, medical and/or faith-based behavioral health counseling • Improved Waist circumference, BMI, A1c, blood pressure, depression, anxiety, addiction, and quality of life	**Clinic level**: • Providers and staff involved with 3D integrated services will move from level 3 and 4 to level 5 **Patient level**: • Improved Waist circumference, BMI, A1c, blood pressure, depression, anxiety, self-management and quality of life

Please note that activities/outputs specific to the Sí Three intervention are in ***bold, italicized text.***

Appendix C: Sí Texas Mid-Point Implementation Evaluation: Key Informant Interview General Guide

INTERVIEW GOALS

• To collect qualitative information about the implementation of the Sí Texas initiative
• To understand whether the intended target population has been reached at each subgrantee site
• To learn whether what was planned for implementation was actually implemented, and to identify facilitators and barriers of adoption
• To learn what has gone well during the initial phase of the Sí Texas project at the subgrantee level and what needs improvement, and to understand plans for making improvements in the future

INTRODUCTION/INFORMED CONSENT

- Thank you for taking the time out of your day to meet with us. My name is [name] I am a researcher at Health Resources in Action, and today I am joined by my colleague [name] who will assist me during our interview.

- Our goal today is to collect perspectives about the implementation of your Sí Texas project. We hope to learn what has gone well during this initial phase of the project. We are also interested in learning about any challenges that may have been encountered during this period, and your perspectives about what's ahead for the program.

- The interview should last approximately 45 minutes to one hour. I want to remind you that this interview is voluntary and confidential. What we talk about in this space stays in this space so feel free to share your opinion openly and honestly without worrying that it will be repeated. You may choose not to answer any questions during the interview and we can stop at any time. Your interview answers will be summarized in a report along with the interviews from other interview participants.

- I will not identify [name of subgrantee], your name, or your organization's name with your responses in any publication. At the end of the study, we will return to many of our interviewees and ask to re-interview them after the program period has ended. However, participating in this interview does not mean you have to participate in a subsequent interview. The final interview is also voluntary.

- Do you have any questions about the study or how your responses will be used? I would also like to record our session today to make sure our notes are complete and correct, but we will delete the recording after we verify and save our notes. We won't use names in our notes. Are you okay with me recording our discussion?

- As a reminder, when you answer a question, please do not use client's/patient's names. We would appreciate you provide more general examples if you would like to describe a specific situation.

INTERVIEW QUESTIONS

1. Key Informant Background
   - What is your current role, and how long have you served in this role? How long have you been with your organization?
   - What are your responsibilities at [subgrantee/organization]?
   - Do you have any responsibilities for running the [name of subgrantee Sí Texas program]? If so, would you tell us about those responsibilities?
   - What was your involvement in the [name of subgrantee Sí Texas program] planning process? What was that process like?

For the remaining questions, the interviewer will select questions to ask based on the person being interviewed and the subgrantee's specific needs/implementation questions. It is recommended that those questions be selected prior to interview.

2. Level of Integrated Behavioral Health
   - What do you understand the goals of the Sí Texas project to be?
   - Prior to the program's implementation, did your program offer both primary care and behavioral health services?
     - What did that look like? To what extent were primary care and behavioral health services connected/coordinated/combined, if at all?
     - [For programs with other integration goals]: To what extent are [services] integrated?
       - Probes: in what way are services integrated? Coordinated? (e.g., IT, workflow)
   - Now that the [name of subgrantee Sí Texas program] has been implemented, to what extent are primary care and behavioral health services connected/coordinated/combined, if at all?
     - How feasible has it been to integrate these services? (If applicable)

3. Program Components and Population
   - How are participants identified for the program? What is/was the enrollment process like?
     - How were participants assigned to the intervention or control group? (For randomized control trials, ask the participant to describe the randomization process.)
     - When a participant enrolls in the program, what happens to them next? Take me through the services and activities that an enrollee receives in the program.
       - Probe: Are warm hand offs between providers a component of the services participants receive? How do those hand offs work? (If applicable)
     - How are behavioral health/health coaches accessed or how do they become involved in patient care?
   - Since beginning enrollment, to what extent has the program been able to deliver all the program services that had been planned as part of the program intervention? (Ask those who had a role in planning the program)
   - Since the program started, has anything changed about the services that intervention group participants received or activities they have access to at your clinic? In what way?
   - To what extent/Have any adjustments been made to program operations or offerings based on your early experience implementing the program?
   - How would you describe the population that your program is serving?

- What are they like in terms of demographics generally? Is this the population it intended to serve?

4. Adoption
   - To-date, what have been the most successful parts of the program? Why?
   - To-date, what have been the least successful parts of the program? Why?
   - Please describe any barriers you or your organization has experienced in implementing the program.
     - In what ways did these barriers affect program implementation? In what ways have you been able to address these barriers?
   - Please describe anything that has helped your organization implement the program.
     - Probes: Is the staff, the facilities, the data systems, outside partners, or other things?
   - What kind of training did you develop/participate in as part of the program?
     - Did this training prepare you for your responsibilities in the program? If not, what was missing from the training?
   - What, if any, concerns have program staff raised about the program? How about non-program staff (if relevant)?
     - What has been the response, if any, to those concerns?

5. Control Group Program-Like Components (if applicable)
   - When a participant is randomized/enrolled in the control/comparison group of your program, what can they expect to receive or participate in terms of services or activities?
   - Since the program started, has anything changed about the services that control group participants received or activities they have access to at your clinic? In what way?
     - Have those changes been experienced by the intervention group? If no, why not?

6. Operations (Choose Clinic or Community as appropriate)

Clinic-based Operations

- In what ways have clinic operation workflow changed due to implementation of your project?
- What do you see as the impact of this workflow change, if any?
  - Have these changes had any effects on patient care for those participants <u>not</u> enrolled in the study? In what way?
- To what extent have information/data systems/your EMR been changed to support the program? Have you added any information/data systems for the project?

Community-based Operations

- How, if at all, has your agency operation workflow changed due to implementation of your project?
- What do you see as the impact of this workflow change, if any?
  - How, if at all have these workflow changes affected client care for those participants <u>not</u> enrolled in the study? In what way?
- To what extent have information/data systems been changed to support the community program? Have you added any information/data systems for the project?

7. Patient and Provider Satisfaction

   [Remind respondent not to identify participants by name or to use any identifying information when giving examples]

- What do you think participants in general would say about the program? Would you mind sharing any general themes from feedback you have heard from participants about the program?
- Have you heard any feedback from providers about program implementation? What are some of the general themes from their feedback been?
- To what extent have there been challenges to retaining primary care, behavioral health, or community-based staff during the course of the [name of subgrantee program]? Why do you think there have been challenges, and what has been done to address those challenges?

8. External Partnerships (if applicable)
    - How would you describe your partnership(s) with external organizations related to this program? What role have these partnerships played in early implementation?
    - How has the partnership been helpful in promoting implementation of program activities?
    - To what extent have there been challenges in building and maintaining productive partnerships to-date?
    - Are there any gaps in program activities that were the responsibility or role of a partner? Would you share with me any steps your organization has taken (or will take) to overcome this gap?

9. Sustainability and Lessons Learned
    - If you could go back in time and change anything about getting the program started, what would that change be? Why?
    - What changes, if any, would you want to make at this point in the program?
    - What lesson have you learned to-date from the early experiences of your program that you would want to share with other organizations thinking of implementing your program in their setting?

10. Closing

Thank you so much for your time. That's it for my questions. Is there anything else that you would like to mention that we didn't discuss today?

Appendix D: Sí Texas Summative Implementation Evaluation: Key Informant Interview General Guide

Sí Texas Summative Implementation Evaluation: Key Informant Interview General Guide

CORE INTERVIEW GOALS

- To understand how primary care and behavioral health services are integrated (in various settings) from the perspective of staff (clinic and non-clinic)
- To identify perceived facilitators and barriers to adoption of the IBH model, including external factors
- To identify program successes, challenges, opportunities for improvement, and lessons learned for sustainability
- To better understand the perceived impact of the program on participants' health and wellbeing.

INTRODUCTION/INFORMED CONSENT (2 MIN)

- Hi, my name is [name] and I am a researcher at Health Resources in Action. I am also joined by my colleague [name] who will assist me during our interview. Thank you for taking the time to speak with us today.

- We are speaking with a variety of people to better understand the implementation of [name of subgrantee Sí Texas program]. We are interested in learning what has worked well, challenges that may have been encountered, and any advice or lessons learned that could inform future planning or sustainability of programs like [name of subgrantee Sí Texas program].

- The interview should last approximately [INSERT TIME: 30-60 minutes]. I want to remind you that this interview is voluntary and confidential. What we talk about in this space stays in this space so please feel free to share your opinions openly and honestly. You may choose not to answer any questions during the interview and we can stop at any time. We are conducting several interviews such as this one and will be writing a summary report that pulls out common themes. We will not identify you in our report or any future publication.

- Do you have any questions about the study or how your responses will be used? I would also like to record our session today to make sure our notes are complete and correct, but we will delete the recording after we verify and save our notes. We won't use names in our notes. Are you okay with me recording our discussion?

- As a reminder, when you answer a question, please do not use client's/patient's names. We would appreciate you provide more general examples if you would like to describe a specific situation.

INTERVIEW QUESTIONS

[**NOTE:** IF INTERVIEWEE PARTICIPATED IN MID-POINT DATA COLLECTION, PLEASE FRAME CONVERSATION AS NEEDED TO ACKNOWLEDGE PREVIOUS DISCUSSION (E.G., since we last interviewed you, what additional changes were made to better connect or coordinate services?)]

Key Informant Background (3 MIN)

1. I'd like to start by asking you a few questions about yourself. Can you tell me about your role in [name of subgrantee Sí Texas program]?
    a. How long have you been involved with the [name of subgrantee Sí Texas program]?
        i. Has anything about your role in the project changed since you started working with [name of subgrantee Sí Texas program]?

Integrated Behavioral Health Program Goals and Activities (10-15 MIN)

2. Now I'd like to talk about the program's goals and its specific activities. What do you see as the goals of [name of subgrantee Sí Texas program]? What were you hoping to achieve for participants?
    a. [SUBGRANTEE SPECIFIC PROBES: How about goals or desired outcomes for the wider community—for example, family members or care givers? Operational goals for [name of subgrantee Sí Texas program] (e.g., improving show rates to appointments, reducing wait times, etc.)]?

3. Can you walk me through the program: after a participant enrolled in the intervention group, what services or activities did they receive?
    a. After a participant enrolled in the control/comparison group, what services or activities did they receive?
    b. What changes, if any, were made to the services or activities offered to intervention participants? How about comparison/control group participants? Why?
        i. How did these changes affect the program?

4. Since implementing the [name of subgrantee Sí Texas program], to what extent have primary care and behavioral health services been connected or coordinated? How have these services been connected or coordinated?
    a. How easy or hard has it been to connect or coordinate these services? Why? (If applicable)
        i. What has made services more or less connected or coordinated?
        ii. What changes were made to better connect or coordinate services?
    b. [SUBGRANTEE SPECIFIC PROBE: How are primary care providers involved in patient care? [OR] How are behavioral health providers/health coaches involved in patient care?]
    c. [SUBGRANTEE SPECIFIC PROBE: Do warm handoffs occur between primary care and behavioral health? How do warm hand offs work? Since the program started, have any changes been made to how warm hand offs work?]

Adoption Facilitators and Barriers (15 MIN)

[NOTE TO INTERVIEWER: FOCUS ON FACILITATORS/BARRIERS TO IMPLEMENTATION NOT OUTCOMES]

5. Next I'd like to talk about your experience with implementing the program or putting it into practice. What worked well about putting the program into practice? Why? [PROBE ON ALL: LEADERSHIP, STAFF, COMMUNICATION, DATA SYSTEMS, EMR, PARTNERSHIPS, TRAINING, AND OTHER SUBGRANTEE SPECIFIC AREAS]
    a. What helped you/your organization implement the program?

6. On the flip side, what has not worked well about putting the program into practice? Why? [PROBE ON ALL: LEADERSHIP, STAFF, COMMUNICATION, DATA SYSTEMS, EMR, PARTNERSHIPS, TRAINING, AND OTHER SUBGRANTEE SPECIFIC AREAS]
    a. What barriers or challenges did you/your organization experience in implementing the program? [PROBE ON EXTERNAL FACTORS (e.g., natural disasters, legislation, funding shifts, political events, etc.)]
        i. In what ways have you been able to address these barriers?

7. [IF NOT YET MENTIONED:] Since the start of the [name of subgrantee Sí Texas program], what changes were made to how the program was implemented? Why? [PROBE ON: WORKFLOW, STAFFING, DATA SYSTEMS/EMR, POLICY, OTHER SUBGRANTEE SPECIFIC AREAS]
    a. How did these changes affect the program?

Provider and Patient Satisfaction (5 MIN)

8. [IF NOT YET MENTIONED:] I'm also interested in your perspective on others' experiences with implementing the program. What feedback have you heard from providers or staff about the process of implementing the program?
    a. How satisfied were providers or staff with the program?
    b. [SPECIFIC SUBGRANTEE PROBE: To what extent did providers or staff buy in to the program? How did this affect implementation?]

9. What feedback have you heard from participants about the process of participating in the program?
    a. [SPECIFIC SUBGRANTEE PROBE: How satisfied were participants with the program?]

Program Impact (5 MIN)

10. In your opinion, how effective was the program at achieving its goals?
    a. How do you think the program affected participants' health?
    b. To what extent do you think the program made an impact on participants' health?
        i. What was the program's impact on participant…? [PROBE ON SPECIFIC IMPACT MEASURES (e.g., diabetes, depression, BMI, etc.)]

11. What events or trends did you see as affecting program impact? (e.g., natural disasters, legislation, funding shifts, political events, etc.)

Sustainability and Lessons Learned (10 MIN)

12. Lastly, I'd like to talk about the future of [name of subgrantee Sí Texas program]. As the Sí Texas project draws to a close, what is the plan for [name of subgrantee Sí Texas program]?[PROBE ON PROGRAM CONTINUATION, REPLICATION, SCALING UP]
    a. Moving forward, how does [subgrantee] plan to improve or enhance the integration of primary care and behavioral health services?

13. If you could start over and implement this program from the very beginning, what changes would you make for the program to be more successful? Why? [PROBE ON DATA SYSTEMS, STAFFING, TRAINING, CLINIC SPACE, FUNDING]
    a. If a similar organization were planning to implement your program from the ground up, what advice would you give them?

14. What suggestions/recommendations do you have to help continue/sustain the positive efforts of [name of subgrantee Sí Texas program]? [PROBE ON PROGRAM REPLICATION, SCALING UP, FUNDING, POLICY CHANGE]

Closing (2 MIN)

Thank you so much for your time. That's it for my questions. Is there anything else that you would like to mention that we didn't discuss today?

Appendix E: Sí Texas Summative Implementation Evaluation: Focus Group Guide

Sí Texas Summative Implementation Evaluation:
Participant Focus Group Core Guide
October 11, 2017

NEIRB 120170278
#96104.0

CORE FOCUS GROUP GOALS

- To better understand the perceived impact of the program on participants' health and wellbeing.
- To assess how satisfied participants are with the services they have received (Note: Included in most but not all subgrantee SEPs)
- To identify perceived facilitators and barriers to participating in the program, including external factors
- To identify participant perceptions of program successes, challenges, and opportunities for improvement

INTRODUCTION (5 MIN)

- My name is [name] and this is my colleague [name] and we are from Health Resources in Action an organization working with [subgrantee name] that provides the [name of program/service/study]. Thank you for taking the time to speak with us today.

- We are talking with a variety of people involved in [name of subgrantee program/service/study] to better understand how the [program/services/study] worked. We are interested in hearing about your experience participating in the [program/services/study] and your ideas about how to make [program/services/study] better in the future. I want everyone to know there are no right or wrong answers to our questions. We want to know your opinions, and those opinions might not all be the same. This is fine. Please feel free to share your opinions, both positive and negative. What you share with us today will in no way affect the care you receive.

- I want to remind you that talking with us in this group is voluntary. You can leave anytime or choose not to answer any question we ask. We also want to do everything we can to make sure what we talk about in the group stays private, so we ask that you not share anything you hear today with anyone outside of the group. This is to make sure everyone feels comfortable sharing their opinions. We will definitely not share anything we hear today with anyone outside the group, but we can't be sure that something you say in the group won't be repeated by someone else in the group.

- We are speaking with several different groups such as this one and will be writing up a report of the general ideas we hear across all of the group. No one's name will be used in our summary. When we write our report we will mention that "some people said this" or "other people said that." No one will be able to tell it was you who said something in our report.

- Our conversation will last about an hour and a half. If you have a cell phone, please turn it off or use vibrate mode. If you need to go to the restroom during the conversation, please feel free to leave, but we'd appreciate it if you would go one at a time.

- [IF INCENTIVE IS OFFERED, OTHERWISE OMIT: Each of you will receive a [$amount] gift card for completing today's group conversation. To receive the gift card, you will need to put your initials

on a receipt for our records and we will give you a copy of that receipt. Our copy of the receipt will be kept private.]

- We would also like to record our session today to make sure our notes are complete and correct, but we will delete the recording after we verify and save our notes. We won't use names in our notes. Is everyone okay with me recording our conversation?

- Do you have any questions before we begin our introductions and conversation?

INTRODUCTION AND WARM-UP (5 MIN)

1. First let's spend a little time getting to know one another. Let's go around the table and introduce ourselves. Please tell me: 1) Your first name; 2) how long you've been in the [program/service/study] and 3) something about yourself – such as what you like to do for fun with your family. [AFTER ALL PARTICIPANTS INTRODUCE THEMSELVES, MODERATOR TO ANSWER QUESTIONS]

PROGRAM RECRUITMENT (10 MIN)

2. Let's get started by talking about how you first found out about the [name of subgrantee program/service/study]. Tell me a little bit about how you were introduced to this [program/service/study].
    a. How did you hear about the [program/service/study]?
    b. Who talked to you about it?
    c. How easy or hard was it to understand the information provided to you about the [program/service/study]?

3. Why did you join the [program/service/study]?
    a. What concerns, if any, did you have about joining the program/service/study?

PARTICIPANT EXPERIENCE: INTERVENTION/CONTROL GROUP (20-30 MIN)

4. I'd now like you to think about your experience as a participant of [name of program/service/study]. If you had to describe the [program/service/study] to a neighbor, what would you say? How would you describe the [name of program/service/study]?
    a. In your own words, what is the purpose/goal of the [name of program/service/study]?
    b. Who is the program/service for (e.g., for people who have diabetes or want to lose weight)?
    c. What services did you receive? What activities did you participate in? [ADD SUBGRANTEE SPECIFIC PROBES HERE]
        i. How often?
    d. How was this program/service/study similar or different to health services you received before the program/service/study?

5. What did you think about the program/service/study? On a scale of 1-10 [USE VISUAL SCALE], how would you rate your experience with the program/service/study? Why? [ADD PROBES ON INTERVENTION/CONTROL COMPONENTS HERE (E.G., CLINIC/COMMUNITY SERVICES, REFERALLS, CARE COORDINATION, COMMUNICATION BETWEEN PROVIDERS, ETC.]
    a. What did you <u>like best</u> about the program/service/study? Why?

   i. In what ways has the program/service/study met your needs?
   ii. What was helpful to you?
b. What did you <u>like least</u> about the program/service/study?
c. What could have made your experience better?

6. What did you think about the program/clinic staff (e.g., how they treated you, how comfortable you felt around them, etc.)?
7. How easy or hard was it to participate in the program/service/study?
   a. What made it <u>easier</u> to participate in the program/service/study?
      i. What helped you participate in the program/service/study? [PROBE: COST, SCHEDULE, LANGUAGE, TRANSPORTATION, INCENTIVES, ETC.]
   b. What made it <u>harder</u> to participate in the program/service/study? [PROBE: COST, SCHEDULE, LANGUAGE, TRANSPORTATION, POLITICAL EVENTS, HURRICANE HARVEY, ETC.]

PROGRAM VALUE/IMPACT (10-15 MIN)

8. How did participating in [name of program/service/study] affect you/your health?
   a. How about other parts of your life? [PROBE ON: WORK, RELATIONSHIPS WITH FAMILY, STRESS, SLEEP, ETC.]

9. How can the program/service/study be improved?
   a. What else could the program/service/study do to improve participants' health?
   b. What could have improved your experience in the [name of program/service/study]?
   c. What's missing? What kinds of services or activities would you want to see offered by the program/service/study?

10. Thinking about your experience in the [name of program/service/study], would you sign up for the program/service again? Why or why not?
    a. Would you recommend this [name of program/service/study] to someone else? Why or why not?

CLOSING/INCENTIVE DISTRIBUTION (2 MIN)

Thank you so much for your time. That's it for my questions. Is there anything else that you would like to mention that we didn't discuss today?

[OPTIONAL: OMIT THE FOLLOWING SECTION IF INCENTIVES NOT BEING USED:
I want to thank you again for your time. To express our thanks to you, we have [$amount] gift cards from [name of vendor, e.g., H-E-B]. [Name of HRiA staff person] has a receipt for you to initial and then he/she will give you your gift card. [DISTRIBUTE INCENTIVES AND HAVE RECEIPT FORMS SIGNED].]

Thank you again. Your feedback is very helpful, and we greatly appreciate your time and for sharing your opinion.

Appendix F: Implementation Evaluation Measures

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
REACH: Did the Sí Three's program reach its intended target population?				
--	Demographic characteristics of participants	Eligibility criteria data	• How would you describe the population that your program is serving? • What are they like in terms of demographics generally? • Is this the population it intended to serve?	None
FIDELITY: What are the components of Sí Three's program and how do these components work "on the ground" at 6 and 12 months? Are these components different than what was planned? If so, why? To what extent did the Mercy clinic implement the Sí Three model with fidelity?				
What are the resources of the program?	Input: Principal investigator	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Program Manager	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Navigators	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Care coordinator	--	What is your current role?	Yes/No
What are the resources of the program?	Input: LPC/pastoral counselor	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Behavioral Coach	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Contracted Evaluators	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Data Entry	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Exercise Coach	--	What is your current role?	Yes/No

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
What are the resources of the program?	Input: Nurse Educator	--	What is your current role?	Yes/No
What are the resources of the program?	Input: Laredo Health Dept	--	How has the partnership been helpful in promoting implementation of program activities?	Yes/No
What are the resources of the program?	Input: Border Region Behavioral Health Center	--	How has the partnership been helpful in promoting implementation of program activities?	Yes/No
What are the resources of the program?	Input: SCAN	--	How has the partnership been helpful in promoting implementation of program activities?	Yes/No
What are the resources of the program?	Input: Faith-based counselors	--	How has the partnership been helpful in promoting implementation of program activities?	Yes/No
What are the resources of the program?	Input: Other organizations	--	How has the partnership been helpful in promoting implementation of program activities?	Yes/No
What are the resources of the program?	Input: Methodist Healthcare Ministries	--	How has the partnership been helpful in promoting implementation of program activities?	Yes/No
What are the program activities and how have they been operationalized?	Activity: Communications with patients about and coordination of internal and external components	• Patient satisfaction with Si Three (by type of service)	Since beginning enrollment, to what extent has the program been able to deliver all the program services that	Record of communication with patient

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
	of patient's behavioral and physical health		had been planned as part of the program intervention?	
What are the program activities and how have they been operationalized?	Activity: Enter all patient data in EMR	• Record of vitalization of blood pressure, height, weight, and waist circumference • Record of blood test results for HbA1c • Record of patient treatment plan created • Number of patients with all intake forms and assessments completed (e.g., PHQ-9, Duke Health Profile, etc.)	• To what extent have information/data systems/your EMR been changed to support the program? • Have you added any information/data systems for the project?	None
What are the program activities and how have they been operationalized?	Activity: Track, monitor, and remind patients of appointments	• Number of patients lost to follow-up • Number of patients whose eligibility status for the study changed after enrollment (e.g., pregnant, suicidal) • Show rate for primary care services • Show rate for behavioral health services	When a participant enrolls in the program, what happens to them next? Take me through the services and activities that an enrollee receives in the program.	None

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
		• Show rate for lifestyle classes/sessions at Mercy • Show rate for referral appointments (total and by type of service) • Number of clinic visits/follow-up visits received (total and by type of service) • Number of referrals created • Receipt of intervention by review of patient attendance		
What are the program activities and how have they been operationalized?	Activity: Diagnosis of diabetes, obesity and hypertension by the navigator.	• Record of vitalization of blood pressure, height, weight, and waist circumference • Record of blood test results for HbA1c • Record of patient treatment plan created	When a participant enrolls in the program, what happens to them next? Take me through the services and activities that an enrollee receives in the program.	None
What are the program activities and how have they been operationalized?	Activity: Administration of surveys to assess behavioral health and spirituality.	• Number of patients with all intake forms and assessments completed (e.g., PHQ-9, Duke Health Profile, etc.)	Since beginning enrollment, to what extent has the program been able to deliver all the program services that had been planned as part of the program intervention?	None

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
What are the program activities and how have they been operationalized?	Activity: Administer patient satisfaction surveys	• Patient satisfaction with Si Three (by type of service)	--	None
What are the program activities and how have they been operationalized?	Activity: Re-assess patients quarterly on physical health measures And semi-annually on behavioral health measures	--	Since beginning enrollment, to what extent has the program been able to deliver all the program services that had been planned as part of the program intervention?	Record of the number of times these assessments were completed and dates
Are the components different than what was planned? If so, why?	Output: Recruit 205 participants into each arm of the study	• Number of target participants—intervention and internal comparison groups • Number of patients screened for participation in the study • Number of patients consented to participate in the study • Number of patients who choose not to participate in the study • Number of patients enrolled in the program – intervention and internal comparison groups	--	None

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
Are the components different than what was planned? If so, why?	Output: Development of a patient care plan (including behavioral health treatment plans)	• Record of patient treatment plan created	Since beginning enrollment, to what extent has the program been able to deliver all the program services that had been planned as part of the program intervention?	None
Are the components different than what was planned? If so, why?	Output: Referral to one or more in-house services (exercise coach, educator, medical and/or faith-based behavioral health counselor) and/or community resources and chronic disease management	• Number of referrals created • Number of Si Three activities/services offered	Since beginning enrollment, to what extent has the program been able to deliver all the program services that had been planned as part of the program intervention?	None, assuming the number of Si Three activities/services offered is done at the individual level
INTEGRATION: What level of Integrated Behavioral Health did Sí Three achieve as a result of implementing the program?				
What level of Integrated Behavioral Health did Mercy Ministries achieve as a result of implementing the program?	IBH Level	Score (measured by IBH Checklist)	--	None
To what extent have providers and program staff adopted the components of the Sí Three program at 6 and 12 months?	Output: Increased understanding of integration	--	• Now that the program has been implemented, to what extent are primary care and behavioral health services connected, coordinated, combined, if at all?	Staff satisfaction/knowledge survey

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
What are the facilitators and barriers to adoption?	Output: Ongoing quality improvement among clinic staff Activity: Clinic-capacity building activities: Phone conferences, case conferences, face-to-face interactions, EPIC training, ongoing feedback and mentoring	--	• Please describe any barriers you or your organization has experienced in implementing the program. • In what ways did these barriers affect program implementation? In what ways have you been able to address these barriers? • Please describe anything that has helped your organization implement the program. • Probes: Is the staff, the facilities, the data systems, outside partners, or other things?	Staff/Administration satisfaction surveys
To what extent do providers buy-in to the program, and how has that buy-in affected implementation?	Activity: Administer staff satisfaction surveys Output: Provider and staff buy-in to model	--	• Have you heard any feedback from providers about program implementation? • What are some of the general themes from their feedback been?	Staff satisfaction surveys
To what extent did the comparison groups receive program-like components?				
--	--	• Number of patients in internal comparison	• When a participant is randomized/enrolled in	None

Research question/subquestions	**Logic Model Elements/Components** *What are we measuring to answer this research question?*	**Quantitative Indicator(s) Captured** *What data is being collected by subgrantee that we could use to capture this?*	**Qualitative Data** *What questions do we ask in our interview protocol to cover this? Do we need to augment our interview protocol to cover gaps?*	**Qualitative/quantitative Indicator(s) Needed** *If gap, what quantitative data do we need?*
		group that receive 1 program-like component • Number of patients in external comparison group that receive more than 1 program-like component	the control/comparison group of your program, what can they expect to receive or participate in terms of services or activities? • Since the program started, has anything changed about the services that control group participants received or activities they have access to at your clinic? In what way? • What do you see as the impact of this workflow change, if any? • Have these changes had any effects on patient care for those participants not enrolled in the study? In what way?	

Appendix G: Loss to Follow-Up/Attrition Tables

Table 69. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Demographic Characteristics among the Intervention and Primary Comparison Groups

	Full Sample (n=410)		Completed Study (n=293)		Did Not Complete Study (n=117)		p-value
Variables	N	%	N	%	N	%	
Sex							
Male	52	12.7	41	14.0	11	9.4	
Female	358	87.3	252	86.0	106	90.6	0.21
Missing	--	--	--	--	--	--	
Ethnicity							
Hispanic	409	99.8	292	99.7	117	100.0	
Non-Hispanic	1	0.2	1	0.3	0	0.0	0.53
Missing	--	--	--	--	--	--	
Age							
Mean	44.1	--	45.8	--	39.6	--	**<0.001**
SD	10.8	--	10.2	--	11.1	--	
18-24	16	3.9	4	1.4	12	10.3	
25-34	55	13.4	31	10.6	24	20.5	
35-44	147	35.9	104	35.5	43	36.8	
45-54	120	29.3	94	32.1	26	22.2	**<0.001**
55-64	67	16.3	56	19.1	11	9.4	
65+	5	1.2	4	1.4	1	0.9	
Missing	--	--	--	--	--	--	
Employment Status							
Employed	194	47.3	141	48.1	53	45.3	
Not Employed	216	52.7	152	51.9	64	54.7	0.61
Missing	--	--	--	--	--	--	
Marital Status							
Unmarried	191	46.7	163	44.4	61	52.6	
Married	218	53.3	130	55.6	55	47.4	0.13
Missing	--	--	--	--	--	--	
Primary Language							
English	50	12.2	29	9.9	21	17.9	
Spanish	360	87.8	264	90.1	96	82.1	**0.02**
Missing	--	--	--	--	--	--	
Smoking Status							
Current Smoker	38	9.2	22	7.5	16	13.6	
Former Smoker	14	3.4	9	3.1	5	4.3	0.12
Never Smoked	358	87.3	262	89.4	96	82.1	
Missing	--	--	--	--	--	--	
Alcohol Consumption							
Yes	83	20.7	53	18.5	30	26.1	0.09

	Full Sample (n=410)		Completed Study (n=293)		Did Not Complete Study (n=117)		p-value
Variables	N	%	N	%	N	%	
No	318	79.3	233	81.5	85	73.9	
Missing	9	--	7	--	2	--	
Spirituality Index							
Mean	48.9	--	49.3	--	47.9	--	
SD	12.1	--	11.8	--	11.8	--	0.29
Missing	--	--	--	--	--	--	

Table 70. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Demographic Characteristics among the Intervention and Secondary Comparison Groups

	Full Sample (n=573)		Completed Study (n=399)		Did Not Complete Study (n=174)		p-value
Measure	N	%	N	%	N	%	
Sex							
Male	131	22.9	84	21.1	47	27.0	
Female	442	77.1	315	79.0	127	73.0	0.12
Missing	--	--	--	--	--	--	
Ethnicity[a]							
Hispanic	570	99.5	397	99.5	173	99.4	
Non-Hispanic	3	0.5	2	0.5	1	0.6	0.91
Missing	--	--	--	--	--	--	
County of Residence[a]							
Cameron	1	0.2	1	0.3	0	0.0	
Hidalgo	365	63.7	256	64.2	109	62.6	0.80
Webb	207	36.1	142	35.6	65	37.4	
Missing	--	--	--	--	--	--	
Age							
18-24	18	3.1	6	1.5	12	6.9	
25-34	50	8.7	29	7.3	21	12.1	
35-44	158	27.6	115	28.8	43	24.7	**0.001**
45-54	180	31.4	129	32.3	51	29.3	
55-64	135	23.6	102	25.6	33	19.0	
65+	32	5.6	18	4.5	14	8.1	
Mean	47.8	--	48.4	--	46.3	--	
SD	11.9	--	11.0	--	13.6	--	0.07
Missing	--	--	--	--	--	--	
Employment Status							
Employed	222	38.7	149	37.3	73	42.0	
Not Employed	351	61.3	250	62.7	101	58.1	0.30

Missing	--	--	--	--	--	--	
Marital Status							
Unmarried	219	38.4	148	37.2	71	41.3	
Married	351	61.6	250	62.8	101	58.7	0.36
Missing	*3*	--	*1*	--	*2*	--	
Primary Language							
English	126	22.0	87	21.8	39	22.4	
Spanish	446	77.8	312	78.2	134	77.0	0.31
SL	1	0.2	0	0.0	1	0.6	
Missing	--	--	--	--	--	--	
Smoking Status							
Current Smoker	47	8.2	25	6.3	22	12.6	
Former Smoker	30	5.2	20	5.0	10	5.8	**0.03**
Never Smoked	496	86.6	354	88.7	142	81.6	
Missing	--	--	--	--	--	--	
Alcohol Consumption							
Yes	120	21.2	76	19.2	44	25.6	
No	447	78.8	319	80.8	128	74.4	0.09
Missing	*6*	--	*4*	--	*2*	--	

Table 71. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Demographic Characteristics among the Intervention Group

	Full Sample (n=207)		Completed Study (n=142)		Did Not Complete Study (n=65)		p-value
Variables	N	%	N	%	N	%	
Sex							
Male	27	13.0	20	14.1	7	58	
Female	180	87.0	122	85.9	10.8	89.2	0.51
Missing	--	--	--	--	--	--	
Ethnicity							
Hispanic	206	99.5	141	99.3	65	100.0	
Non-Hispanic	1	0.5	1	0.7	0	0.0	0.50
Missing	--	--	--	--	--	--	
Age							
Mean	43.8	--	46.3	--	38.2	--	**<0.001**
SD	11.3	--	10.6	--	11.1	--	
18-24	9	4.3	1	0.7	8	12.3	
25-34	29	14.0	13	9.2	16	24.6	
35-44	75	36.2	54	38.0	21	32.3	
45-54	61	29.5	45	31.7	16	24.6	**<0.001**
55-64	29	14.0	25	17.6	4	6.2	
65+	4	1.9	4	2.8	0	0.0	
Missing	--	--	--	--	--	--	

	Full Sample (n=207)		Completed Study (n=142)		Did Not Complete Study (n=65)		p-value
Variables	N	%	N	%	N	%	
Employment Status[a]							
Employed	89	43.0	60	42.3	29	44.6	
Not Employed	118	57.0	82	57.7	36	55.4	0.75
Missing	--	--	--	--	--	--	
Marital Status							
Unmarried	93	45.2	60	42.3	33	51.6	
Married	113	54.8	82	57.7	31	48.4	0.21
Missing	--	--	--	--	--	--	
Primary Language							
English	26	12.6	16	11.3	10	15.4	
Spanish	181	87.4	126	88.7	55	84.6	0.41
Missing	--	--	--	--	--	--	
Smoking Status[a]							
Current Smoker	23	11.1	13	9.2	10	15.4	
Former Smoker	7	3.4	4	2.8	3	4.6	0.31
Never Smoked	177	85.5	125	88.0	52	80.0	
Missing	--	--	--	--	--	--	
Alcohol Consumption							
Yes	43	21.4	25	68.7	18	28.6	
No	158	78.6	113	81.9	45	71.4	0.09
Missing	6	--	4	--	2	--	
Spirituality Index							
Mean	47.6	--	48.0	--	46.8	--	
SD	11.9	--	11.9	--	12.1	--	0.49
Missing	--	--	--	--	--	--	

[a]Fisher's Exact test was used due to cells having expected count less than 5

Table 72. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Demographic Characteristics among the Primary Comparison Group

	Full Sample (n=203)		Completed Study (n=151)		Did Not Complete Study (n=52)		p-value
Variables	N	%	N	%	N	%	
Sex							
Male	25	12.3	21	13.9	4	7.7	
Female	178	87.7	130	86.1	48	92.3	0.24
Missing	--	--	--	--	--	--	
Ethnicity							
Hispanic	203	100.0	151	100.0	52	100.0	--

	Full Sample (n=203)		Completed Study (n=151)		Did Not Complete Study (n=52)		p-value
Variables	N	%	N	%	N	%	
Non-Hispanic	0	0.0	0	0.0	0	0.0	
Missing	--	--	--	--	--	--	
Age							
Mean	44.3	--	45.4	--	41.3	--	**0.02**
SD	10.3	--	9.8	--	11.2	--	
18-24	7	3.5	3	2.0	4	7.7	
25-34	26	12.8	18	11.9	8	15.4	
35-44	72	35.5	50	33.1	22	42.3	
45-54	59	29.1	49	32.5	10	19.2	**0.048**
55-64	38	18.7	31	20.5	7	13.5	
65+	1	0.5	0	0.0	1	1.9	
Missing	--	--	--	--	--	--	
Employment Status							
Employed	105	51.7	81	53.6	24	46.2	
Not Employed	98	48.3	70	46.4	28	53.9	0.35
Missing	--	--	--	--	--	--	
Marital Status							
Unmarried	98	48.3	70	46.4	28	53.9	
Married	105	51.7	81	53.6	24	46.2	0.35
Missing	--	--	--	--	--	--	
Primary Language							
English	24	11.8	13	8.6	11	21.2	
Spanish	179	88.2	138	91.4	41	78.9	**0.02**
Missing	--	--	--	--	--	--	
Smoking Status[a]							
Current Smoker	15	7.4	9	6.0	6	11.5	
Former Smoker	7	3.4	5	3.3	2	3.9	0.40
Never Smoked	181	89.2	137	90.7	44	84.6	
Missing	--	--	--	--	--	--	
Alcohol Consumption							
Yes	40	20.0	28	18.9	12	23.1	
No	160	80.0	120	81.1	40	76.9	0.52
Missing	3	--	3	--	0	--	
Spirituality Index							
Mean	50.2	--	50.5	--	49.3	--	
SD	12.1	--	12.4	--	11.4	--	0.55
Missing	--	--	--	--	--	--	

[a]Fisher's Exact test was used due to cells having expected count less than 5

Table 73. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Demographic Characteristics among the Secondary Comparison Group

	Full Sample (n=366)		Completed Study (n=257)		Did Not Complete Study (n=109)		p-value
Measure	N	%	N	%	N	%	
Sex							
Male	104	28.4	64	24.9	40	36.7	
Female	262	71.6	193	75.1	69	63.3	**0.02**
Missing	--	--	--	--	--	--	
Ethnicity[a]							
Hispanic	364	99.5	256	99.6	108	99.1	
Non-Hispanic	2	0.5	1	0.4	1	0.9	0.51
Missing	--	--	--	--	--	--	
County of Residence[a]							
Cameron	1	0.3	1	0.4	0	0.0	
Hidalgo	365	99.7	256	99.6	109	100.0	0.99
Missing	--	--	--	--	--	--	
Age							
18-24	9	2.5	5	2.0	4	3.7	
25-34	21	5.7	16	6.2	5	4.6	
35-44	83	22.7	61	23.7	22	20.2	0.19
45-54	119	32.5	84	32.7	35	32.1	
55-64	106	29.0	77	30.0	29	26.6	
65+	28	7.7	14	5.5	14	12.8	
Mean	50.1	--	49.6	--	51.2	--	
SD	11.6	--	11.1	--	12.7	--	0.24
Missing	--	--	--	--	--	--	
Employment Status							
Employed	133	36.3	89	34.6	44	40.4	
Not Employed	233	63.7	168	65.4	65	59.6	0.30
Missing	--	--	--	--	--	--	
Marital Status							
Unmarried	126	34.6	88	34.4	38	35.2	
Married	238	65.4	168	65.6	70	64.8	0.88
Missing	--	--	--	--	--	--	
Primary Language[a]							
English	100	27.3	71	27.6	29	26.6	
Spanish	265	72.4	186	72.4	79	72.5	0.44
SL	1	0.3	0	0.0	1	0.9	
Missing	--	--	--	--	--	--	
Smoking Status							
Current Smoker	24	6.6	12	4.7	12	11.0	
Former Smoker	23	6.3	16	6.2	7	6.4	0.08
Never Smoked	319	87.2	229	89.1	90	82.6	
Missing	--	--	--	--	--	--	

Alcohol Consumption							
Yes	77	21.0	51	19.8	26	23.9	
No	289	79.0	206	80.2	83	76.2	0.39
Missing	--	--	--	--	--	--	

[a]Fisher's Exact test was used due to cells having expected count less than 5

Table 74. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Health Impact Measures among the Intervention and Primary Comparison Groups

	Full Sample (n=410) Mean (SD)	Completed Study (n=293) Mean (SD)	Did Not Complete Study (n=117) Mean (SD)	p-value
BMI	32.9 (6.6)	33.2 (6.5)	32.3 (6.8)	0.20
Systolic[b]	**124.6 (17.5)**	**126.1 (18.1)**	**120.8 (15.2)**	**0.01**
Diastolic[b]	**74.2 (9.8)**	**74.9 (10.0)**	**72.4 (9.0)**	**0.02**
Waist Circumference: Males	42.2 (5.1)	41.5 (3.8)	45.0 (8.3)	0.23
Waist Circumference: Females	43.6 (5.6)	44.0 (5.6)	42.8 (5.6)	0.07
General Health	71.1 (17.2)	71.2 (17.7)	70.8 (15.8)	0.81
Nonparametric Tests[a]	Median (SD)	Median (SD)	Median (SD)	*p*
PHQ-9	**4.0 (5.5)**	**4.0 (5.3)**	**5.0 (5.9)**	**0.03**
GAD-7	4.0 (5.3)	3.0 (5.1)	4.0 (5.6)	0.06
HbA1c	6.6 (1.9)	6.7 (1.7)	6.2 (2.3)	0.06
CAGE-AID	0.0 (0.6)	0.0 (0.6)	0.0 (0.6)	0.47

Note: Bold denotes statistical significance (p-value < 0.05)

[a] The Wilcoxon rank sum test was used to examine non-normally distributed data

[b] *Sample size for blood pressure is 292 for those who completed the study due to 1 participant missing data at 12-month follow-up*

Table 75. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Health Impact Measures among the Intervention and Secondary Comparison Group

	Full Sample (n=573) Mean (SD)	Completed Study (n=399) Mean (SD)	Did Not Complete Study (n=174) Mean (SD)	p-value
BMI	34.6 (7.5)	34.9 (7.4)	34.0 (7.6)	0.13
Systolic	**128.8 (18.9)**	**130.2 (18.9)**	**125.7 (18.6)**	**0.01**
Diastolic	**79.0 (10.6)**	**79.6 (10.6)**	**77.6 (10.6)**	**0.04**
Waist Circumference: Males	40.8 (6.2)	41.1 (6.4)	40.2 (6.1)	0.45
Waist Circumference: Females	41.5 (5.8)	41.7 (5.9)	40.9 (5.7)	0.21
Non-Parametric Tests[a]	Median (SD)	Median (SD)	Median (SD)	p-value
PHQ-9	2.0 (4.7)	2.0 (4.6)	3.0 (4.9)	0.67
General Health	80.0 (17.5)	80.0 (18.2)	80.0 (15.9)	0.46
GAD-7	1.0 (4.4)	1.0 (4.4)	2.0 (4.6)	0.50
HbA1c	6.7 (2.0)	6.9 (1.9)	6.3 (2.3)	0.24

Note: Bold denotes statistical significance (p-value < 0.05)

[a] The Wilcoxon rank sum test was used to examine non-normally distributed data; these results aligned with t test results.

Table 76. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Health Impact Measures among the Intervention Group

	Full Sample (n=207) Mean (SD)	Completed Study (n=142) Mean (SD)	Did Not Complete Study (n=65) Mean (SD)	p-value
BMI	33.2 (7.1)	33.7 (7.2)	32.1 (6.7)	0.12
Systolic	**125.3 (18.4)**	**128.2 (19.8)**	**119.0 (12.8)**	**0.001**
Diastolic	**74.9 (10.1)**	**76.1 (10.5)**	**72.1 (8.7)**	**0.01**
Waist Circumference: Males	41.5 (4.2)	41.0 (3.9)	42.9 (5.3)	0.35
Waist Circumference: Females	43.7 (6.0)	44.3 (6.0)	42.5 (5.7)	0.06
General Health	67.7 (17.5)	67.3 (18.3)	68.6 (15.9)	0.64
Nonparametric Tests[a]	Median (SD)	Median (SD)	Median (SD)	*p*
PHQ-9	5.0 (6.0)	5.5 (6.0)	5.0 (6.2)	0.13
GAD-7	5.0 (5.6)	5.0 (5.6)	5.0 (5.8)	0.27
HbA1c	6.5 (2.0)	6.7 (1.9)	6.2 (2.5)	0.49
CAGE-AID	0.0 (0.6)	0.0 (0.7)	0.0 (0.6)	0.16

Note: Bold denotes statistical significance (p-value < 0.05)

[a] The Wilcoxon rank sum test was used to examine non-normally distributed data

Table 77. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Health Impact Measures among the Primary Comparison Group

	Full Sample (n=203) Mean (SD)	Completed Study (n=151) Mean (SD)	Did Not Complete Study (n=52) Mean (SD)	p-value
BMI	32.5 (6.0)	32.5 (5.7)	32.5 (6.9)	0.84
Systolic[b]	123.9 (16.5)	124.1 (16.1)	123.1 (17.6)	0.69
Diastolic[b]	73.5 (9.5)	73.7 (9.5)	72.7 (9.50	0.48
Waist Circumference: Males	43.0 (6.0)	41.9 (3.7)	48.0 (11.9)	0.38
Waist Circumference: Females	43.5 (5.1)	43.7 (5.2)	43.2 (5.1)	0.54
General Health	74.5 (16.2)	74.9 (16.5)	73.5 (15.4)	0.59
Nonparametric Tests[a]	Median (SD)	Median (SD)	Median (SD)	*p*
PHQ-9	3.0 (4.6)	3.0 (4.2)	3.0 (5.4)	0.11
GAD-7	3.0 (4.7)	2.0 (4.4)	4.0 (5.3)	0.07
HbA1c	6.7 (1.6)	6.7 (1.5)	6.0 (2.1)	0.05
CAGE-AID	0.0 (0.6)	0.0 (0.6)	0.0 (0.6)	0.93

Note: Bold denotes statistical significance (p-value < 0.05)p-value

[a] The Wilcoxon rank sum test was used to examine non-normally distributed data

[b] *Sample size for blood pressure is 150 for those who completed the study due to 1 participant missing data at 12-month follow-up*

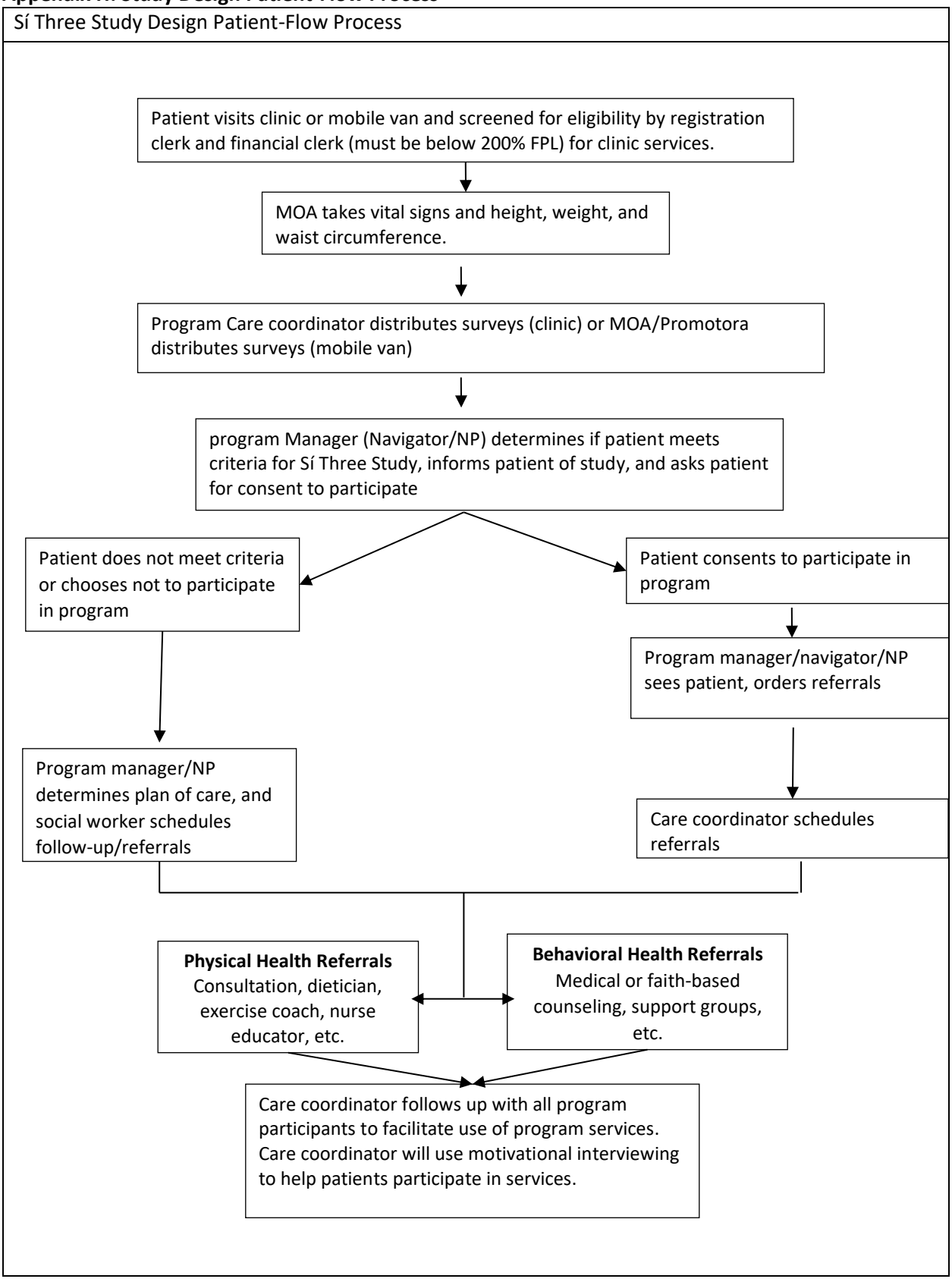
Table 78. Analysis of Participants Who Completed the Study Compared to Lost to Follow-up on Health Impact Measures among the Secondary Comparison Group

	Full Sample (n=366) Mean (SD)	Completed Study (n=257) Mean (SD)	Did Not Complete Study (n=109) Mean (SD)	p-value
BMI	35.4 (7.6)	35.5 (7.5)	35.1 (7.8)	0.53
Systolic	130.8 (18.9)	131.3 (18.3)	129.7 (20.4)	0.47
Diastolic	81.3 (10.2)	81.5 (10.1)	80.8 (10.3)	0.59
Waist Circumference – Males	40.6 (6.6)	41.1 (6.9)	39.8 (6.2)	0.34
Waist Circumference – Females	40.1 (5.4)	40.3 (5.3)	39.8 (5.5)	0.54
Non-Parametric Tests[a]	Median (SD)	Median (SD)	Median (SD)	p-value
PHQ-9	1.0 (2.5)	1.0 (2.6)	1.0 (2.3)	0.61
General Health	86.7 (15.4)	86.7 (16.0)	83.3 (14.0)	0.30
GAD-7	0.0 (4.7)	0.0 (2.3)	0.0 (1.9)	0.83
HbA1c	6.8 (2.0)	7.0 (1.9)	6.4 (2.3)	0.33

Note: Bold denotes statistical significance (p-value < 0.05)p-value

[a] The Wilcoxon rank sum test was used to examine non-normally distributed data; these results aligned with t test results.

Appendix H: Study Design Patient-Flow Process



Appendix I: Study Group Comparison Table

For additional clarity, the table below outlines the practices and services for the intervention group (Sí Three program participants), primary comparison group (Mercy Ministries' usual care), and secondary comparison group (Edcouch clinic).

Practices and services the intervention group, primary comparison group and secondary comparison group receives			
	Intervention group	**Primary comparison group**	**Secondary comparison group**
Impact measure survey instruments • PHQ-9 • GAD-7 • CAGE-AID • Duke Health Profile • Spirituality Index of Well-being	Participants surveyed at baseline, 3, 6, 9 and 12 months	Participants surveyed at baseline, 6 months and 12 months	Participants surveyed at baseline, 6 months and 12 months *Note, participants will not receive the Spirituality Index of Well-being or CAGE-AID.*
Patient satisfaction survey	Surveys completed at 6 months and 12 months	Not administered to this group	Not administered to this group
Care coordination	Care Coordinator contacts and schedules all follow-up appointments and keeps in contact with all study participants	None	None
Healthcare provider visit	Scheduled by participant or care coordinator at baseline, 3 months, 6 months, 9 months, and 12 months	Scheduled by participant or social worker as ordered by NP	Scheduled by participant or ordered by MD
Referrals to behavioral health	Care Coordinator or NP/navigator schedules with LPC/Pastoral Counselor	RN schedules appointment with 2 part-time counselors (current wait list)	Edcouch staff schedules an appointment with LPC in Mercedes. Alton participants see LPC at that clinic if LPC is available. If not, Alton staff schedules an appointment with LPC in Mercedes.

Referrals to health education and nutrition educator	Care Coordinator schedules session with MSN Nurse Educator	RN schedules class at Mercy with BSN Nurse Educator	Scheduled to see a trained employee (Non-nurse)
Referral to exercise class	Care Coordinator schedules exercise class with Exercise Coach	Community Class (led by community contracted person twice-a-week non-scheduled at will	None

Appendix J: Explanation of Eligibility Criteria for the Shared-Comparison Sites

As mentioned in the SEP three Sí Texas SIF evaluations were sharing a pool of patients at two comparison clinics to draw a comparison group sample. Three interventions— NCDV, UT-RGV, and Mercy Clinic—proposed to construct their comparison group sample using data collected from study participants at the NCDV Edcouch and Alton Clinics. The SEP proposed that propensity score matching would be used to construct a secondary (external) comparison group. Due to near completeness of data at both Mercy and the NCDV Edcouch clinics and an insufficient number of variables to match on between the primary and secondary comparison groups, propensity score matching was not used to construct the secondary comparison group. To construct the secondary comparison group for the Sí Three program study, all Edcouch participants who met the Mercy Sí Three program eligibility criteria were included in that group.

Eligibility Criteria for the Secondary Comparison and Intervention Groups by Clinic

Group	Secondary Comparison Group[1]		Intervention Site		
Clinic	Edcouch	Alton	NCDV (all of the following)	UT-RGV (any one of the following)	Mercy (any one of the following)
County	Any county in the Rio Grande Valley	Any county in the Rio Grande Valley	Hidalgo or Starr	Not an eligibility criterion	Not an eligibility criterion
A1c	≥6.5	≥6.5	≥6.5	≥6.5	>7.0
BMI	Any value[2]	Any value[2]	≥30.0	≥30.0	>30.0
PHQ-9 score	Any score[2]	Any score[2]	Not an eligibility criterion	≥5	>5
Blood pressure	Any blood pressure[2]	Any blood pressure[2]	Not an eligibility criterion	140/90 or higher	>140/90
GAD-7	Any score[2]	Not collected	Not an eligibility criterion	5 or higher	> 5
Waist Circumference	Any value[2]	Not collected	Not an eligibility criterion	Not an eligibility criterion	>40 in (men) > 35 in (women)
Cage-AID	Not collected	Not collected	Not an eligibility criterion	Not an eligibility criterion	>2

[1] The Edcouch and Alton Clinics eligibility criteria requires patients to have **one or more** of the listed eligibility criteria.

[2] This is not an explicit eligibility criterion, rather it is a measure that will be collected on all participants and the Alton and Edcouch Clinics.

Appendix K: Patient-Centered Integrated Behavioral Health Care Checklist

Patient-Centered Integrated Behavioral Health Care Principles & Tasks

AIMS CENTER
Advancing Integrated Mental Health Solutions

About This Tool

This checklist was developed in consultation with a group of national experts (http://bit.ly/IMHC-experts) in integrated behavioral health care with support from The John A. Hartford Foundation, The Robert Wood Johnson Foundation, Agency for Healthcare Research and Quality, and California HealthCare Foundation. For more information, visit: http://bit.ly/IMHC_principles.

The core principles of effective integrated behavioral health care include a patient-centered care team providing evidence-based treatments for a defined population of patients using a measurement-based treat-to–target approach.

Principles of Care	We apply this principle in the care of ... of our patients: None	Some	Most/All
1. Patient-Centered Care			
Primary care and behavioral health providers collaborate effectively using shared care plans.	☐	☐	☐
2. Population-Based Care			
Care team shares a defined group of patients tracked in a registry. Practices track and reach out to patients who are not improving and mental health specialists provide caseload-focused consultation, not just ad-hoc advice.	☐	☐	☐
3. Measurement-Based Treatment to Target			
Each patient's treatment plan clearly articulates personal goals and clinical outcomes that are routinely measured. Treatments are adjusted if patients are not improving as expected.	☐	☐	☐
4. Evidence-Based Care			
Patients are offered treatments for which there is credible research evidence to support their efficacy in treating the target condition.	☐	☐	☐
5. Accountable Care			
Providers are accountable and reimbursed for quality care and outcomes.	☐	☐	☐

Page 2

Core components and tasks are shared by effective integrated behavioral health care programs. The AIMS Center Integrated Care Team Building Tool (http://bit.ly/IMHC-teambuildingtool) can help organizations build clinical workflows that incorporate these core components and tasks into their unique setting.

Core Components & Tasks	None of our patients receive this service	Some of our patients receive this service	Most/All of our patients receive this service
1. Patient Identification and Diagnosis			
Screen for behavioral health problems using valid instruments	☐	☐	☐
Diagnose behavioral health problems and related conditions	☐	☐	☐
Use valid measurement tools to assess and document baseline symptom severity	☐	☐	☐
2. Engagement in Integrated Care Program			
Introduce collaborative care team and engage patient in integrated care program	☐	☐	☐
Initiate patient tracking in population-based registry	☐	☐	☐
3. Evidence-Based Treatment			
Develop and regularly update a biopsychosocial treatment plan	☐	☐	☐
Provide patient and family education about symptoms, treatments, and self management	☐	☐	☐
Provide evidence-based counseling (e.g., Motivational Interviewing, Behavioral Activation)	☐	☐	☐
Provide evidence-based psychotherapy (e.g., Problem Solving Treatment, Cognitive Behavior Therapy, Interpersonal Therapy)	☐	☐	☐
Prescribe and manage psychotropic medications as clinically indicated	☐	☐	☐
Change or adjust treatments if patients do not meet treatment targets	☐	☐	
4. Systematic Follow-up, Treatment Adjustment, and Relapse Prevention			
Use population-based registry to systematically follow all patients	☐	☐	☐
Proactively reach out to patients who do not follow-up	☐	☐	☐
Monitor treatment response at each contact with valid outcome measures	☐	☐	☐
Monitor treatment side effects and complications	☐	☐	☐
Identify patients who are not improving to target them for psychiatric consultation and	☐	☐	☐
Create and support relapse prevention plan when patients are substantially improved	☐	☐	
5. Communication and Care Coordination			
Coordinate and facilitate effective communication among providers	☐	☐	☐
Engage and support family and significant others as clinically appropriate	☐	☐	☐
Facilitate and track referrals to specialty care, social services, and community-based resources	☐	☐	☐
6. Systematic Psychiatric Case Review and Consultation			
Conduct regular (e.g., weekly) psychiatric caseload review on patients who are not improving	☐	☐	☐
Provide specific recommendations for additional diagnostic work-up, treatment changes, or	☐	☐	☐
Provide psychiatric assessments for challenging patients in-person or via telemedicine	☐	☐	☐
7. Program Oversight and Quality Improvement			
Provide administrative support and supervision for program	☐	☐	
Provide clinical support and supervision for program	☐	☐	
Routinely examine provider- and program-level outcomes (e.g., clinical outcomes, quality of care, patient satisfaction) and use this information for quality improvement	☐	☐	

Appendix L: Patient Health Questionnaire – 9 (PHQ-9)

PATIENT HEALTH QUESTIONNAIRE-9 (PHQ-9)

Over the last 2 weeks, how often have you been bothered by any of the following problems? *(Use "✔" to indicate your answer)*	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9. Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

FOR OFFICE CODING 0 +_____+_____+ _____

=Total Score: _____

If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all D | Somewhat difficult D | Very difficult D | Extremely difficult D

Developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues, with an educational grant from Pfizer Inc. No permission required to reproduce, translate, display or distribute.

Appendix M: Generalized Anxiety Disorder – 7 (GAD – 7)

Generalized Anxiety Disorder 7-item (GAD-7) scale

Over the last 2 weeks, how often have you been bothered by the following problems?	Not at all sure	Several days	Over half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it's hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid as if something awful might happen	0	1	2	3
Add the score for each column	+	+	+	
Total Score *(add your column scores)* =				

If you checked off any problems, how difficult have these made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all _____
Somewhat difficult _____
Very difficult _____
Extremely difficult _____

Source: Spitzer RL, Kroenke K, Williams JBW, Lowe B. A brief measure for assessing generalized anxiety disorder. *Arch Inern Med.* 2006;166:1092-1097.

Appendix N: CAGE – AID

The CAGE Questionnaire Adapted to Include Drugs (CAGE-AID)

When thinking about drug use, include illegal drug use and the use of prescription drug use other than prescribed.

	YES	NO
1) Have you felt you ought to cut down on your drinking or drug use?	☐	☐
2) Have people annoyed you by criticizing your drinking or drug use?	☐	☐
3) Have you felt bad or guilty about your drinking or drug use?	☐	☐
4) Have you ever had a drink or used drugs first thing in the morning to steady your nerves or to get rid of a hangover (eye-opener)?	☐	☐

SCORE_____

CAGE Source: Brown, RL, Rounds, LA. Conjoint screening questionnaires for alcohol and other drug abuse: Criterion validity in a primary care practice. *Wisconsin Medical Journal*. (1995) 94 (3) 135-140.

Appendix O: Spirituality Index of Wellbeing

Spirituality Index of Well-Being

Which statement best describes your feelings and choices?

Strongly Agree 1	Agree 2	Neither Agree nor Disagree 3	Disagree 4	Strongly Disagree 5

1)	There is not much I can do to help myself.	1	2	3	4	5
2)	Often, there is no way I can complete what I have started.	1	2	3	4	5
3)	I can't begin to understand my problems.	1	2	3	4	5
4)	I am overwhelmed when I have personal difficulties and problems.	1	2	3	4	5
5)	I don't know how to begin to solve my problems.	1	2	3	4	5
6)	There is not much I can do to make a difference in my life.	1	2	3	4	5
7)	I have not found my life's purpose yet.	1	2	3	4	5
8)	I don't know who I am, where I came from, or where I am going.	1	2	3	4	5
9)	I have a lack of purpose in my life.	1	2	3	4	5
10)	In this world, I don't know where I fit in.	1	2	3	4	5
11)	I am far from understanding the meaning of life.	1	2	3	4	5
12)	There is a great void in my life at this time.	1	2	3	4	5

SCORE_____

Source: Daaleman TP, Frey BB. The spirituality index of well-being: a new instrument for health-related quality of life research. *Ann Fam Med*. 2004; 2:499-503.

Appendix P: Duke Health Profile

FORM A: FOR SELF-ADMINISTRATION BY THE RESPONDENT (revised 4-2000)

DUKE HEALTH PROFILE (The DUKE)

Date Today:_____ Name: _____ ID Number:_____
Date of Birth:_____ Female____ Male____

INSTRUCTIONS: Here are some questions about your health and feelings. Please read each question carefully and check (√) your best answer. You should answer the questions in your own way. There are no right or wrong answers. (Please ignore the small scoring numbers next to each blank.)

		Yes, describes me exactly	Somewhat describes me	No, doesn't describe me at all
1.	I like who I am	12	11	10
2.	I am not an easy person to get along with	20	21	22
3.	I am basically a healthy person	32	31	30
4.	I give up too easily	40	41	42
5.	I have difficulty concentrating	50	51	52
6.	I am happy with my family relationships	62	61	60
7.	I am comfortable being around people	72	71	70

TODAY would you have any physical trouble or difficulty:

		None	Some	A Lot
8.	Walking up a flight of stairs	82	81	80
9.	Running the length of a football field	92	91	90

DURING THE PAST WEEK: How much trouble have you had with:

		None	Some	A Lot
10.	Sleeping	102	101	100
11.	Hurting or aching in any part of your body	112	111	110
12.	Getting tired easily	122	121	120
13.	Feeling depressed or sad	132	131	130
14.	Nervousness	142	141	140

DURING THE PAST WEEK: How often did you:

		None	Some	A Lot
15.	Socialize with other people (talk or visit with friends or relatives)	150	151	152
16.	Take part in social, religious, or recreation activities (meetings, church, movies, sports, parties)	160	161	162

DURING THE PAST WEEK: How often did you:

		None	1-4 Days	5-7 Days
17.	Stay in your home, a nursing home, or hospital because of sickness, injury, or other health problem.	172	171	170

MANUAL SCORING FOR THE DUKE HEALTH PROFILE

Item		Raw Score*
8	=	_____
9	=	_____
10	=	_____
11	=	_____
12	=	_____
Sum	=	_____ x 10 =

PHYSICAL HEALTH SCORE ☐

Item		Raw Score*
1	=	_____
4	=	_____
5	=	_____
13	=	_____
14	=	_____
Sum	=	_____ x 10 =

MENTAL HEALTH SCORE ☐

Item		Raw Score*
2	=	_____
6	=	_____
7	=	_____
15	=	_____
16	=	_____
Sum	=	_____ x 10 =

SOCIAL HEALTH SCORE ☐

GENERAL HEALTH SCORE

Physical Health score	=	_____
Mental Health score	=	_____
Social Health score	=	_____
Sum	=	_____ ÷ 3 =

☐

PERCEIVED HEALTH SCORE

Item		Raw Score*
3	=	_____ x 50 =

☐

Item		Raw Score*
1	=	_____
2	=	_____
4	=	_____
6	=	_____
7	=	_____
Sum	=	_____ x 10 =

SELF-ESTEEM SCORE ☐

To calculate the scores in this column the raw scores must be revised as follows:
If 0, change to 2; if 2, change to 0; if 1, no change.

Item		Raw Score*	Revised
2	=	_____	_____
5	=	_____	_____
7	=	_____	_____
10	=	_____	_____
12	=	_____	_____
14	=	_____	_____
Sum	=	_____	x 8.333 =

ANXIETY SCORE ☐

Item		Raw Score*	Revised
4	=	_____	_____
5	=	_____	_____
10	=	_____	_____
12	=	_____	_____
13	=	_____	_____
Sum	=	_____	x 10 =

DEPRESSION SCORE ☐

Item		Raw Score*	Revised
4	=	_____	_____
5	=	_____	_____
7	=	_____	_____
10	=	_____	_____
12	=	_____	_____
13	=	_____	_____
14	=	_____	_____
Sum	=	_____	x 7.143 =

ANXIETY-DEPRESSION (DUKE-AD) SCORE ☐

PAIN SCORE

Item		Raw Score*	Revised
11	=	_____	_____ x 50 =

☐

DISABILITY SCORE

Item		Raw Score*	Revised
17	=	_____	_____ x 50 =

☐

* Raw Score = last digit of the numeral adjacent to the blank checked by the respondent for each item. For example, if the second blank is checked for item 10 (blank numeral = 101), then the raw score is "1", because 1 is the last digit of 101.

Final Score is calculated from the raw scores as shown and entered into the box for each scale. For physical health, mental health, social health, general health, self-esteem, and perceived health, 100 indicates the best health status, and 0 indicates the worst health status. For anxiety, depression, anxiety-depression, pain, and disability, 100 indicates the worst health status and 0 indicates the best health status.

Missing Values: If one or more responses is missing within one of the eleven scales, a score cannot be calculated for that particular scale.

SPANISH (UNITED STATES) FORMULARIO A: PARA AUTO-ADMINISTRACIÓN POR LA PERSONA QUE RESPONDE (revisado 4-2000)

PERFIL DE SALUD DE DUKE (El Duke)

Fecha de hoy:_____ Nombre: _____Número de identificación: _____

Fecha de nacimiento: _____ Sexo: Femenino ☐ Masculino ☐

INSTRUCCIONES: Estas son algunas preguntas sobre su salud y sus sentimientos. Por favor, lea cada pregunta cuidadosamente y marque (✓) la respuesta más apropiada para usted. Usted debe contestar las preguntas a su manera. No hay respuestas correctas ni incorrectas. (Por favor, ignore los pequeños números al lado de cada línea).

	Sí, me Describe exactamente	Me describe más o menos	No, no me describe de ninguna manera
1. Me gusta quien soy	12	11	10
2. No me llevo bien con otros fácilmente	20	21	22
3. Soy básicamente una persona saludable	32	31	30
4. Me doy por vencido(a) muy fácilmente	40	41	42
5. Tengo dificultad en concentrarme	50	51	52
6. Yo estoy contento(a) con mis relaciones familiares	62	61	60
7. Me siento cómodo(a) alrededor de otras personas	72	71	70

¿Tendría HOY alguna dificultad o problema físico:	Ninguna	Alguna	Mucha
8. Al subir un tramo de escaleras?	82	81	80
9. Al correr la distancia de un campo de fútbol americano (100 yardas / 91 metros)?	92	91	90

DURANTE LA ÚLTIMA SEMANA: ¿Cuánta dificultad ha tenido con:	Ninguna	Alguna	Mucha
10. Dormir?	102	101	100
11. Dolor en alguna parte de su cuerpo?	112	111	110
12. Cansarse fácilmente?	122	121	120
13. Sentirse deprimido(a) o triste?	132	131	130
14. Nerviosismo?	142	141	140

DURANTE LA ÚLTIMA SEMANA: ¿Con qué frecuencia:	No, en absoluto	A veces	Muchas veces
15. Pasó tiempo con otras personas (por ejemplo, hablar o visitar con amigos o parientes)?	160	151	152
16. Participó en actividades sociales, religiosas, o recreativas (por ejemplo, reuniones, iglesia, cine, deportes, fiestas)?	160	161	162

DURANTE LA ÚLTIMA SEMANA: ¿Con qué frecuencia:	No, en absoluto	1-4 días	5-7 días
17. Se quedó en su casa, en la casa de ancianos, o en el hospital debido a enfermedad, lesión, o cualquier otro problema de salud?	172	171	170

173

Appendix Q: Patient Satisfaction Survey

Satisfaction Survey Navigator (Nurse Practitioner)

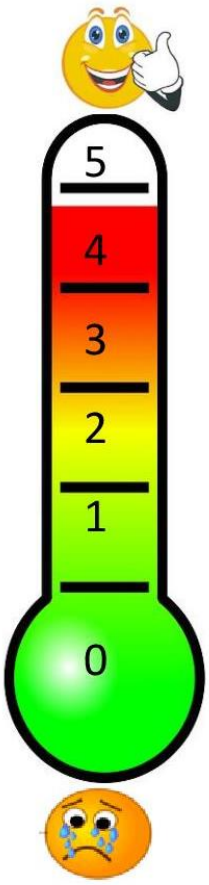
(Circle YOUR choice)

1. When I needed an appointment, I could get one soon enough.
   0 1 2 3 4 5

2. The staff was friendly to me.
   0 1 2 3 4 5

3. The wait for services was not too long.
   0 1 2 3 4 5

4. After talking to my Nurse Practitioner, I know why and how to take my medication(s).
   0 1 2 3 4 5



5. After talking to my Nurse Practitioner, I know something I can do to make myself healthier.
   0 1 2 3 4 5

6. I am going to be able to do the things the Nurse Practitioner explained to me at home.
   0 1 2 3 4 5

7. The Nurse Practitioner answered any questions I had.
   0 1 2 3 4 5

8. Overall, I was satisfied with my care at Mercy Clinic.
   0 1 2 3 4 5

Social Innovation Funded program

Appendix R: Satisfaction Survey Results

Table 79. Patient Satisfaction - Nurse Practitioner

Question	Mid-point			Endpoint		
	# of responses	# responding 4 or 5*	% responding 4 or 5	# of responses	# responding 4 or 5	% responding 4 or 5
When I needed an appointment, I could get in one soon enough.	90	89	99	78	75	96
The staff was friendly to me.	90	90	100	78	76	97
The wait for services was not too long.	90	83	92	78	70	90
After talking to my Nurse Practitioner, I know why and how to take my medication(s).	90	88	98	78	75	96
After talking to my Nurse Practitioner, I know something I can do to make myself healthier.	90	89	99	78	74	95
I am going to be able to do the things the Nurse Practitioner explained to me at home.	90	90	100	78	74	95
The Nurse Practitioner answered questions I had.	90	90	100	78	74	95
Overall, I was satisfied with my care at Mercy Clinic.	90	90	100	78	75	96

* Note: 4 generally corresponds with satisfied and 5 corresponds with highly satisfied. Mercy used a pictoral scale for patient satisfaction forms.

Table 80. Patient Satisfaction - Educator

Question	Mid-point			Endpoint		
	# of responses	# responding 4 or 5	% responding 4 or 5	# of responses	# responding 4 or 5	% responding 4 or 5
When I needed a class, I could get in one soon enough.	92	83	90	21	20	95
The staff was friendly to me.	92	87	95	21	21	100
The wait for services was not too long.	92	68	86	21	20	95
After talking to my Nurse Educator, I know why and how to take my medication(s).	92	85	92	21	21	100
After talking to my Nurse Educator, I know something I	92	88	96	21	21	100

can do to make myself healthier.						
I am going to be able to do the things the Nurse Educator and I planned for me to do.	92	80	93	21	21	100
The Nurse Educator answered questions I had.	92	87	95	21	21	100
Overall, I was satisfied with my care at Mercy Clinic.	92	88	96	21	21	100

Table 81. Patient Satisfaction - Behavioral Health Consultant

Question	Mid-point			Endpoint		
	# of responses	# responding 4 or 5	% responding 4 or 5	# of responses	# responding 4 or 5	% responding 4 or 5
When I needed an appointment, I could get in one soon enough.	16	14	88	32	32	100
The staff was friendly to me.	16	14	88	32	32	100
The wait for services was not too long.	16	14	88	32	30	94
After talking to my LPC/Behavioral Counselor, I know taking my medication is important.	16	13	81	32	32	100
After talking to my LPC/Behavioral Health Counselor, I am able to make changes in my life-style choices.	16	14	88	32	31	97
I am going to be able to identify things that trigger a change in my behavior.	16	13	81	32	32	100
The LPC/Behavioral Counselor answered questions I had.	16	14	88	32	32	100
Overall, I was satisfied with my care at Mercy Clinic.	16	14	88	32	32	100

Table 82. Patient Satisfaction - Exercise Coach

Question	Mid-point*		
	Number of responses	Number responding 4 or 5	% Responding 4 or 5
When I needed a class, I could get in one soon enough.	24	21	88
The staff was friendly to me.	24	22	92
The wait for services was not too long.	24	21	88
After talking to my Exercise Coach, I know how to do my exercise(s) and take my pulse.	24	21	88
After talking to my Exercise Coach I can do my exercises to make myself healthier.	24	21	88
I am going to be able to exercise like the Exercise Coach showed me at home.	24	20	83
The Exercise Coach answered questions I had.	24	22	92
Overall, I was satisfied with my care at Mercy Clinic.	24	22	92

*Endpoint data not presented (n<5)

Patient Satisfaction - Spiritual Counselor

- Data not presented (n<5), but overall 100% satisfaction with the spiritual counselor services.

Table 83. Staff Satisfaction - Clinic Staff

Question	Mid-point			Endpoint		
	# of responses	# responding 4 or 5	% Responding 4 or 5	# of responses	# responding 4 or 5	% Responding 4 or 5
When I need to ask a question, I know who to ask and understand the response.	28	27	96	27	26	96
The provider and other staff are friendly to me and reflect the Mercy values.	28	27	96	27	26	96
The patients do not have to wait a long time to be seen by the provider(s).	28	13	46	26	16	62
Patients understand their plan of care.	28	18	72	24	21	88
After talking to the patient, treatment plans are carried-out and appointments made.	28	21	88	22	22	100
I can go to the patient record (EPIC) and find the	28	22	96	22	20	91

documentation and follow-up of the patient's treatment plan.						
There is evidence the patient understands the treatment plan as ordered by the Nurse Practitioner.	28	16	67	23	22	96
Overall, I am satisfied with the care patients receive at Mercy.	28	25	89	26	25	96