John Snow, Inc.

IOPCM ROI Analysis: Phase II
Outcomes Evaluation of IOPCM Program with Additional Analyses

Prepared for Partnership HealthPlan of California

Janauary 27, 2017

Table of Contents

Introduction

Some of the biggest challenges facing health care today include rapidly escalating health care costs, fragmented health care delivery, and overuse and/or misuse of the health care system. In this environment, the need for innovative approaches to health care delivery is more important than ever. Since September 2012, Partnership HealthPlan of California (PHC), a non-profit community-based health care organization providing Medi-Cal coverage for over 500,000 members across 14 Northern California counties, has invested in piloting an innovative Intensive Outpatient Case Management (IOPCM) program. IOPCM aims to improve care coordination and management, patient experience, and quality, while reducing total cost of care for the highest cost members. PHC works with federally qualified health centers (FQHCs) to deliver IOPCM, using one of three delivery model types: (1) a facility-based nurse coordinating care centrally for a panel of patients, (2) a facility-based nurse distributing time across care teams, or (3) an externally contracted, off-site care manager.

PHC's early analyses of the first IOPCM cohort had promising yet inconclusive results due to limitations of small sample size and concerns about the lack of an appropriate comparison group. Nevertheless, PHC opted to continue the program at existing sites and to expand it to several additional sites starting in October 2013. PHC's investment in IOPCM comes at a timely moment. The California Department of Health Care Services (DHCS) is planning to roll out the Health Homes for Patients with Complex Needs program for Medi-Cal and dual-eligible beneficiaries in most PHC counties and San Francisco beginning in July 2017, and in 18 other counties (2 of which are PHC counties) in two subsequent phases. Given this phased roll out and a limited term of enhanced federal match for the program, lessons learned from IOPCM can inform the implementation and evaluation of the State's Health Homes program.

In September 2015, PHC contracted with John Snow, Inc. (JSI) to conduct a two-phase return on investment (ROI) analysis of IOPCM. In Phase I, JSI conducted descriptive statistics on sample claims data and a literature review to identify rigorous and defensible approaches to identify control groups. At the end of Phase I, JSI recommended propensity score matching followed by a difference-in-difference analysis as the evaluation design. This method is rigorous and realistic given real-world data availability, and is increasingly being used in case management evaluations and studies of health care spending and utilization that use claims data. After consulting with PHC, JSI moved into Phase II of the work and applied the agreed upon methods to conduct an outcomes evaluation of PHC's IOPCM program. This report presents a summary of key findings from JSI's outcomes evaluation analysis.

Summary of Key Findings

Below, we summarize key findings that emerged from our analysis.

1) **The propensity score method was effective in identifying an appropriate control group.** Applying propensity score-based matching improved the distribution of baseline factors (e.g., demographics and other factors hypothesized to influence receipt of intervention) between intervention patients and controls. In other words, matching was effective in minimizing baseline differences between eligible controls and intervention patients. This made the two study groups more similar, mitigating concerns of selection bias and increasing the confidence that observed intervention effects can be attributed to the intervention and not to baseline differences between the two groups. Matches were successfully identified for 73% of the intervention sample and all subsequent analyses were conducted on the matched cohort. All results should be interpreted with caution, however, since they could be influenced by unmeasured confounders for which propensity score matching does not control.

2) **Both intervention and control patients experienced a downward shift in utilization with a notable differential in changes in the second year**. By the start of the second intervention year, IOPCM patients showed a downward shift in inpatient admissions and ED visits and a gradual plateau in outpatient visits. Importantly, though both control and intervention patients experienced downward shifts in inpatient and ED utilization over time, intervention patients experienced more dramatic shifts that were sustained over time, while controls' utilization tended to fluctuate. For example, despite intervention patients having significantly higher ED visits per thousand member years (PTMY) compared to controls in the pre-intervention period, intervention patients' ED visits PTMY declined over time and were lower than controls' ED visits during the second year—a trend that continued over time.

3) **Differential utilization patterns for some subgroups of patients suggest that absence of psychotic illness may make patients more susceptible to intervention impact, especially over time**. The IOPCM intervention may have a differential impact on patients based on their diagnoses. For example, we examined utilization patterns for two discrete cohorts of intervention and control patients: those who had never had a psychotic mental illness diagnosis in the study period and those who had been diagnosed at least once with a psychotic mental illness during the study period. Our analyses showed that among patients **not** ever diagnosed with psychotic mental illness intervention, ED visits PTMY for intervention patients, although higher at the pre 0-6 month period, steadily declined in the post-intervention period and remained down even at 30 months. In contrast, ED visits PTMY for control patients **not** ever diagnosed with psychotic mental illness fluctuated, going down in the first year but then rising back up in the second year and remaining on the upward trend. We also found that among patients **ever** diagnosed with a psychotic mental health condition, inpatient admissions PTMY fluctuated for both intervention and controls. Importantly, among patients **not** ever diagnosed with psychotic mental illness, inpatient admissions PTMY for intervention patients were comparable to that of controls at most six-month time periods and dropped below that of controls in the 19-24 month period, remaining on the downward trend. In contrast, inpatient admissions PTMY for control patients started

to increase in the second year. These findings, although not statistically significant, suggest that a lack of psychotic illness may make patients more susceptible to the impact of the intervention, especially over time, while the presence of psychotic illness may delay the intervention impact on patients' utilization and may make patients' more prone to rebounds in utilization even when the intervention is having a net effect.

4) **Intervention effect on utilization was not statistically significant but is directionally meaningful, especially over time.** We conducted difference-in-difference analysis using multivariable regression to assess whether receiving the intervention is associated with utilization after controlling for factors hypothesized and/or tested in previous studies to be associated with utilization outcomes of interest. Our analyses shows that after controlling for potential confounders, the difference-in-differences or the difference between the rate of change in utilization for intervention patients compared to the rate of change in utilization for controls showed directionally meaningful, although not statistically significant, intervention effects. For inpatient admissions, a favorable intervention effect was seen as early as Year 1 post intervention: being in the intervention for one year is associated with 168 fewer inpatient admissions PTMY for intervention patients compared to controls after controlling for baseline inpatient admissions (0-12 months), risk scores, model delivery type, and aid categories. For ED visits, a favorable intervention effect was seen by Year 2: being in the intervention for two years was associated with 1039 fewer ED visits PTMY for interventions compared to controls after controlling for ED visits in the pre-0-6 month period, risk scores, model delivery type, and aid categories. Ultimately, for both the utilization and cost evaluations, a larger cohort followed over a longer period of time would be needed to demonstrate statistically significant results.

5) **Savings from cost avoidance, although not immediate, may accumulate over time.** Financial modeling applying cost assumptions to utilization trends suggests that the IOPCM program may result in cost avoidance over time. For example, financial modeling for ED visit costs showed that ED costs were lower for intervention patients starting in the second year and continued to drop, resulting in an annual ED visit cost avoidance for a hypothetical cohort of 1000 intervention patients compared to 1000 controls of $116,678 by one and half years post intervention implementation. Annualized savings from cost avoidance for the same hypothetical intervention cohort increased to $313,531 in the first half of Year 3 of the intervention. Cost avoidance tied to inpatient utilization fluctuated more for the same hypothetical cohorts of intervention and control patients, going from $474K in annualized cost avoidance in the first half of Year 2, to an additional $1.1M in annualized costs for intervention patients in the latter half of Year 2, to $747K in annualized cost avoidance for intervention patients by the first half of Year 3. It is notable, that cost avoidance, even if it is uncertain whether it was due to the intervention or to chance, still represents real avoided spending for the health plan.

6) **There may be some positive changes in patient quality and experience over time; however, a larger sample size and comparable data for a control group are needed to make inferences.** Descriptive analysis of assessment scores data was conducted for patients who had assessment score data at all three assessment visits. The following assessment scores were analyzed: SF12 physical health composite score (SF12 PCS), SF12 mental health composite score (SF12 MCS), Patient Activation Measure (PAM),

PHQ2, PHQ9 and Audit C. There was no meaningful change in the mean SF12 PCS score, while there was a slight improvement in the mean SF12 MCS score, increasing from a mean of 36.6 at the initial visit to 39.9 at the final visit. The mean PAM score, however, decreased from 63.8 to 57.9, suggesting a decline in patient knowledge, skill and confidence in ability to manage health. For Audit-C, there was no change in the proportion of patients with a positive Audit C score (indicating no change in the prevalence of alcohol use disorders). For depression screening, the proportion of patients screening positive for depression declined from 60% at the initial visit to 43% at the final visit. However, the proportion of patients screening positive for severe depression increased from 23% at the initial visit to 30% at the final visit. Caution must be exercised while making inferences from these data given the small sample sizes and discrepancies in the ways these tools may have been administered across health centers. For example, unlike standard administration practices where only patients with a positive PHQ2 screen (a score of 3 or higher) receive a PHQ9 screen, in the IOPCM data the cohort of patients with PHQ2 scores at all three assessment points (n=30) also received a PHQ9 assessment score irrespective of their PHQ2 score. Specifically, 9 patients with a PHQ2 score less than 3 at all of their three assessment visits received a PHQ9 assessment. Ultimately, larger sample sizes and comparable data for a control group are needed to assess the intervention effect on patient quality and experience.

7) **Due to the small sample size of homeless patients, it is hard to make any conclusions about the intervention effect on homeless patients.** In the matched study sample of 519 patients, 15 intervention patients and 13 controls respectively were identified as homeless. For this subgroup of patients, ED visits PTMY decreased for both intervention patients and controls in the post-intervention period, with a steeper decrease for intervention patients up to the post 13-18 months period, followed by a steep increase in utilization in the post 19-24 month period. Inpatient admissions similarly declined for both intervention patients and controls in the post intervention period. At the start of the second year post-intervention, however, the pattern shifted with inpatient admissions PTMY for controls increasing while inpatient admissions PTMY for interventions declining to zero. Although these are both interesting findings, caution should be exercised while making any inferences about the intervention effect given the small sample size for this subgroup analysis.

8) **Differential utilization patterns among patients with versus without substance abuse suggest that the intervention may take longer to impact patients with substance abuse.** Comparing utilization patterns among intervention patients with versus without substance abuse, we found that ED visits PTMY for intervention patients **with** substance abuse decreased to levels below that of controls in the post 19-24 month period, while ED visits PTMY for intervention patients **without** substance abuse decreased to levels below that of controls six month earlier (in the post 13-18 month period). This finding suggests that it may take longer for IOPCM to impact patients with substance abuse. We also found that among patients diagnosed **with** substance abuse, inpatient admissions PTMY fluctuated for both controls and intervention patients. Inpatient admissions PTMY for patients **without** substance abuse tracked fairly closely for intervention patients and controls. Importantly, however, at each time interval, inpatient admissions PTMY were substantially lower in absolute terms for patients **without** substance abuse compared to those **with** substance abuse.

9) **Differential utilization patterns were observed for subgroups of patients based on the prevalence or absence of certain disease conditions. The findings are summarized in the table below.**

Disease Condition	ED Visits	Inpatient Admissions
CHF	A more consistent pattern of decreased ED visits over time among intervention patients **with** versus **without** CHF suggests that IOPCM may have a differential impact on ED visits based on the prevalence of CHF.	A more consistent pattern of decreased inpatient admissions over time among intervention patients **with** versus **without** CHF suggests that IOPCM may have a differential impact on inpatient admissions based on the prevalence of CHF.
COPD	ED visits declined over time for all IOPCM patients relative to controls irrespective of COPD prevalence. Importantly, intervention and control patients' utilization slopes crossed six months later for patients **with** versus **without** COPD, suggesting that IOPCM may take longer to impact patients with COPD.	Inpatient admissions PTMY tracked closely for intervention patients and controls irrespective of COPD diagnosis, suggesting that there is no differential program impact on inpatient admissions based on the prevalence/absence of COPD.
Hypertension	A more consistent pattern of decreased ED visits over time among intervention patients **with** versus **without** hypertension suggests that IOPCM may have a differential impact on ED visits based on the prevalence of hypertension.	A more consistent pattern of decreased inpatient admissions over time among intervention patients **with** versus **without** hypertension suggests that IOPCM may have a differential impact on inpatient admissions based on the prevalence of hypertension.
Diabetes	ED visits PTMY decreased for all intervention patients irrespective of diabetes diagnosis. However, over a two-year period, patients without diabetes may have experienced a steeper decline in utilization suggesting a slight differential program impact based on the absence of diabetes.	One year into the intervention, patients with diabetes had consistently lower utilization levels than controls, while utilization fluctuated for intervention patients without diabetes. This finding suggests that there may be a slight differential program impact based on the prevalence of diabetes**.**

Objectives

The goals of the evaluation were to: (1) support PHC in assessing the success of the IOPCM program and PHC's return on investment, and (2) allow PHC to inform the State's Health Home implementation and evaluation.

The specific objectives of JSI's outcome evaluation efforts were to:

1. Identify an appropriate control group comparable to the intervention group on demographics and other baseline variables hypothesized to affect receipt of intervention.
2. Assess how intervention patients and controls compare on health care utilization (inpatient admissions, outpatient visits, ED visits, and 30-day readmissions) pre and post intervention.
3. Examine longitudinal trends in health care utilization to assess how long a patient needs to be in the intervention to start seeing reductions in health care utilization, and whether these reductions in utilization were sustained over time.
4. Assess whether receiving the intervention is associated with utilization after controlling for factors hypothesized and/or tested in previous studies to be associated with utilization outcomes of interest.

This report presents a summary of the analytic approach and key findings from JSI's analysis for each objective.

Analytic Approach

Data used for the outcomes evaluation study included claims and enrollment data, risk scores data and IOPCM program data from September, 2011 to December, 2015. Due to small sample size per IOPCM site, JSI and PHC collectively decided to aggregate data across sites and analyzed the data as a single intervention and control group. Per discussions with PHC and consistent with standard assessment durations used in program evaluations, a 'six-month minimum continuous enrollment' (in IOPCM for intervention patients/in PHC for controls) was applied as the analysis inclusion criteria, reducing the sample size from 920 to 676. Propensity score-based matching was then conducted using multivariable logistic regression to identify an appropriate control group. Propensity score model building was an iterative process. The first model included an extensive list of variables determined in consultation with PHC and based on Phase I analysis. Consistent with standard statistical practices, the model accomplishing the desired level of prediction with as few predictor variables as possible was selected as the final model to estimate propensity scores. The final propensity matching model included the following variables: age (continuous), receipt of psychotic or non-psychotic mental health diagnosis in the pre-intervention period (binary), recent hospitalization (0-10 days before intervention enrollment) (binary), and homelessness status (binary) (See appendix for more details). Matches were successfully identified for 73% of the intervention sample, and the final sample size for the outcomes analysis was 519 (262 intervention patients and 257 controls).

All subsequent analyses were conducted on the matched cohort. Bivariate statistics were conducted to compare intervention patients to controls on: demographics, 12-months pre intervention characteristics (subsequently referred to as 'baseline'), and health care utilization.

For each individual patient in the study, we divided their membership and claims data into six-month time periods for the pre and post intervention years. For intervention patients, membership months and health care utilization in each interval was calculated relative to the individual patient's date of IOPCM enrollment. For control patients, these calculations were based on the date of IOPCM initiation (or contract start date) at the site that they were part of. To study utilization trends over time, total utilization per thousand member years (PTMY) was calculated for each six-month interval from baseline to 30 months post intervention. Per discussion with PHC, costs associated with utilization were calculated by multiplying total utilization in each interval by a proxy for average cost per ED visit and per inpatient admission.

Difference-in-difference analysis was conducted using multivariable regression to estimate the intervention effect on utilization over time after controlling for potential confounders, unobservable differences between delivery models, and potential regression to the mean. Sensitivity analysis was conducted to examine changes in the intervention effect, if any, after removing outliers from the matched cohort. Additionally, sensitivity analysis was also conducted using alternative risk adjustment methods to assess if intervention effects were different using other approaches as compared to the difference-in-difference with multivariable regression approach. This included: 1) standard multivariable regression analysis on the non-matched cohort (n=676 with six-month minimum continuous enrollment), and propensity score risk-adjustment which entailed using propensity scores as weights in a weighted outcomes regression. Analysis methods are described in detail in the Appendix.

Results

1. Propensity score method was effective in identifying an appropriate control group.

Table 1 describes baseline demographics, mental/behavioral health diagnosis, and other key characteristics of the intervention and control groups before and after propensity score-based matching. After propensity score-based matching, there were no statistically significant differences between the two groups. In other words, the matching was effective in minimizing baseline differences between eligible controls and intervention patients. This makes the two groups become more similar, mitigates concerns of selection bias, and increases the confidence that observed intervention effects can be attributed to the intervention and not to baseline differences between the two groups.

The one exception where differences remained before and after matching was in risk scores, which changed from being significant at the $p<.05$ level pre-matching to being significant at the $p<.10$ level after matching. However, the distribution of risk scores across the two groups did not shift in any particular direction, making it difficult to conclude that overall one group remained sicker/healthier than the other after matching.

Table 1. Frequency distribution of demographics and key variables hypothesized to influence treatment assignment for unmatched cohort and propensity matched cohorts

	Unmatched Cohort (n=676)			Propensity score matched cohort (n=519)		
Variable Name	Intervention Patients	Control Patients	Statistical Significance	Intervention Patients	Control Patients	Statistical Significance
	(n=318)	(n=358)		(n=262)	(n=257)	
	% or mean			% or mean		
Age			*			
18-39 years	13%	28%		15%	16%	
40-49 years	17%	20%		19%	20%	
50-59 years	49%	37%		47%	44%	
60-69 years	21%	15%		19%	21%	
Gender						
Female	58%	52%		58%	55%	
Risk Scores			*			**
0 to less than 2	24%	38%		28%	31%	
2 to less than 4	35%	24%		34%	26%	
4 to less than 6	15%	18%		16%	22%	
6 to less than 8	12%	10%		12%	11%	
8 to less than 10	7%	7%		5%	8%	
10 or higher	7%	3%		7%	4%	
Baseline ever received Mental Health/Substance Abuse diagnosis (12 months pre intervention)						
Psychotic mental health diagnosis	29%	18%	*	23%	21%	
Other non-psychotic mental health diagnosis	57%	43%	*	54%	52%	
Substance abuse diagnosis	54%	49%		53%	49%	
Comorbidity (substance abuse and mental health diagnosis)	33%	28%		31%	30%	
Aid Category			*			
LIHP /CMSP or Disabled^	82%	76%		83%	75%	
Family	8%	14%		9%	12%	
Aged Dual/Disabled Dual	2%	1%		2%	1%	
BCCTP	0%	1%		0%	2%	
Aged	8%	8%		7%	11%	
Hospitalization						
Ever hospitalized 0-10 days pre enrollment	8%	2%	*	5%	3%	
Homeless						
Identified as homeless	5%	11%	*	6%	5%	
Baseline membership duration (12 months pre intervention) (mean)	10.4	9.4	*	10.0	10.0	

*P<.05

**P<.05

^ LIHP/CMSP OR Disabled was combined as per discussion with PHC. At the Phase I presentation when JSI shared results showing that disproportionately more controls were categorized as LIHP/CMSP while more intervention patients were categorized as Disabled, PHC said that the State had financial incentives to keep members in the LIHP/CMSP category rather the Disabled category and they did not expect the two aidcat categories to converge over time. JSI created a combined LIHP/CMSP or Disabled category to improve matching on the aidcat variable.

Table 2 describes total baseline utilization before and after propensity-score based matching. As desired, matching did not influence utilization outcomes. The intervention group had significantly higher total ED and outpatient visits per thousand member years (PTMY) before and after matching ($p<.05$). While expected for a high-utilizer program, this difference does create some limitation in concluding that the control group was as equivalent to the intervention group as it would be in a randomized controlled trial.

Table 2. Total baseline utilization per thousand member years (PTMY) unmatched cohort and propensity matched cohort

	Unmatched Cohort (n=676)			Propensity score matched cohort (n=519)		
Utilization	**Intervention Patients (n=318)**	**Control Patients (n=358)**	**Statistical Significance**	**Intervention Patients (n=262)**	**Control Patients (n=257)**	**Statistical Significance**
Total baseline utilization PTMY (12 months pre intervention)						
Inpatient admissions	1162	1167		1124	1107	
ED visits	4217	3553	*	4330	3390	*
OP visits	16308	11708	*	16326	12825	*
30-day readmissions	266	189		280	169	

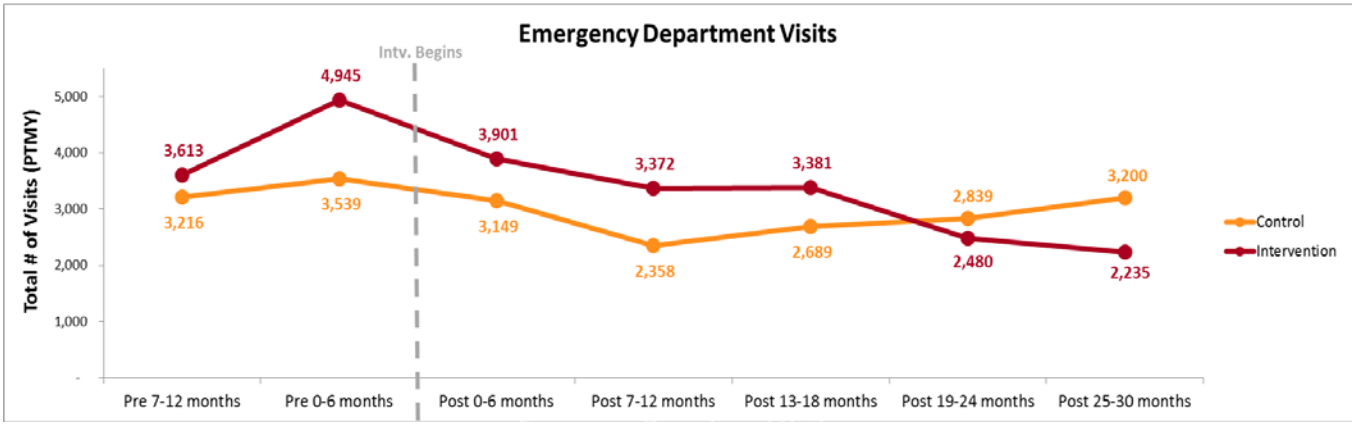
*p<.05

2. Both intervention and control patients experienced a downward shift in utilization with a notable differential in changes in the second year.

Figures 1- 3 depict the shift in total utilization PTMY from baseline to 30 months post intervention. Figure 4 depicts the shift in proportions of intervention versus control patients with 30-day readmissions from baseline to 30 months post intervention. The timeline is depicted on the x-axis and was divided into six-month intervals to demonstrate the subtle utilization shifts taking place over time.

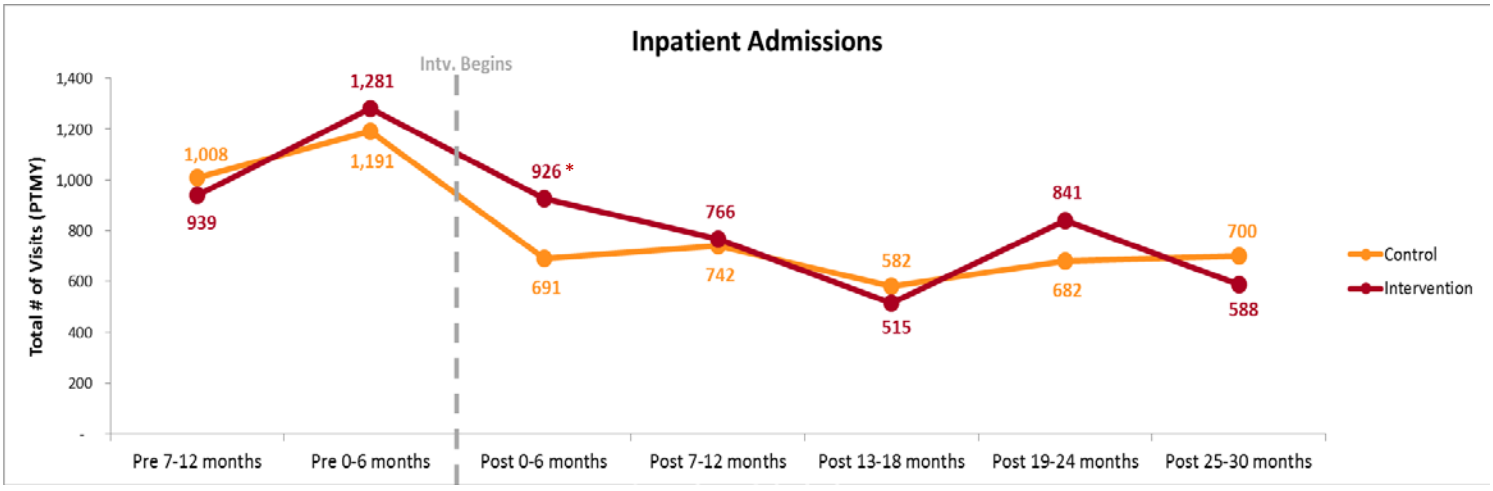
ED Visits. For intervention patients, after an initial increase from the pre 7-12 month to pre 0-6 month period, ED visits PTMY gradually declined in the post intervention period. Control patients' ED visits follow a similar downward trend. However, at 19-24 months, or during the second year, the direction switches with intervention patients' ED utilization PTMY being lower than that of control patients. In other words, if there were 1,000 interventions and 1,000 control members experiencing the second half of Year 2 utilization rates for a year, intervention patients would have 2.4 ED visits per patient while controls would have 2.8 ED visits per patient. This favorable trend is sustained over the next six-month interval, and the difference between the two groups becomes greater (2.2 ED visits per patient per year for intervention patients versus 3.2 ED visits per patient per year for controls).

Figure 1. Emergency department visits per thousand member years (PTMY) over time.



Inpatient Admissions. In terms of inpatient utilization, interventions and controls had more similar utilization at each six-month interval. Utilization rates also fluctuated in the post intervention period. Intervention patients' utilization started to decline in the post 0-6 month interval but then went back up in the post 19-24 month period and then again declined in the post 25-30 month period. During the post 13-18 month interval intervention patients' utilization was lower than that of controls (0.515 inpatient admissions per member for interventions versus 0.582 admissions per member for controls). Controls had similar fluctuation; inpatient utilization declined in the post 0-6 month period, increased in the post 7-12 month period, declined in the 13-18 month period to increase again in the post 19-24 month period and remained on the rise. This fluctuation in both groups could be attributed to decreasing sample size in the longitudinal data (see Study Limitations for further discussion).

Figure 2. Inpatient admissions per thousand member years (PTMY) over time.

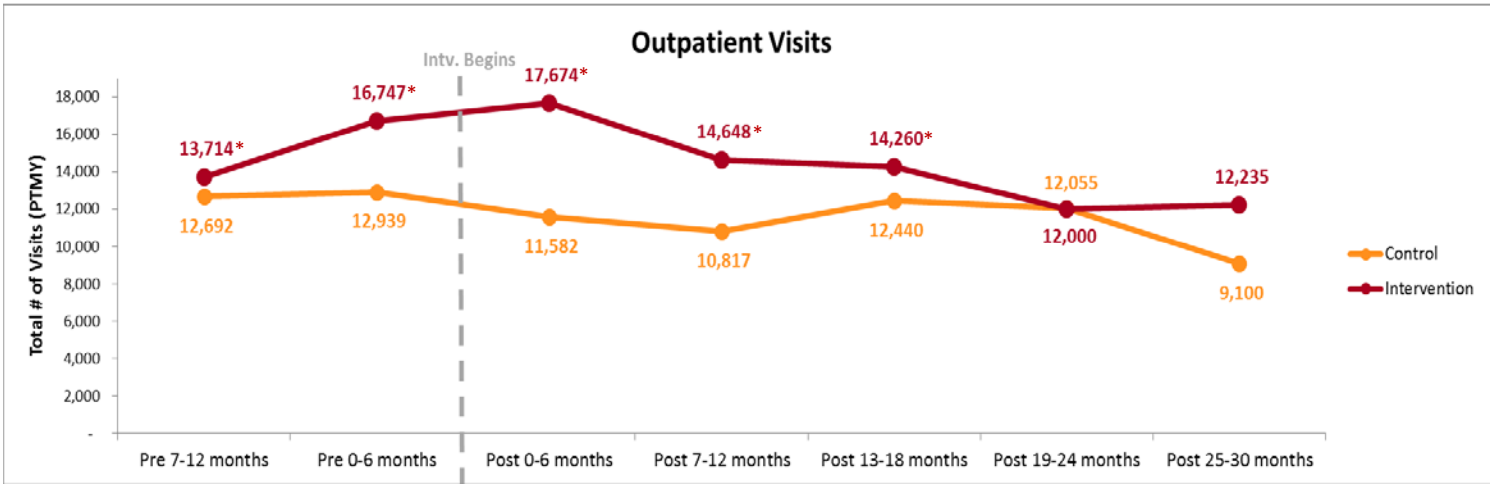


*p<.05

Outpatient Visits. Intervention patients had consistently higher total outpatient visits PTMY as compared to controls. The difference between the two groups was highest at the post 0-6 month interval (17,674 visits PTMY for intervention patients versus 11,582 visits PTMY for control patients), gradually declining thereafter. This

finding is not unexpected. Given the design of the IOPCM program, we would expect that with better identification and management of disease conditions, there would be a shift in sources of utilization away from ED visits and inpatient admissions towards outpatient visits. As such the outpatient visit utilization rate would remain higher for interventions patients compared to controls even over time.

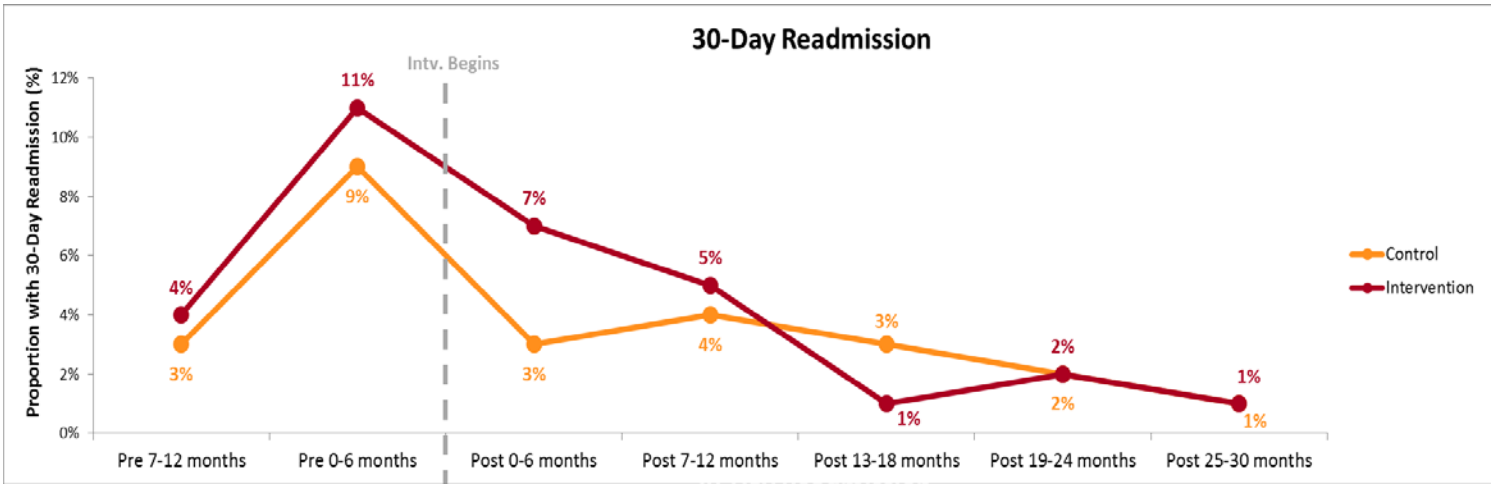
Figure 3. Outpatient visits per thousand member years (PTMY) over time.



*p<.05

30-Day Readmissions. The proportion of patients with 30-day readmissions declined for both intervention and control patients in the post-intervention period. At the post 13-18 month period, fewer intervention patients had 30-day readmissions compared to controls (1% versus 3%). Proportion of controls with 30-day readmission steadily went down while there was some fluctuation for the intervention patients. The fluctuation may be a result of the drop in sample size in the longitudinal data.

Figure 4. Proportion of patients with 30-day readmissions over time.



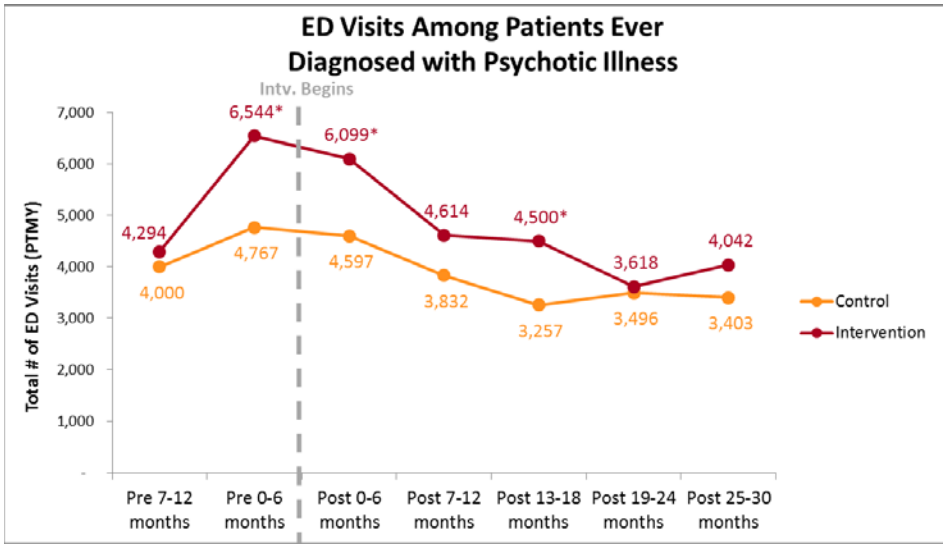
*p<.05

3. Differential utilization patterns for some subgroups of patients suggest absence of psychotic illness may make patients more susceptible to intervention impact, especially over time.

Based on the literature and expert consultations we hypothesized that the prevalence of psychotic mental illness may be associated with more stubbornly high utilization patterns that are both harder and longer to impact. To examine the potentially differential impact of the IOPCM intervention based on psychotic mental health diagnoses, utilization trends were compared for patients 'ever diagnosed with psychotic mental illness' in the study data versus patients 'not ever diagnosed with psychotic mental illness' in the study data.

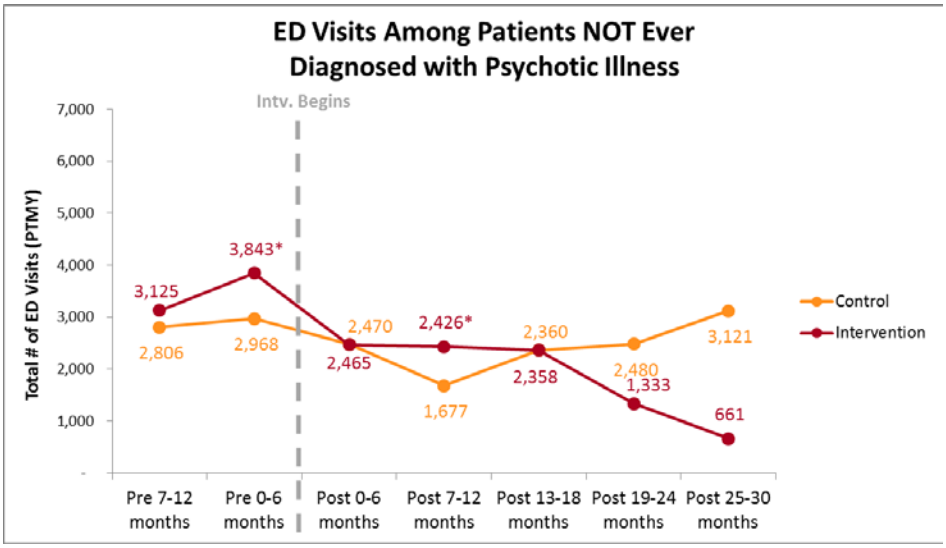
ED Visits. Figure 5 shows that among patients ever diagnosed with a psychotic mental health condition, ED visits PTMY declined for both intervention and control patients in the post-intervention period, though intervention patients had consistently higher ED visits PTMY in each six-month period. Importantly, Figure 6 shows that that among patients **not** ever diagnosed with psychotic mental illness, ED visits PTMY for intervention patients, although higher at the pre 0-6 month period, declined in the post-intervention period in year 1, dropping rapidly at the post 19 month period and remaining on the declining trend. ED visits PTMY for control patients in the subgroup of patients **not** ever diagnosed with psychotic mental illness fluctuated, going down in the first year but then rising back up in the second year and remaining on the upward trend. These findings, although not statistically significant, suggest that a lack of psychotic illness may make patients more susceptible to the impact of the intervention, especially over time.

Figure 5. ED visits among patients ever diagnosed with psychotic illness



*p<.05

Figure 6. ED visits among patients not ever diagnosed with psychotic illness

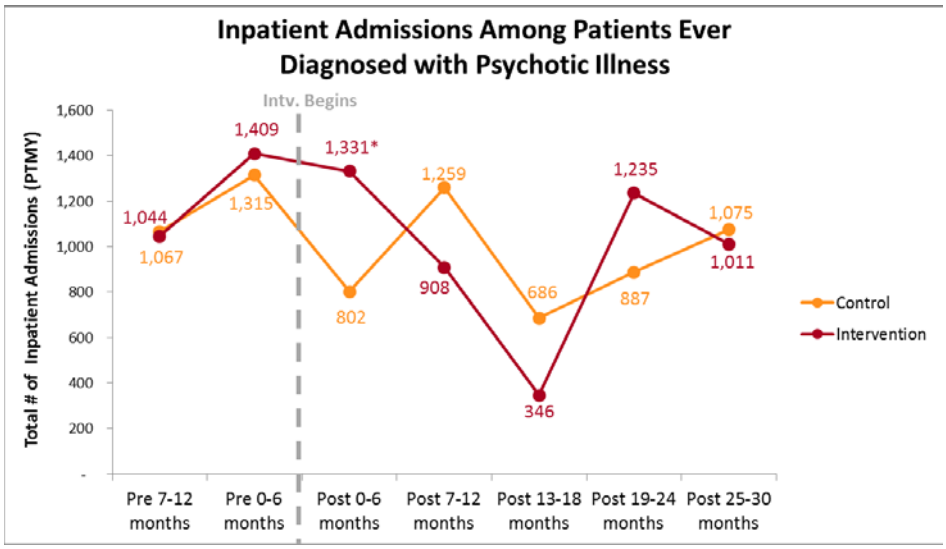


*p<.05

Inpatient admissions. Figure 7 shows that among patients **ever** diagnosed with a psychotic mental health condition, inpatient admissions PTMY for controls and interventions patients fluctuated quite a bit. Inpatient admissions PTMY were higher for intervention patients compared to controls in the first 6-months post-intervention but then started to decline, dropping substantially below that of controls in the post 7-12 and 13-18 month intervals only to rise back up in the post 19-24 months period. Inpatient admissions PTMY for controls **ever** diagnosed with a psychotic mental illness fluctuated similarly, increasing at the post 7-12 month period, dropping in the post 13-18 month period and rising back up again. This finding suggests that the intervention impact on patients with psychotic illness may take longer and be more cyclical. Interestingly, Figure 8 shows that among patients **not** ever diagnosed with psychotic mental illness, inpatient admissions PTMY for intervention

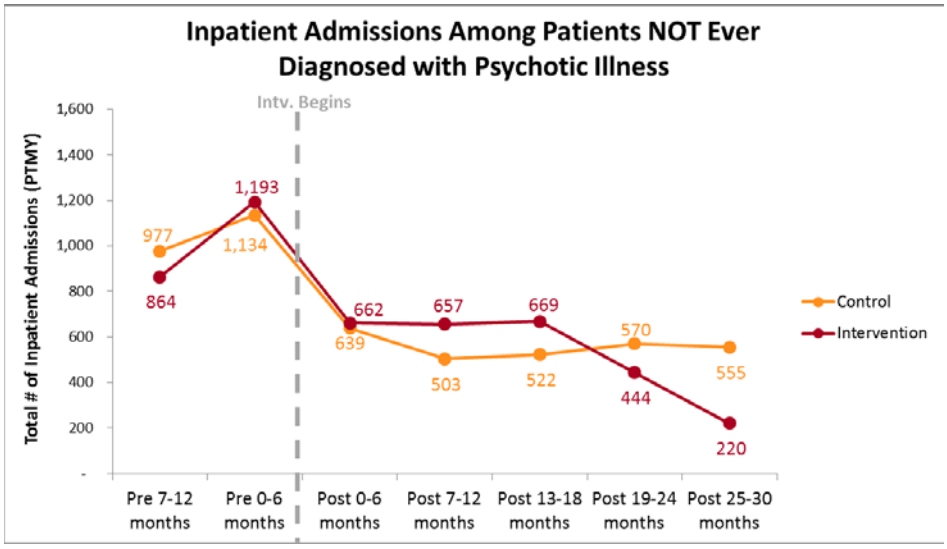
patients were comparable to that of controls at most six-month time periods, rose slightly above controls in 7-18 post intervention, and dropped below that of controls in the 19-24 month period, remaining on the downward trend. Inpatient admissions PTMY for control patients **not** ever diagnosed with psychotic mental illness, after an initial decline, started to increase, albeit gradually, in the post 13-18 months period and continued to fluctuate.

Figure 7. Inpatient admissions among patients ever diagnosed with psychotic illness



*p<.05

Figure 8. Inpatient admissions among patients not ever diagnosed with psychotic illness



*p<.05

4. Intervention effect on utilization was not statistically significant but is directionally meaningful, especially over time.

The results presented in this section were calculated using multivariable regression methods and should be interpreted differently from the utilization trends analysis presented above. Important differences include:

1) Utilization trends analyses are based on six-month intervals: pre 7-12 months to post 25-30 months. Data were analyzed despite the sample size drop in each post six-month period to provide a glimpse of the downward utilization trends that occur over time. The main multivariate regression model was conducted on the combined 12 months pre and post-intervention period to increase sample size and power to detect an intervention effect and to ensure uniformity across pre/post time periods (12 months each) to make interpretation easier. Models for the second intervention year were analyzed for six-month intervals due to an approximately 34% drop in sample size from post 13-18 months to post 19-24 months. Regression results, though not statistically significant, are included below as they indicate the potentially favorable intervention effect that may occur with larger sample sizes and over longer time periods that may even show statistical significance.
2) Utilization analysis results presented above measure total utilization PTMY. Multivariable regression results presented below for ED visits, inpatient admissions, and outpatient visits are the coefficients (odds ratio for 30-day readmissions) obtained from a regression model that controlled for pre-intervention utilization and potential confounders of the intervention effect (e.g., risk scores and aid categories). Multivariable regression results presents for the 30-day readmission outcome represents the regression-adjusted odds ratio. Multivariable regression models do not control for factors that were included in the propensity score model (i.e., age, receipt of psychotic or non-psychotic mental health diagnosis in the pre-intervention period, recent hospitalization (0-10 days before intervention enrollment), and homelessness status), but include other factors hypothesized to confound the intervention effect. This includes: risk scores, aidcat categories, and model delivery type.
3) Utilization analyses results are presented as utilization rates PTMY for intervention patients and controls at each time period. Regression coefficients represent the difference-in-difference (DID) of the means of the two groups. In other words, the DID model calculates the change from pre-12 months to post-12 months for controls and interventions separately, and then estimates the difference in the change between the two groups, indicating whether the change is greater or smaller for interventions compared to controls. The DID model results estimates how the rate of change for interventions compares to the rate of change for controls over two time periods after controlling for other factors that may influence the intervention effect.

Difference-in-difference analysis using multivariable regression revealed directionally meaningful but not statistically significant differences in utilization comparing intervention patients to controls (Tables 3 and 4).
P-values and confidence intervals are not shown since intervention effects were not statistically significant.

ED Visits. Table 3 shows that being in the intervention for one year is associated with 382 higher ED visits PTMY for intervention patients compared to controls after controlling for baseline ED visits (0-12 months), risk scores, model delivery type, and aid categories. Results were similar for the 13-18 month period; being in the

intervention for 18 months is associated with 132 higher ED visits PTMY for intervention patients compared to controls after controlling for ED visits in the pre-0-6 month period and other confounders. Repeating the model for the post 19-24 month period revealed different results: being in the intervention for two years was associated with 1039 fewer ED visits PTMY for interventions compared to controls after controlling for ED visits in the pre-0-6 month periods and other confounders.

Inpatient admissions. Table 3 shows that being in the intervention for one year (0-12 months) is associated with 168 fewer inpatient admissions PTMY for intervention patients compared to controls after controlling for baseline inpatient admissions (0-12 months), risk scores, model delivery type, and aid categories. In other words, a group of 1,000 intervention patients being continuously insured for a year would be predicted to have 168 fewer hospital admissions as compared to 1,000 controls. There is some fluctuation over time likely resulting from loss of sample size in subsequent time intervals. Being in the intervention for 18 months is associated with 97 higher inpatient admissions PTMY for intervention patients compared to controls after controlling for inpatient admissions in the pre-0-6 month period and other confounders. The intervention effect becomes favorable again in the post 19-24 month period. Being in the intervention for 24 months is associated with 27 fewer inpatient admissions PTMY for intervention patients compared to controls after controlling for inpatient admissions in the pre-0-6 month period and other confounders.

Table 3. Estimated effect of IOPCM intervention on utilization. Numbers for utilization shown as PTMY.

	Difference- in- difference coefficients PTMY		
	Post 0-12 Months (n=492)	Post 13-18 Months (n=275)	Post 19-24 Months (n=182)
ED visits[1]	382	132	-1,039
Inpatient admissions[1]	-168	97	-27

[1]Controling for receipt of intervention, baseline utilization, model type, risk scores and aid category

30-Day Readmissions. Table 4 shows that being in the intervention for a year controlling for model delivery type and inpatient admissions in the post 0-12 month period is associated with a higher odds of having a 30-day readmission for interventions compared to controls (Odds Ratio=1.4). In other words, intervention patients were 40% more likely to have a 30 day-readmission compared to controls. The odds decreased to .1 in the post 13-18 month period, implying that intervention patients were 90% less likely to have a 30-day readmission compared to controls. The odds increased in the 19-24 month period to 1, indicating no difference between groups.

Table 4. Estimated effect of IOPCM intervention on utilization. Numbers shown as odds ratio.

	Difference- in- difference Odds Ratio		
	Post 0-12 Months (n=492)	Post 13-18 Months (n=275)	Post 19-24 Months (n=182)
30-day readmissions[2]	1.4	0.1	1.0

[2]Odds or likelihood of having a 30-day readmission in the follow-up period comparing intervention patients to controls controlling for model type and number of hospitalizations in the follow-up period

5. Savings from cost avoidance, although not immediate, may accumulate over time

Costs associated with changes in utilization were calculated making several assumptions about average PHC expenses for different types of utilization. Applying these cost assumptions to utilization trends shows that the IOPCM program may result in cost avoidance over time. It should be noted that the financial modeling is simply an estimation of how the observed utilization trends would translate into costs using standardized cost assumptions. The financial modeling is provided to give a sense that even if there is a greater than 5% chance that the observed differences were due to chance, the difference in costs to PHC would still be real and meaningful in dollar terms. Figures 9, 10 and Table 5 show modeled annualized costs and cost differences for ED visits and inpatient admissions over time. The modeled costs are representative of the costs for a hypothetical cohort of 1,000 control patients and 1,000 intervention patients over time using their calculated utilization rates PTMY, which, in turn, are based on actual utilization observed in each six-month period. In other words, using the utilization rates for control and intervention patients in six-month periods, annualized costs were estimated by projecting what each group's costs would be in a year in which their utilization rates for each six-month period were maintained for the rest of that year. Annualized estimates are presented to maintain consistency with figures presented throughout the report, as well as with standardized measures in the field. To estimate the modeled costs experienced in each six-month period, costs in Table 5 would simply need to be halved.

ED Visit Costs. Financial modeling for ED visit costs showed that ED costs were lower for intervention patients starting in the second year and continued to drop, resulting in an annual ED visit cost avoidance for a hypothetical cohort of 1000 intervention patients compared to 1000 controls of $116,678 by one and half years post intervention implementation. Annualized cost avoidance for the same hypothetical intervention cohort increased to $313,531 in the first half of Year 3 of the intervention.

Inpatient Admission Costs. Financial modeling for inpatient admission costs shows that cost avoidance tied to inpatient utilization fluctuated more for the same hypothetical cohorts of intervention and control patient. It went from $474K in annualized cost avoidance in the first half of Year 2, to an additional $1.1M in annualized costs for intervention patients in the latter half of Year 2, to $747K in annualized cost avoidance for intervention patients by the first half of Year 3. Ultimately, for both the utilization and cost assessments, a larger cohort followed over a longer period of time would be needed to demonstrate statistically significant results.

Figure 9. Modeled ED visit costs per thousand member years (PTMY) over time.

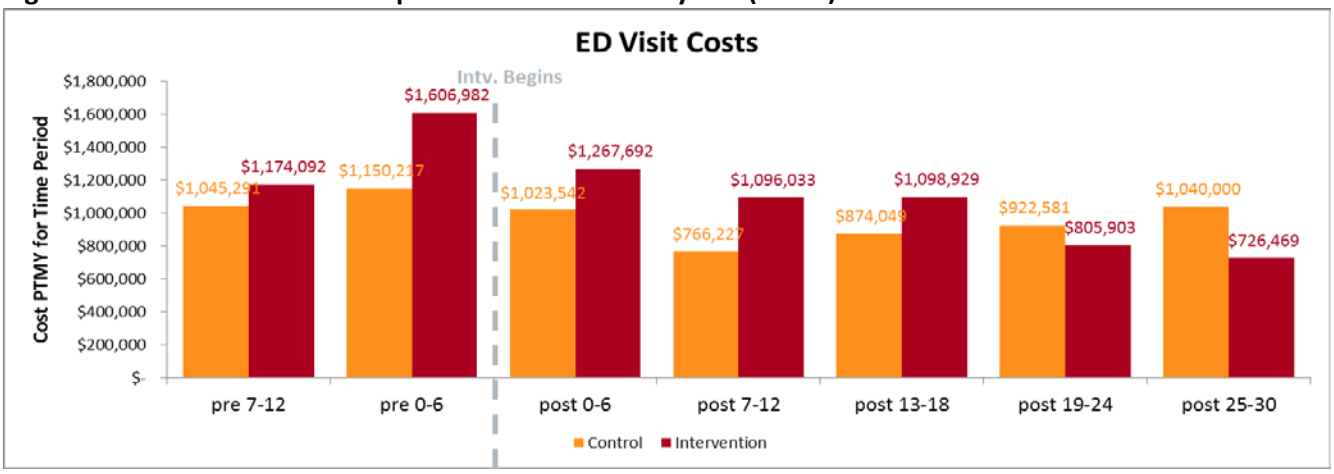
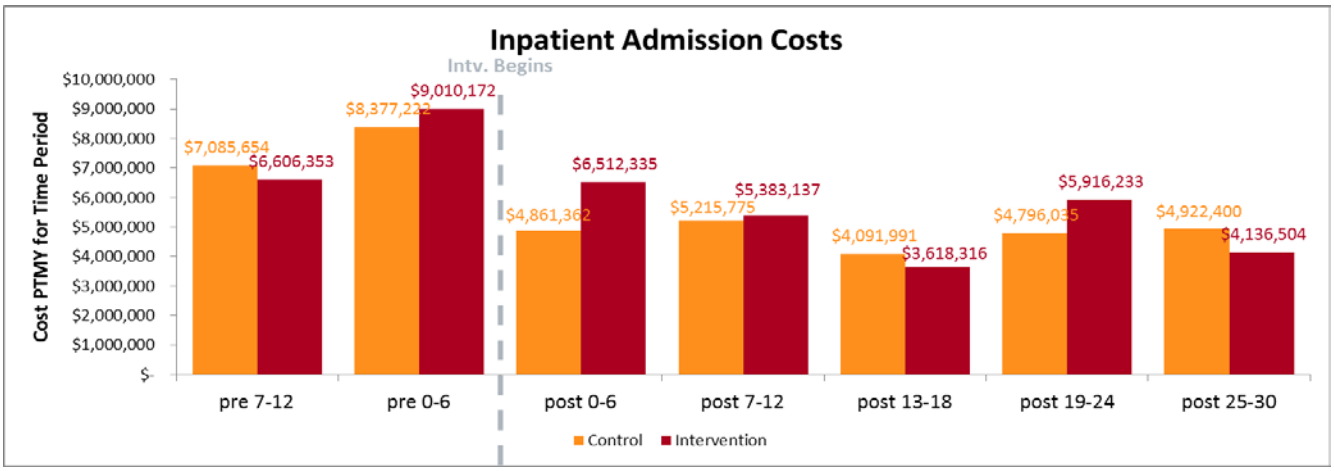


Figure 10. Modeled inpatient admission costs per thousand member years (PTMY) over time.



Hospitalization costs were calculated using a proxy cost of $7,032/stay (assumed stay of 2 days, $3516/day)

Table 5. Estimated annualized costs and cost differences for a hypothetical cohort of 1,000 control and 1,000 intervention patients pre- and post- intervention.

EMERGENCY DEPARTMENT*							
Time Period	pre 7-12	pre 0-6	post 0-6	post 7-12	post 13-18	post 19-24	post 25-30
Control Cost	1,045,291	1,150,217	1,023,542	766,227	874,049	922,581	1,040,000
Intervention Cost	1,174,092	1,606,982	1,267,692	1,096,033	1,098,929	805,903	726,469
Difference (C - I)	-128,801	-456,765	-244,150	-329,807	-224,881	116,678	313,531
INPATIENT ADMISSION**							
Time Period	pre 7-12	pre 0-6	post 0-6	post 7-12	post 13-18	post 19-24	post 25-30
Control Cost	7,085,654	8,377,222	4,861,362	5,215,775	4,091,991	4,796,035	4,922,400
Intervention Cost	6,606,353	9,010,172	6,512,335	5,383,137	3,618,316	5,916,233	4,136,504
Difference (C - I)	479,301	-632,950	-1,650,973	-167,362	473,676	-1,120,198	785,896

*ED costs were calculated using a proxy cost of $325/visit

**Hospitalization costs were calculated using a proxy cost of $7,032/stay (assumed stay of 2 days, $3516/day)

Analysis Limitations

Propensity score matching, though effective in minimizing baseline differences between intervention and control patients, resulted in a 27% loss in sample size (i.e. no matches for 27% of the intervention patients). This is a limitation common to all matching methods. If good matches were not available for some intervention patients, these patients were removed from the analysis. To minimize the loss of sample from propensity score matching, matching was conducted on rounded propensity score ranges instead of matching on exact propensity scores. Importantly, matches were found for 73% of the intervention sample which is comparable to other studies[1,2,3,4,and 5]. Propensity score matching typically is unable to find matches for patients at the tail ends of the propensity score distribution. This implies that it is possible that intervention patients with the highest probability of being in the intervention were not included in the analysis since they had no suitable controls. These individuals could also have been among the highest users who may have had an even greater effect from being in the intervention. Their intervention effects are not reflected in the results since appropriate matches could not be found and they were removed from the analysis. To address this concern, we conducted standard multivariable regression on the non- matched cohort, including all factors that were included in the propensity model and other factors hypothesized to influence intervention effects. There were no differences in the intervention effects as compared to the difference-in-difference analysis conducted on the propensity score-matched cohort either directionally or in terms of statistical significance.

A second limitation with propensity scores is that while propensity score-based matching will improve the distribution of measurable factors between intervention and comparison groups, some unmeasured factors will likely remain. Selection bias for such factors that cannot be observed (e.g., patient engagement, adherence, attitude, relationship with provider etc.) may persist even after matching.

A third limitation of the analysis was diminishing sample sizes with more than 1 year of post-intervention data. Since IOPCM was initiated across sites on an ongoing basis from October 2013 through October 2015 and JSI's analysis used data until December 2015, analysis of the intervention effect over time for the full sample was limited. There was a substantial drop in data availability on member months and claims data beyond the one year post-intervention period. Sample size fell to 53% of the original matched sample in the 13-18 months post-intervention period, and to 35% of the original matched sample in the next six-month interval. To address this limitation, utilization and costs analyses were conducted on six-month intervals to make maximum use of the available data. For the difference-in-difference analysis, two post-intervention six-month intervals were combined (0-6 months and 7-12 months), but outcomes in the 13-18 and 19-24 month time periods were modelled separately to minimize loss in sample size and resulting power to detect an intervention effect. Regression models were not conduced for subsequent time periods due to substantial loss of sample size and resulting lack of power (Table 7 in the appendix describes the drop in sample size).

Finally, eligible control patients included participants across sites that were identified by physicians for intervention, approached to participate, but declined to participate. This group is therefore not entirely comparable to the group of intervention patients as in a randomized control trial and implies that to some extent selection bias could remain.

Conclusions

In summary, the outcome evaluations study was able to establish a rigorous methodology for retrospectively creating a control group for the IOPCM program. The study showed declines in ED visits and inpatient admissions for IOPCM patients and controls in Year 1, and promising shifts in intervention patients' ED and inpatient admission utilization in Year 2, resulting in intervention patients' utilization trending below controls' utilization in Year 2 and beyond. The difference-in-difference regression analysis did not reveal statistically significant differences between intervention and control patients pre and post intervention. However, directional results were promising and could be meaningful, both to patients' human experience and to the health plan's finances.

Importantly, these findings are consistent with evaluations of similar programs. Rosenthal et al. evaluated the effects of the Rochester Medical Home Initiative (RMHI) and the Rhode Island Chronic Care Sustainability Initiative pilot program on primary care, inpatient and ED utilization, and total cost of care[6, 7] using similar analysis methods as JSI's outcomes evaluation. Even after two to three years and with substantially higher sample sizes (over 30,000 member months) they found only non-significant but downward trends in ED visits and inpatient admissions. The evaluation of the Northeastern Pennsylvania Chronic Care Initiative showed significant reductions in hospital admissions and ED visits, but only by year three of the intervention [8]. Several other comparable programs for care management and care coordination have similarly showed downward trends in utilization with the magnitude of change being significantly greater over time.[9, 10]

Another noteworthy issue is the growing attention around the excessive reliance on p-values alone to make inferences about epidemiological studies and program evaluations.[11] Another form of scientific reasoning known as the Bayesian Induction method is gaining importance. Instead of relying entirely on p-values to deem program success or failure, this method recommends examining intervention effects from your study in conjunction with prior studies that evaluated the same causal effect, and additionally incorporating contextual and subject matter evidence to make inferences about programs. When compared to other programs, results from JSI's outcome evaluations study, even though not statistically significant, are very much in line with and in some cases even better than results from outcomes evaluations of comparable high-utilizer programs. [12]

In conclusion, the IOPCM program is likely to have a positive impact on the population it is serving, but more data would be needed to find statistically significant results. The longitudinal analysis suggested that patience may be necessary to see results, an important finding given that the Health Homes program will be scrutinized before it has been in effect for two years. PHC longitudinal data for IOPCM patients that become Health Homes patients could be integral to showing the longer-term effects of case coordination and care management. Indeed, other innovative health care delivery interventions that target the highest cost and most vulnerable populations, and that aim to change utilization behavior seem to require investment over more than 18 months to realize and sustain improvements in utilization, health outcomes, patient experience and quality of care.[8] IOPCM and Health Homes will need diligent data collection and analysis over time to track what is and isn't working, remembering that an early assessment may deem the program as a failure too soon, while a longer-term horizon may enable assessment of the true value and ROI of the program.

Using the available data, JSI has conducted a quantitative analysis that is as rigorous as possible. JSI recommends supplementing these results with qualitative analysis to gain a fuller picture of program impact. Suggestions for future qualitative research include: 1) in-depth interviews/focus group discussions with IOPCM patients and controls, including those with shorter and longer intervention exposure; 2) in-depth interviews with clinical and non-clinical providers of the different delivery models.

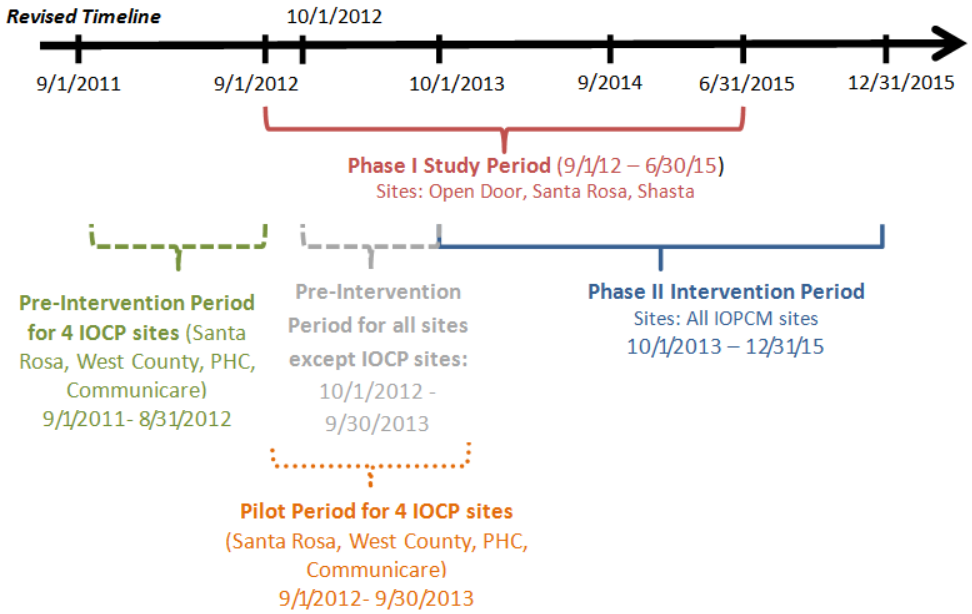
Appendix: Data and Methods

Sample. Data for JSI's outcomes evaluation were made available by PHC and included member eligibility, claims data, risk score data and IOPCM program data from September 2011 through December 2015. The initial dataset included 924 participants from 12 sites; 4 participants were removed after consultation with PHC due to missing program data (i.e. inability to determine intervention/control status). Control patients included participants from all sites who were identified by physicians for intervention, approached to participate, but declined to participate.

Due to small sample size per site, JSI and PHC collectively decided to aggregate data across sites and analyzed the data as a single intervention and control group. To prepare the database, JSI integrated claims files with program data and member eligibility data, created unique encounters for outcomes of interest (e.g., inpatient admissions, ED visits, OP visits and 30-day readmissions) using the agreed upon coding algorithms and created new variables (e.g., 30-day readmission, recent hospitalization 0-10 days before enrollment etc.). Database aggregation was conducted using SAS 9.4 and analysis was conducted using StataSE 13.

As reflected in Figure 11, per discussion with PHC and in order to maximize sample size, JSI treated the intervention start date as 9/1/2012 for participants from the four IOCP pilot sites, and 10/1/2013 for participants from all other sites. September 2011-August 2012 was treated as the baseline period for IOCP site patients, and October 2012-September 2013 was treated as the baseline period for non-IOCP site patients. September 2012-December 2015 was treated as the Phase II intervention period for patients from IOCP sites, while October 2013-December 2015 was treated as the Phase II intervention period for patients from non-IOCP sites.

Figure 11. Timeline



Per discussions with PHC and consistent with standard assessment durations used in program evaluations, a 'six-month minimum continuous enrollment' (in IOPCM for intervention patients/in PHC for controls) was applied as the analysis inclusion criteria. This six month criteria reduced the sample size to 676 (318 controls and 358 intervention patients).

Propensity score-based matching analysis. Propensity score-based matching was conducted on participants meeting the 'six-month minimum continuous enrollment' criteria to identify an appropriate control group, and to minimize baseline differences between eligible controls and intervention patients. This entailed multivariable logistic regression modeling to estimate a propensity score for each individual, i.e. the probability of receiving the intervention conditional on variables hypothesized to influence intervention assignment. Model building was an iterative process. The initial list of variables to be included in the model was determined based on findings from Phase I (variables where the two groups were found to be statistically significantly different at the $p<.05$ level) and in consultation with PHC. The first model included all demographic variables and other baseline characteristics where the two groups differed and/or that were hypothesized to influence intervention assignment. To improve the model fit, the model was repeated using different functional forms of the covariates (e.g., age as a continuous variable, risk scores as a continuous risk score variable, indicator variables for the two largest aid categories: disability and LIHP/CMSP). Likelihood ratio testing was conducted after each model to identify the best fit and most parsimonious model. Consistent with standard statistical methods, only variables significant at the $p<.05$ level in the full model were included in the final model. Table 6 below lists variables used in the full and final multivariable regression models.

Table 6. Variables included used in the full and final multivariable regression models.

Variables	Full Model	Final Model
Age	Yes	Yes; as continuous variable
Gender (binary; 1 if female, 0 if male)	Yes	
Risk Score	Yes	
Ever received non-psychotic mental health diagnosis in baseline (binary; yes/no)	Yes	Yes
Ever received psychotic mental health diagnosis in baseline (binary; yes/no)	Yes	Yes
Ever received substance abuse diagnosis in baseline (binary; yes/no)	Yes	
Ever received comorbidities diagnosis in baseline (substance abuse AND mental health) (binary; yes/no)	Yes	
Aidcat (5 categories)	Yes	
Recent hospitalization 0-10 days pre enrollment (binary; yes/no)	Yes	Yes
Identified as homeless (binary; yes/no)	Yes	Yes
Had 12 months PHC membership pre enrollment (binary; yes/no)	Yes	

Matching was conducted on propensity score strata instead of exact propensity scores to minimize loss of sample (i.e. propensity scores were categorized into ranges and one control was identified for each intervention patient based on having a propensity score within the same range). Matches were successfully identified for 73% of the intervention sample, and the final sample size for the outcomes analysis was 519 (262 intervention patients and 257 controls). Intervention group participants for whom an appropriate match could not be found among the control pool had to be removed from the analysis. Bivariate testing using Chi-square tests for categorical variables and Student *t-tests* for continuous variables was conducted to ensure that propensity-score based matching improved the baseline distribution of demographic and other baseline variables between the intervention and control groups.

Utilization and cost trends analysis. To study utilization patterns over time and maximize use of available data, we established six-month intervals starting from baseline through 36 months. For intervention patients, membership months and health care utilization in each interval was calculated relative to each individual patient's date of IOPCM enrollment. For control patients, these calculations were based on based on the date of IOPCM initiation (or contract start date) at the site that they were part of. Since participants had varying membership durations, the sample size available per six-month interval varied and declined substantially after 24 months post intervention. Table 7 depicts the drop in sample size availability over time. Since controls did not have any member months in the 31-36 six-month intervals, comparisons between the two groups were limited at this point, and this interval was removed from all analyses.

Table 7. Sample size availability over time

Time period	Total N with data/member months N, (% of 519)	Intervention N, (% of 262)	Control N, (% of 257)
7-12 pre intervention	450, (87%)	229, (87%)	221, (86%)
0-6 pre intervention	492, (95%)	256, (98%)	236, (92%)
0-12 pre intervention	492, (95%)	256, (98%)	236, (92%)
0-6 post intervention	519, (100%)	262, (100%)	257, (100%)
7-12 post intervention	483, (93%)	246, (94%)	237, (92%)
0-12 post intervention	509, (98%)	258, (98%)	251, (97%)
13-18 post intervention	275, (53%)	142, (54%)	133, (52%)
19-24 post intervention	182, (35%)	70, (27%)	112, (44%)
25-30 post intervention	110, (21%)	35, (13%)	75, (29%)
31-36 post intervention	31, (6%)	31, (12%)	0, (0%)

Utilization patterns were examined over these intervals for the following health outcomes: inpatient admissions, emergency department (ED) visits, outpatient visits, and 30-day readmissions. For each interval and by intervention and control group, claims for each outcome were aggregated, divided by

member months in that interval for each respective group, and multiplied by 12,000 to calculate total utilization per thousand member years (PTMY).

Costs associated with utilization were estimated using several assumptions. Costs per ED visit cost was estimated at $325 (estimate provided by PHC). As per discussion with PHC during Phase I, to estimate inpatient admission costs, cost per hospital stay was used instead of cost per hospital day since the intervention is aimed at reducing hospital admissions and cannot influence hospital days once a patient is admitted. Cost per hospital day at $3,516 (estimate provided by PHC) was multiplied by the median length of stay for the study sample (2 days) to get a cost per hospital stay of $7,302.

Difference-in-difference analysis. We conducted difference-in-difference analysis using the ANCOVA regression method. The main independent variable was whether or not the patient was in the intervention or control group and the second key variable was baseline utilization. We also controlled for model delivery type to adjust for unobservable differences between delivery models, as well as other variables that were not included in the final propensity score model but that were hypothesized to confound the intervention effect including risk scores and aidcat categories. One model was conducted per outcome and each model was repeated for three time periods (0-12 months post intervention, 13-18 months post-intervention and 19-24 months post-intervention).

The regression equation can be summarized as:

Follow-up utilization = constant+ a*(baseline utilization) +b*(intervention/control group) +c*(delivery model) + d*(other confounders)

Here coefficient b is the main effect of interest—the estimated difference in means between the intervention and control group after controlling for baseline utilization. This model was repeated for several intervals (0-12 month post intervention, 13-18 months post intervention and 19-24 months post intervention). Each model used uniform baseline and post-intervention periods e.g. 0-12 months baseline was used as the baseline for the 0-12 months post intervention, while the 0-6 months baseline period was used as the baseline for the 13-18 months and 19-24 months post intervention to ensure uniformity across pre/post time periods (12 months and 6 months respectively).

This method was selected after exploring other risk adjustment techniques including: 1) post-intervention analysis that examines the intervention effect at post time periods after controlling for confounders, and 2) change score analysis which entails subtracting the baseline value from the post value to assess the intervention effect on changes in utilization. Both of these methods, while robust, have several limitations that are better addressed by the ANCOVA method[13, 14]. First, ANCOVA adjusts each patient's post intervention value for their baseline value but, unlike change score analysis, has the benefit of being unaffected by baseline differences between groups. Second, it is considered a better method when dealing with regression to the mean, which has been a concern for the IOPCM program. If there is regression to the mean, we would expect intervention patients' utilization to go down over time, and post-intervention analysis does not take the change into account and may provide an

underestimate of the effect. Change score analysis, while taking the baseline into consideration, does not control for baseline imbalances due to regression to the mean and may overestimate the intervention effect. Third, ANCOVA has a higher statistical power to detect an intervention effect size compared to other methods. Fourth, ANCOVA is preferred over the change analysis method when there is medium correlation between baseline and follow-up values, as is the case with these IOPCM data.

Sensitivity analysis. We examined the distribution of utilization values over two time periods (baseline and 12 months post intervention) to describe the range of utilization values, identify outliers (i.e. users with higher than expected utilization values), and determine treatment of outliers. Upon examining the outliers, we found that the outliers for ED and inpatient admissions were different people i.e., there was no uniform set of patients who were outside of the normal range for all health care utilization outcomes of interest and could be simply removed from all analysis. The difference-in-difference regression models were repeated for inpatient admission and ED visit utilization after removing the top 5% high utilizers in the pre-intervention period in each category. First, we removed patients with the top 5% highest inpatient admissions. Next, we put these patients back into the model and removed patients with the top 5% highest ED visits. Results showed that removing patients with the highest inpatient admission in the pre-intervention period impacted not just the treatment effect for inpatient admissions but also ED visits. Similarly, removing patients with the top highest ED visits in the pre-intervention period impacted not just the treatment effect for ED visits but also inpatient admissions. These findings are not surprising given the correlation between inpatient admission and ED visits. Importantly, there were no statistically significant or directional changes in the intervention effect and we therefore decided to retain all patients in the analysis for the following reasons:

1. Removing outliers introduces bias. Removing the highest utilizers may, in fact, introduce bias that could be interpreted as regression to the mean bias. For example, as noted above, our analyses showed that removing patient with highest ED visits affected the intervention effect for inpatient admissions not just the intervention effect for ED visits. To avoid introducing bias, highest users were not excluded from the final analyses.

2. Including the highest utilizers in the analysis is a more robust evaluation approach and the distributions are not unreasonable given the program goals. Since the intervention targets high utilizers it is important to study the shifts in utilization, if any, when these users are included in the analysis. The range of values seems reasonable given the fact that the intervention targets patients at risk for high service utilization; all high values may not necessarily be outliers. Importantly, distribution of utilization values were positively skewed (i.e., more observations had lower values) and fairly similar in the pre- and post-intervention periods and for interventions and controls.

Alternative risk adjustment approaches were also explored. We conducted standard multivariable regression analysis on the non-matched cohort (n=676 with six-month minimum continuous enrollment). We also conducted propensity score risk-adjustment using propensity scores as weights in a weighted outcomes regression. There were no differences in the results either directionally or in terms of statistical significance.

Additional Analysis

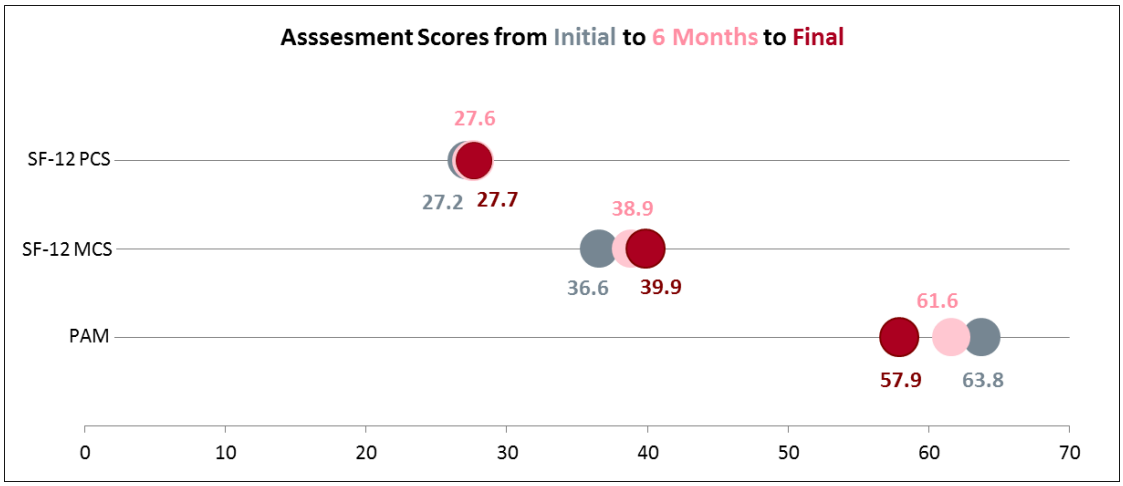
After submission of JSI's final report and presentation of findings, PHC requested additional analyses. Findings from these analyses are summarized in this section.

Assessment Score Analysis

PHC provided JSI with assessment scores data for IOPCM intervention patients. These data were reviewed for quality and completeness to assess the feasibility of examining the association between assessment scores and health care utilization. There was a large drop in data availability from the initial to six-month to final visit. Per consultation with PHC, the assessment score analysis was restricted to patients who had data at all three assessment visits (initial visit, six-month visit, and final visit). JSI conducted analysis for the following assessment scores: SF12 physical health composite score (SF12 PCS), SF12 mental health composite score (SF12 MCS), Patient Activation Measure (PAM), PHQ2, PHQ9 and Audit C. Findings from the descriptive analysis are summarized below.

SF12 PCS, SF12 MCS and PAM scores. SF12 measures functional physical and mental health and well-being from the patient's perspective. Scores range from 0-100, with a higher score indicating better physical/mental health functioning. PAM scores assess individual knowledge, skill, and confidence in managing health and health care utilization. Scores range from 0-100, with a higher score indicating better knowledge, skill, and confidence for managing one's health. For the cohort of patients (n=37) with SF12 scores at all three assessment visits, there was no meaningful change in SF12 PCS scores, while there was a slight improvement in SF12 MCS scores increasing from a mean of 36.6 at the initial visit to 39.9 at the final visit. For the cohort of patients with PAM scores at all three assessment visits (n=38), mean PAM score decreased from 63.8 to 57.9, indicative of a decline in patient knowledge, skill and confidence in managing one's health. Overall, while these data suggest that there may be some changes over time in patients' perceptions of quality and experience, the ability to attribute these changes to the IOPCM program is limited by both the small sample size and lack of comparable data for the control group.

Figure 11. Change in mean scores from initial to final visit for SF12 PCS, SF12 MCS and PAM



Note: Scores can range from 0-100 but the scale has been truncated at 70 in order to better depict the changes over time

Audit C scores. Audit C is a brief alcohol screen to identify patients who are hazardous drinkers or have active alcohol use disorders. Audit C is scored on a scale of 0-12. In men, a score of 4 or higher considered positive, optimal in identifying alcohol use disorders, while in women, a score of 3 or higher is considered positive. The higher the Audit C score, the more likely it is that the patient's drinking is affecting his/her health and safety. In the dataset provided to JSI, Audit C scores were available only for the first two visits (initial and six-month visit), and 15 men and 25 women patients had data for both visits. Data for these patients were analyzed. There were no changes over time in the proportion of patients with a positive Audit C score. For male patients, at both the initial and six-month assessment visit, 6% had a positive Audit C score (a score of 4 or higher) and 94% had a negative score (a score between 0 and 3). For women patients, at both the initial and six-month assessment visit 4% had a positive Audit C score (a score of 3 or higher) and 96% had a negative score (a score between 0 and 2). Given the small sample size and lack of data for a comparable control group, it is hard to conclude whether the program had an effect on patients' alcohol use disorders as measured by Audit C.

PHQ2 and PHQ9. PHQ2 is a depression screening tool to measure the frequency of depressed mood and anhedonia (interest or pleasure in doing things) over the past two weeks. A PHQ2 score ranges from 0-6, where a score of 3 or higher is considered a positive depression screen. PHQ9 is a tool to measure depression severity. A PHQ9 score ranges from 0-27, where a score of 15-19 is considered moderately severe and a score of 20-27 is considered severe. Typically PHQ2 is the first step in depression screening administered to all patients, followed by PHQ9 for depression severity assessment only in patients receiving a positive PHQ2 score (3 or higher). JSI analyzed PHQ2 and PHQ9 scores for patients with data at all three assessment visits. Per discussion with PHC, PHQ9 scores were categorized as scores less than 15 (not severe) and 15 or higher (severe).

Figures 12 and 13 depict changes in PHQ2 and PHQ9 scores over time. The data indicate that for the cohort of patients with PHQ2 scores for all three assessment visits (n=30), the proportion screening positive for depression (a score of 3 or higher) declined from 60% to 43%. However, among the cohort

of patients being assessed for depression severity using the PHQ9 assessment tool and with data at all three assessment visits (n=43), the proportion with severe depression (as score of 15 or higher) increased from 23% at the initial visit to 30% at the final visit. Caution must be exercised while making inferences from these data given the small sample sizes and the potential discrepancies in the ways these tools may have been administered across health centers. As earlier noted, typically PHQ2 is administered to all patients for depression screening while PHQ9 is administered only to patients with a positive PHQ2 screen (a score of 3 or higher). However, in the IOPCM sample, we observed that the cohort of patients with PHQ2 scores at all three assessment points (n=30) also received a PHQ9 assessment score irrespective of their PHQ2 score. Specifically, 9 patients had a PHQ2 score less than 3 at all of their three assessment visits but still received a PHQ9 assessment.

Figure 12. PHQ2 scores across three visits

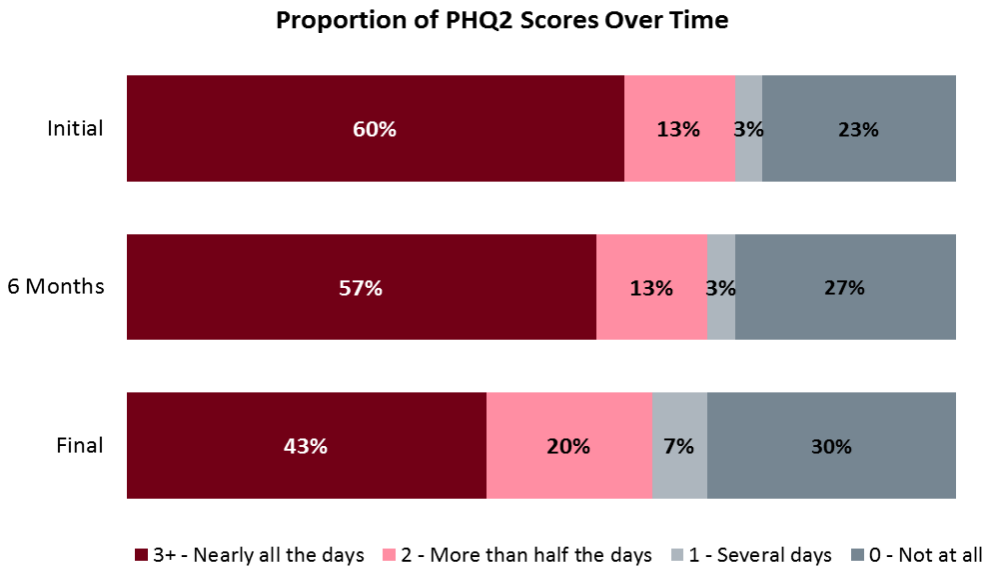
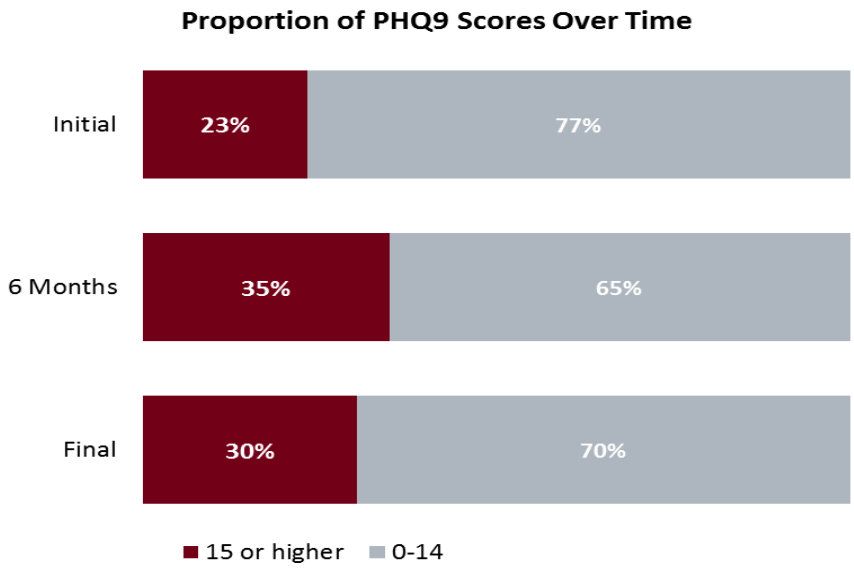


Figure 13. PHQ9 scores across three visits



Utilization Analysis for subgroups of interest

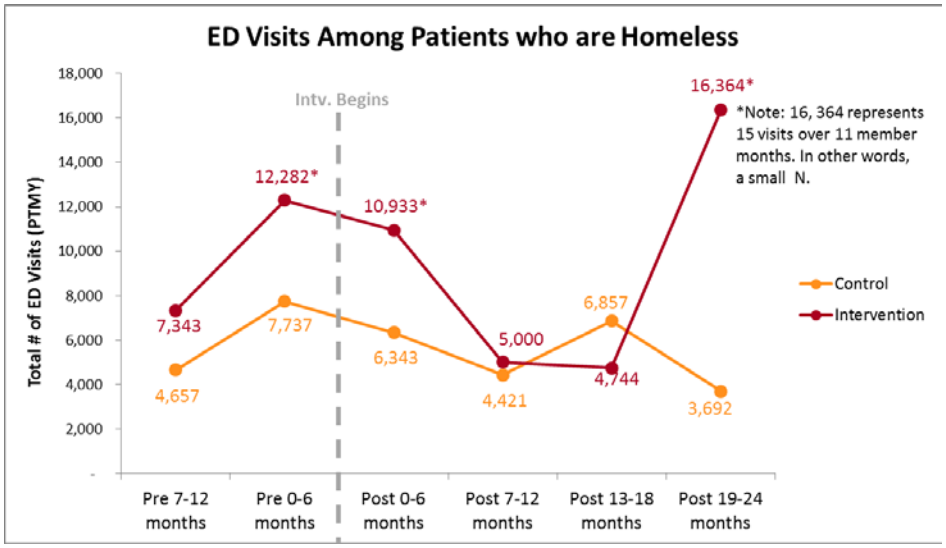
Per PHC's request, JSI conducted utilization analysis among identified subgroups of interest. The subgroups included those identified as homeless versus not homeless, those with and without substance abuse diagnosis, and those with and without diagnosis of the following disease conditions: congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), hypertension, and diabetes. Key findings from the subgroup analysis are summarized below.

Homelessness

Homelessness. Due to the small sample size of homeless patients, it is hard to make conclusions about IOPCM's impact on homeless patients. In the matched study sample, 28 patients were identified as homeless (15 intervention patients and 13 controls respectively).

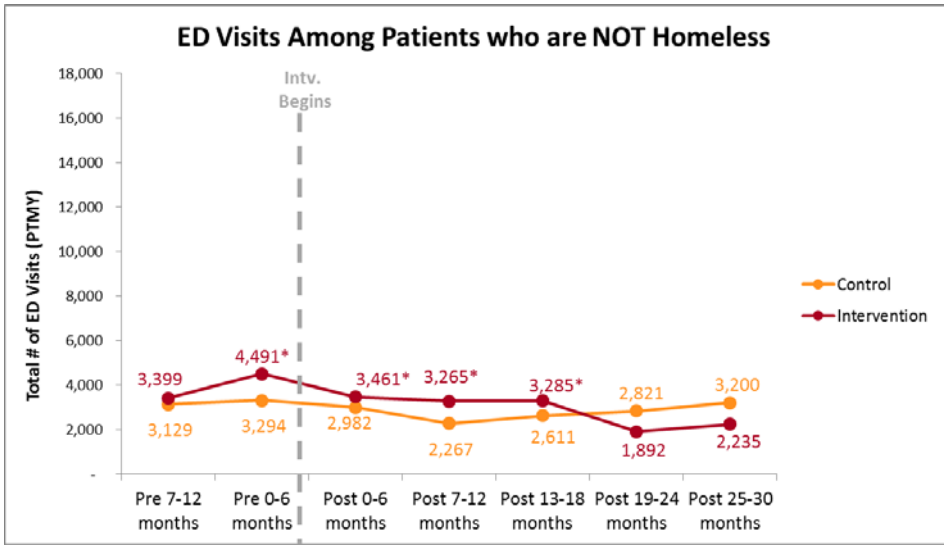
ED Visits. Figure 14 shows that for the subgroup of patients identified as homeless, ED visits PTMY decreased for both control and intervention patients in the post-intervention period, with a steeper decrease for intervention patients up to the post 13-18 months period, followed by a steep increase in utilization in the post 19-24 month period. Among patients who were **not** homeless (Figure 15), which is largely the full study sample for the IOPOCM outcomes analysis, ED visit patterns were, not surprisingly, similar to that of the full study sample. ED visits PTMY decreased consistently for intervention patients compared and controls, and the direction switched at the post 19-24 month interval with intervention patients' ED utilization PTMY being lower than that of controls. Given the small sample size of homeless patients, it is ill-advised to make inferences about the lack of impact of IOPCM on homeless patients.

Figure 14. ED visits per thousand member years (PTMY) among patients identified as homeless (15 intervention patients and 13 controls)



*p<.05

Figure 15. ED visits per thousand member years (PTMY) among patients not identified as homeless (247 intervention patients and 244 controls)



*p<.05

Inpatient admissions. Figure 16 shows that for the subgroup of patients identified as homeless, for the first year post-intervention, inpatient admissions declined for both intervention and control patients. At the start of the second year post-intervention the pattern shifted, inpatient admissions PTMY for controls increased while inpatient admissions PTMY for interventions declined to zero. Among patients who were **not** homeless (Figure 17), utilization patterns were, not surprisingly, very similar to that of the full study sample. Although the fact that inpatient admissions among patients who were homeless declined to zero is an important finding, caution should be exercised while making inferences given the small sample size for this subgroup analysis.

Figure 16. Inpatient admissions per thousand member years (PTMY) among homeless patients (15 intervention patients and 13 controls)

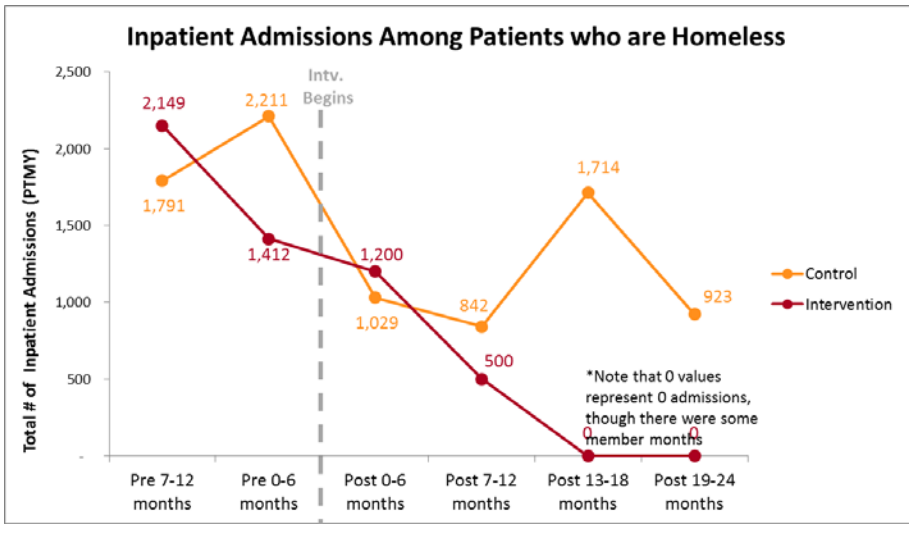
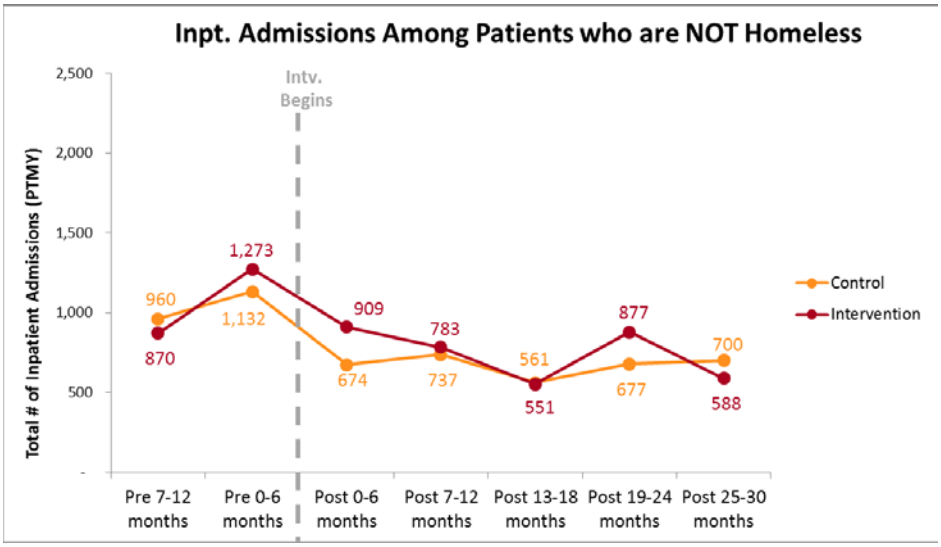


Figure 17. Inpatient admissions per thousand member years (PTMY) among not homeless (247 intervention patients and 244 controls)

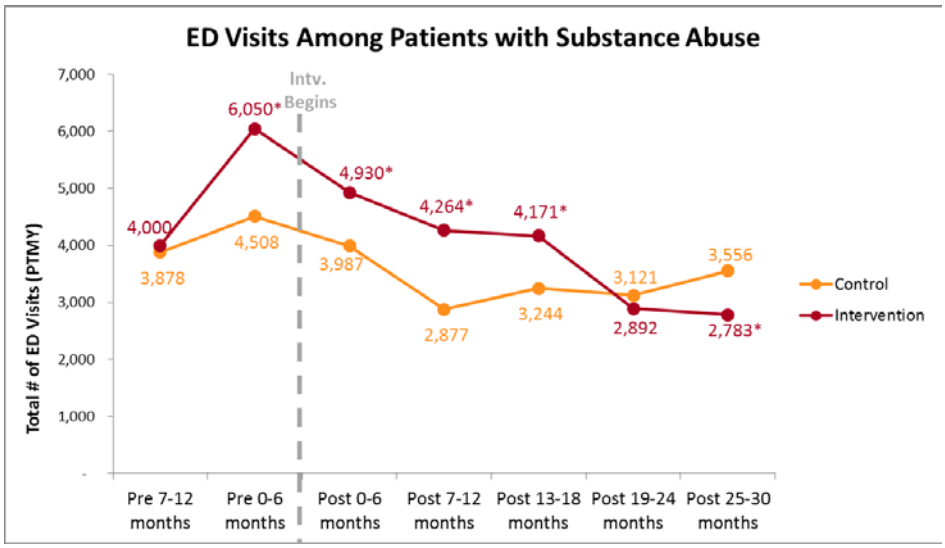


Substance Abuse

Substance Abuse. For ED visits, differential utilization patterns among patients with versus without substance abuse suggests that the intervention may take longer to impact patients with substance abuse.

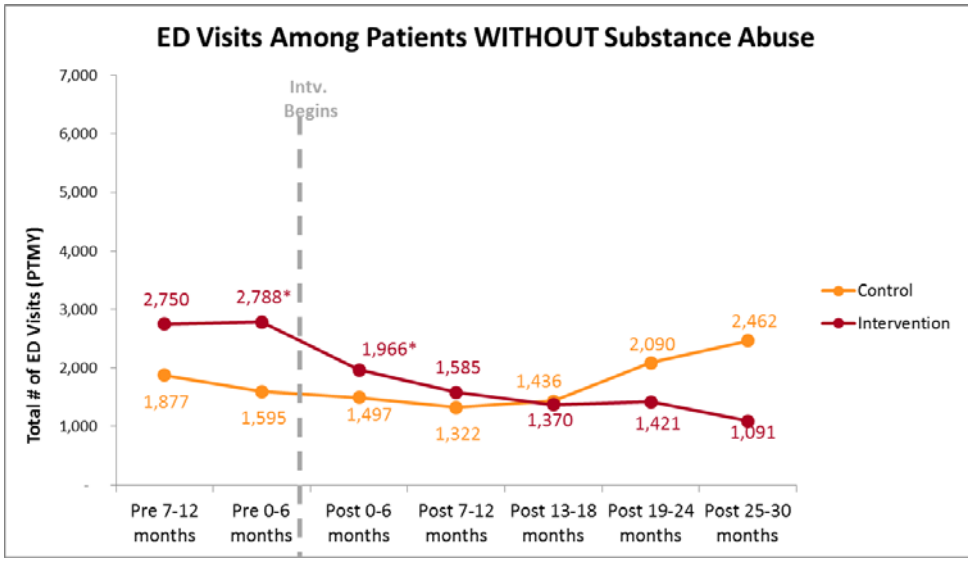
ED Visits. Two thirds of the study sample (66%) had ever received a substance abuse diagnosis. As such, the utilization patterns of patients with substance abuse diagnosis were fairly similar to that of the full study sample (similar to figure 1), with ED visits PTMY declining for both intervention and control patients over time, with the direction switching favorably for intervention patients in the post 19-24 month period. Comparing utilization patterns among intervention patients with versus without a substance abuse diagnosis (Figures 18 and 19), ED visits PTMY for intervention patients **with** substance abuse decreased to levels below that of controls in the post 19-24 month period, while ED visits PTMY for intervention patients **without** substance abuse decreased to levels below that of controls six month earlier (in the post 13-18 month period). This finding suggests that it may take longer for the IOPCM program to impact patients with substance abuse.

Figure 18. ED visits per thousand member years (PTMY) among patients with substance abuse (171 intervention patients and 172 controls)



*p<.05

Figure 19. ED visits per thousand member years (PTMY) among patients without substance abuse (91 intervention patients and 85 controls)

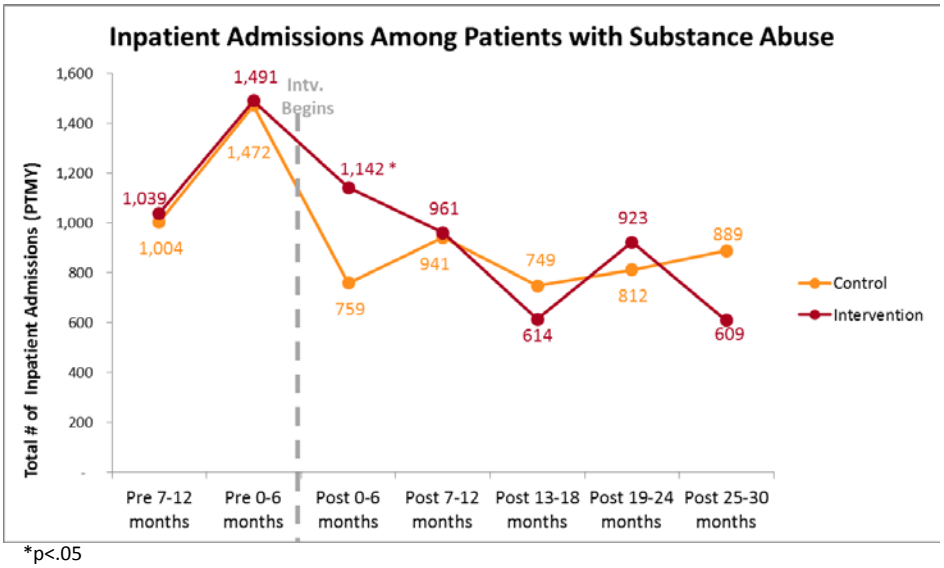


Substance Abuse. The most notable difference comparing patients with versus without substance abuse was lower inpatient admissions PTMY for patients without substance abuse.

Inpatient admissions. As stated above, two thirds of the study sample (66%) had ever received a substance abuse diagnosis, and, as such, the utilization patterns of patients with substance abuse diagnosis were fairly similar to that of the full study sample (similar to figure 2). Inpatient admissions for

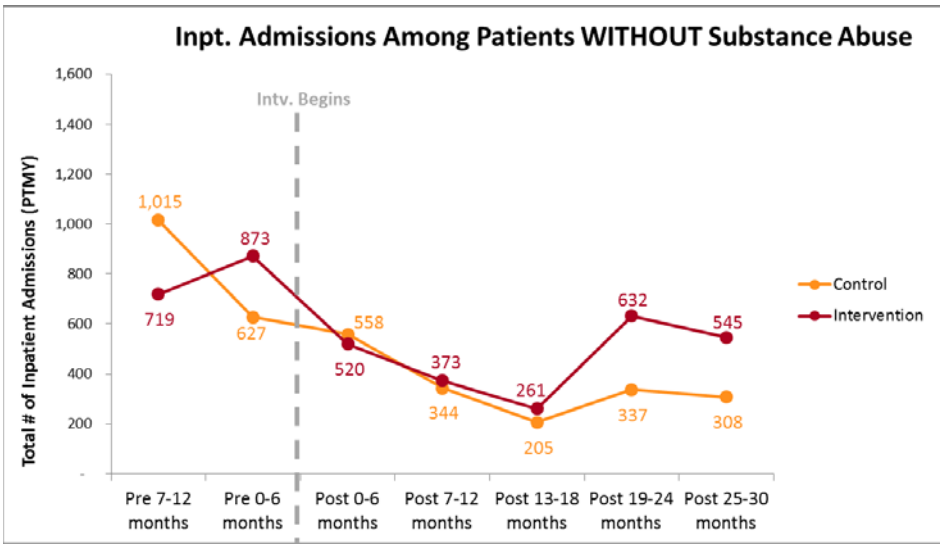
patients **with** substance abuse, for both control and intervention patients fluctuated over time, as shown in Figure 20. Inpatient admissions PTMY for patients **without** substance abuse (1/3 of the study sample) (Figure 21) followed similar patterns, but at each time interval the inpatient admission PTMY were lower in absolute terms for patients without substance abuse compared to those with substance abuse.

Figure 20. Inpatient admissions PTMY among patients with substance abuse (171 intervention patients and 172 controls)



*p<.05

Figure 21. Inpatient admissions PTMY among patients without substance abuse (91 intervention patients and 85 controls)



Disease Conditions

It is important to note that per discussion with PHC, disease conditions were not included as a matching variable in the propensity score analysis which was conducted to identify an appropriate control group for the outcomes evaluation. Table 8 shows the distribution of specific disease conditions across the two study groups. The data indicate that hypertension and diabetes were more prevalent among intervention patients; a statistically significantly higher proportion of intervention patients had ever received a hypertension (74%) or diabetes (54%) diagnosis as compared to controls. There was no difference between the study groups with respect to the prevalence of CHF and COPD.

Table 8. Distribution of disease conditions across study groups

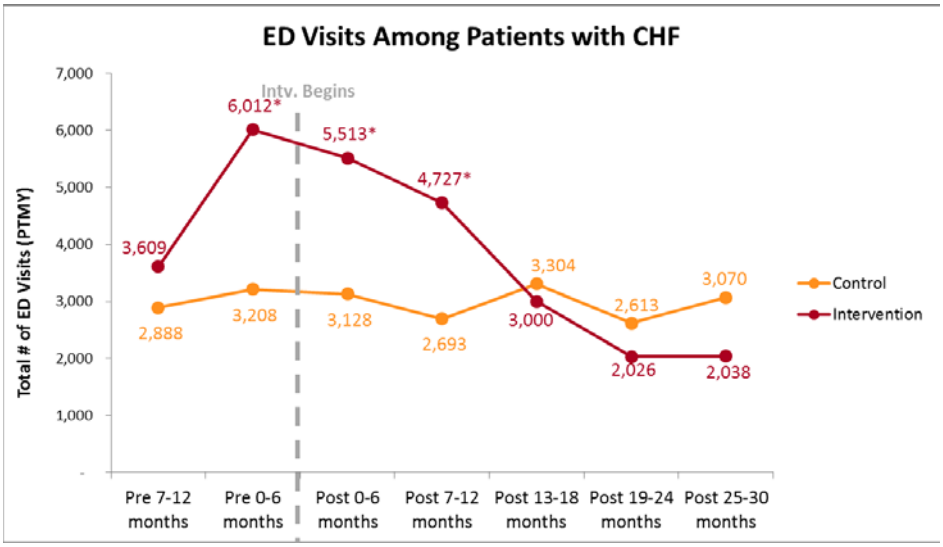
Disease Conditions	Intervention patients %	Control patients %
Congestive heart failure (CHF)	33	30
Chronic obstructive pulmonary disease (COPD)	65	62
Hypertension	74*	62
Diabetes	54*	41

*P<.05

CHF. A more consistent pattern of decreased ED visits over time among intervention patients with versus without CHF suggests that IOPCM may have a differential impact on ED visits based on the prevalence of CHF.

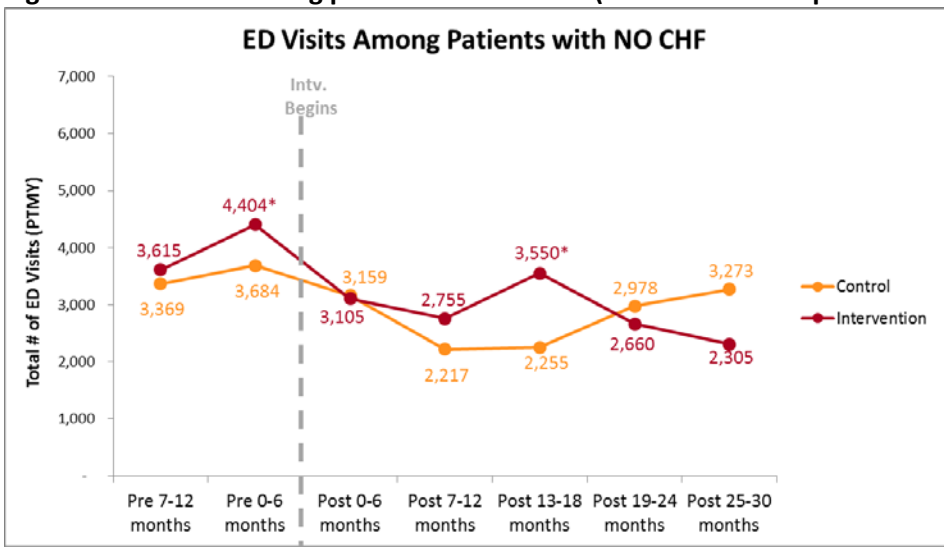
ED Visits. About a third of the sample had ever been diagnosed with CHF. Though ED visit PTMY started out higher for intervention patients **with** CHF compared to those **without** CHF in the pre-intervention period and first year post-intervention, over time there was a more consistent pattern of decreased ED visits PTMY for patients **with** CHF as compared to those **without** CHF. By the second year, intervention patients with CHF had fewer ED visits than their respective control group, and the downward trend continued over time. In contrast, ED visits among intervention patients **without** CHF fluctuated, declining at first, rising back up in the second year, and falling again in the next six-month period. Overall, the data suggest that for a clinically controllable condition such as CHF, the intervention has a favorable effect over time (Figures 22 and 23).

Figure 22. ED visits among patients with CHF (86 intervention patients and 78 controls)



*p<.05

Figure 23. ED visits among patients without CHF (176 intervention patients and 179 controls)

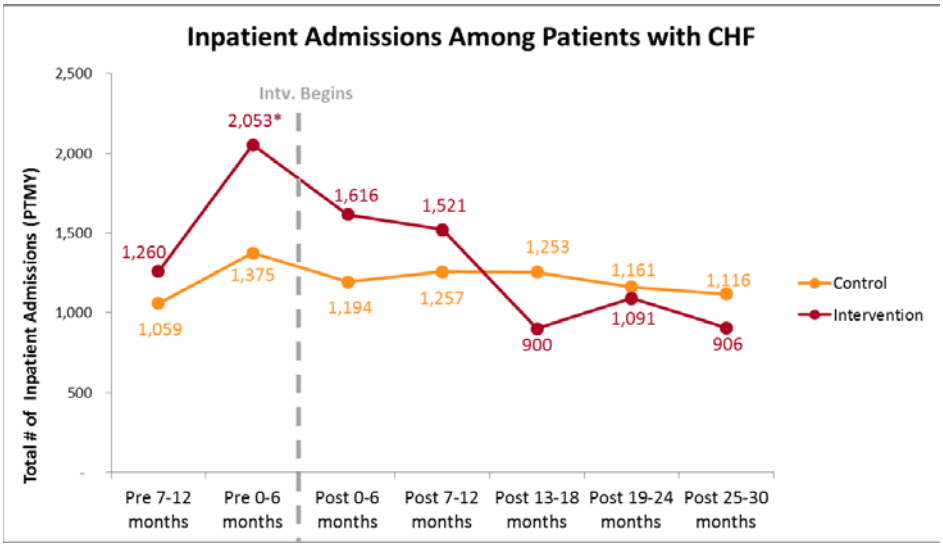


CHF. A more consistent pattern of decreased inpatient admissions over time among intervention patients with versus without CHF suggests that IOPCM may have a differential impact on inpatient admissions based on the prevalence of CHF.

Inpatient admissions. Comparing inpatient utilization trends among patients with versus without CHF, (figures 24 and 25) the data suggest that there was a more consistent pattern of decreased inpatient admissions among intervention patients **with** versus without CHF. Figure 24 shows that, among patients with CHF, intervention patients had steadily decreasing utilization from the pre 0-6 month period to the post 13-18 month period, at which point the intervention group's utilization becomes lower than that of the control group and remains lower for the remainder of the time studied. In contrast, for patients **without** CHF, inpatient admissions PTMY for interventions and controls tracked more closely and

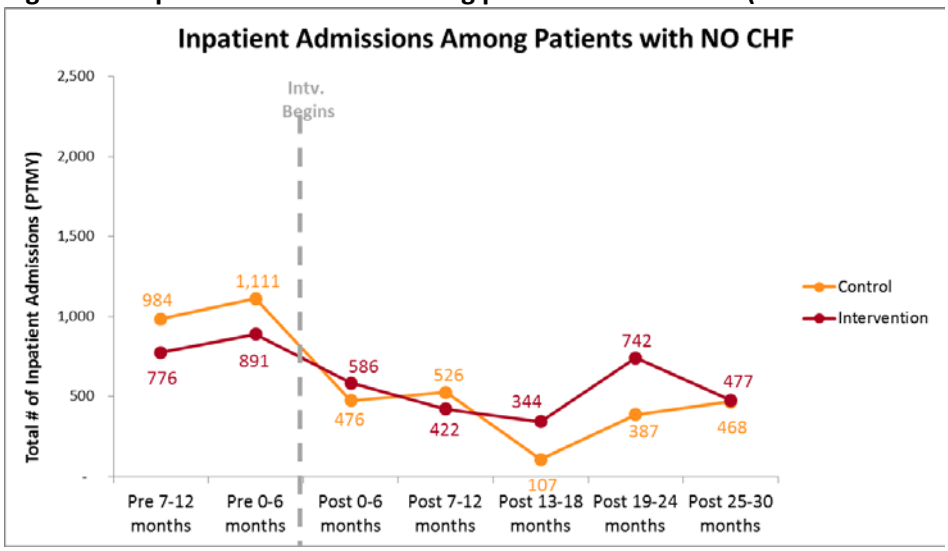
fluctuated over time, rising at the post 19-24 month period and dropping back down in the next six-month period but remaining higher than the levels in the post 13-18 month period (Figure 25).

Figure 24. Inpatient admissions among patients with CHF (86 intervention patients and 78 controls)



*p<.05

Figure 25. Inpatient admissions among patients without CHF (176 interventions and 179 controls)

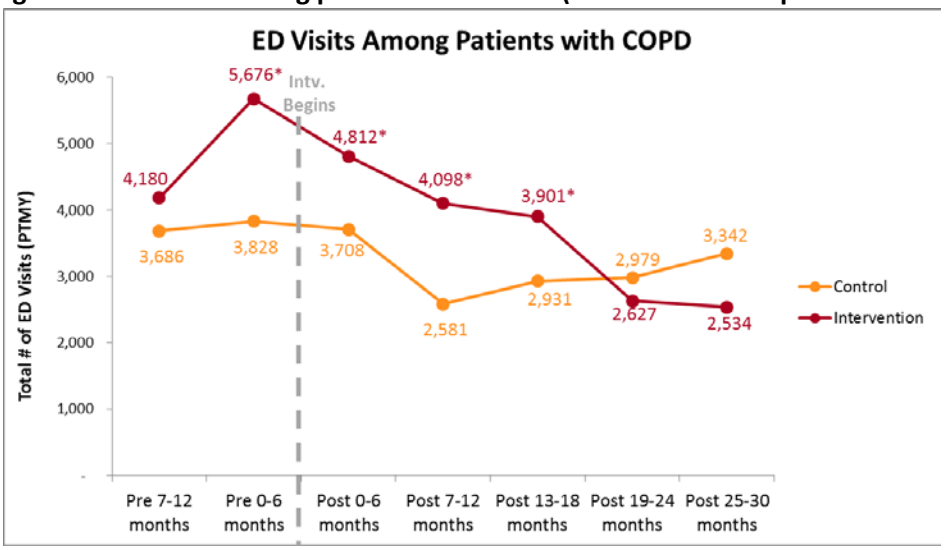


COPD. ED visits PTMY declined over time for all IOPCM patients relative to controls irrespective of COPD prevalence. However, intervention and control patients' utilization slopes crossed six months later for patients with versus without COPD suggesting that the intervention may take longer to impact patients with COPD.

ED Visits. With the majority of the study sample being diagnosed with COPD, utilization patterns among patients with COPD were similar to the full study population (similar to Figure 1). For intervention patients, after an initial increase from the pre 7-12 month to pre 0-6 month period, ED visits PTMY

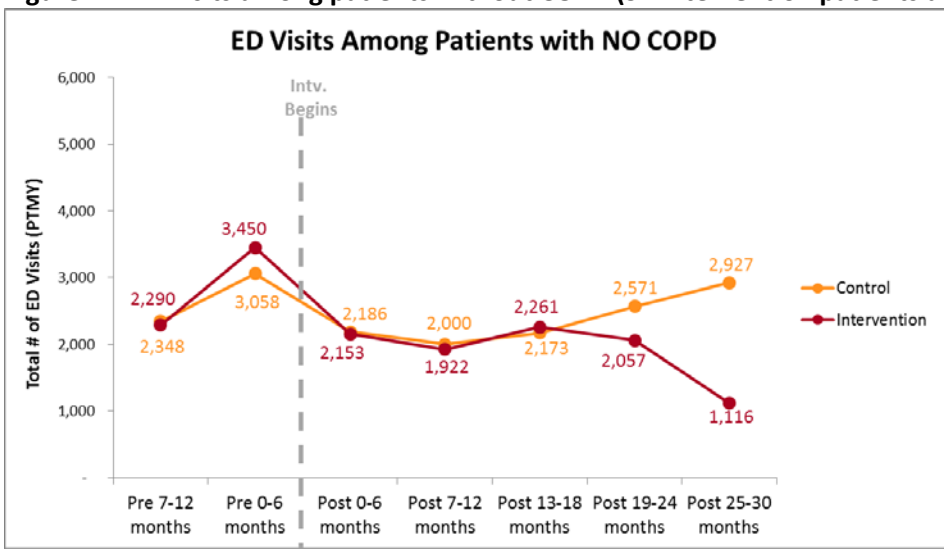
gradually declined in the post intervention period (Figure 26). Comparing utilization patterns among intervention patients **with** versus **without** COPD (Figures 26 and 27), however, we see that ED visits PTMY for intervention patients **with** COPD decreased to levels below that of controls in the post 19-24 month period, while ED visits PTMY for intervention patients **without** COPD were comparable to and then lower than that of controls' levels six month earlier (in the post 13-18 month period). This finding suggests that it may take longer for IOPCM to impact patients with COPD.

Figure 26. ED visits among patients with COPD (171 intervention patients and 160 controls)



*p<.05

Figure 27. ED visits among patients without COPD (91 intervention patients and 97 controls)



COPD. Inpatient admissions PTMY tracked closely for intervention patients and controls irrespective of COPD diagnosis suggesting that there is no differential program impact based on prevalence/absence of COPD.

Inpatient admissions. The utilization patterns among patients with COPD were similar to that of the full sample (Figure 28 is similar to Figure 2). Among patients with COPD, inpatient admissions PTMY tracked closely for controls and intervention patients. Over time, utilization for both interventions and controls decreased between 40-50%. Figure 29 shows inpatient admissions PTMY among patients **without** COPD; here, both groups saw a decrease in inpatient admissions PTMY in the post period up to post 13-18 months, at which point both groups experienced a sharp increase in their utilization. The fluctuation in both groups could be attributed to smaller sample size in the without COPD group.

Figure 28. Inpatient admissions among patients with COPD (171 intervention patients and 160 controls)

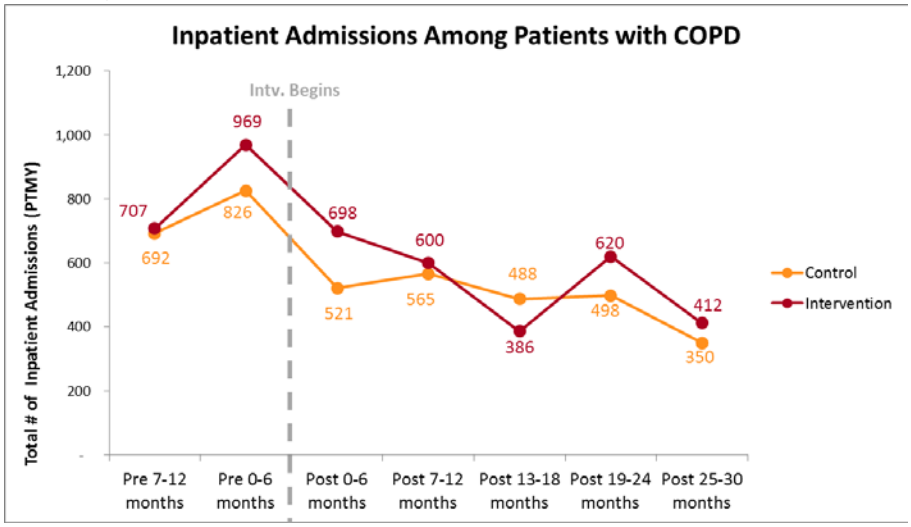
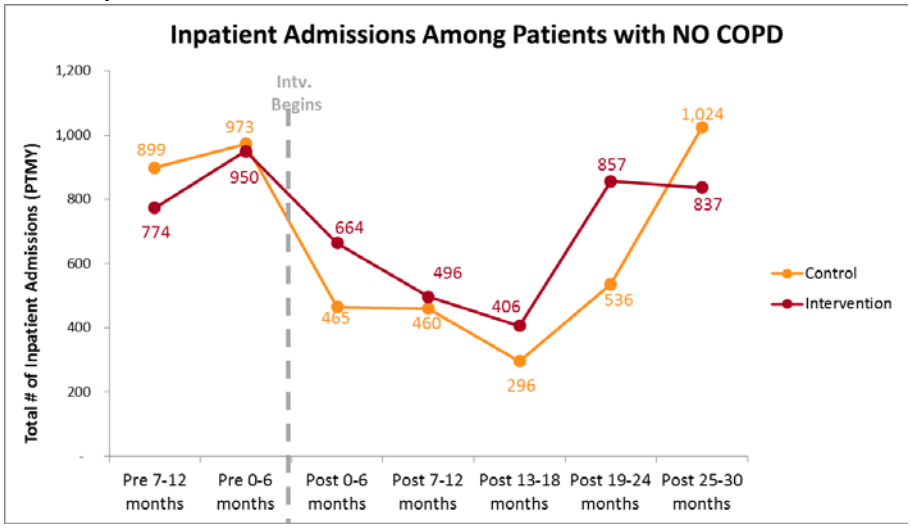


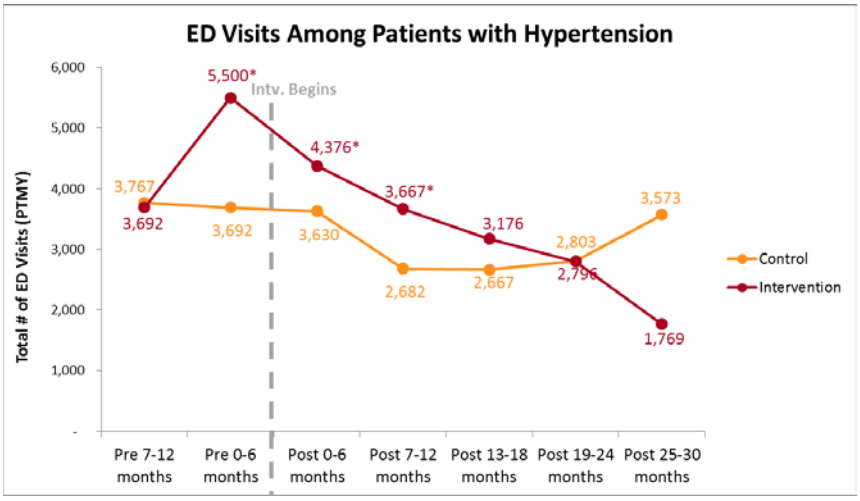
Figure 29. Inpatient admissions among patients without COPD (91 intervention patients and 97 controls)



Hypertension. A more consistent pattern of decreased ED visits over time among intervention patients with versus without hypertension suggests that IOPCM may have a differential impact on ED visits based on the prevalence of hypertension.

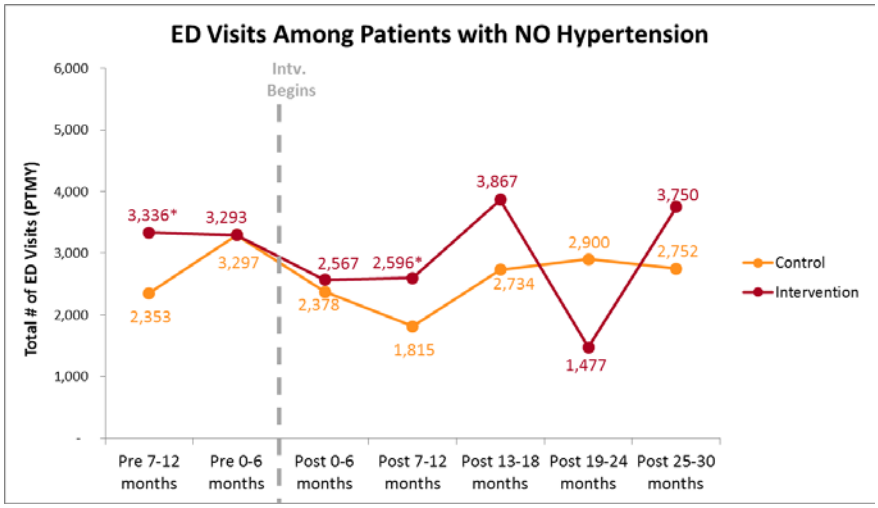
ED Visits. Two thirds of the control group and three fourths of the intervention group had ever received a diagnosis for hypertension. As such, ED utilization among patients with hypertension was similar to that of the full study population (figure 30 similar to figure 1). Importantly, over time there was a more consistent pattern of decreased ED visits PTMY for patients **with** versus **without** hypertension. By the post 19-24 month period, intervention patients with hypertension had fewer ED visits than their respective control group and the downward trend continued. In contrast, ED visits among intervention patients **without** hypertension fluctuated, declining at first and then rising back up in the second year and falling again in the next six-month period. Overall, the data suggest that for a clinically controllable condition such as hypertension, the intervention has a favorable effect over time (Figures 30 and 31).

Figure 30. ED visits among patients with hypertension (194 intervention patients and 158 controls)



*p<.05

Figure 31. ED visits among patients without hypertension (68 intervention patients and 99 controls)



Hypertension. A more consistent pattern of decreased inpatient admissions over time among intervention patients with versus without hypertension suggests that IOPCM may have a differential impact on inpatient admissions based on the prevalence of hypertension.

Inpatient admissions. Comparing inpatient utilization trends among patients **with** versus **without** hypertension, (figures 32 and 33) the data suggest that there was a more consistent pattern of decreased inpatient admissions among intervention patients **with** hypertension compared to intervention patients **without** hypertension, suggesting a favorable intervention effect. Figure 32 shows that, among patients with hypertension, intervention patients had steadily decreasing utilization. For controls with hypertension, after an initial decline, inpatient utilization PTMY started to increase in the post 13-18 month period and remained upward. Among patients **without** hypertension (figure 33), utilization patterns tracked closely for control and intervention patients until the post 19-24 month period, at which point the intervention group experienced a sharp increase in utilization and did not recover to its previously low levels in the post 25-30 month period. Controls also experienced an increase in utilization but not as dramatic. Some of this variability may be explained by smaller sample sizes for the without hypertension subgroup.

Figure 32. Inpatient admissions among patients with hypertension (194 interventions & 158 controls)

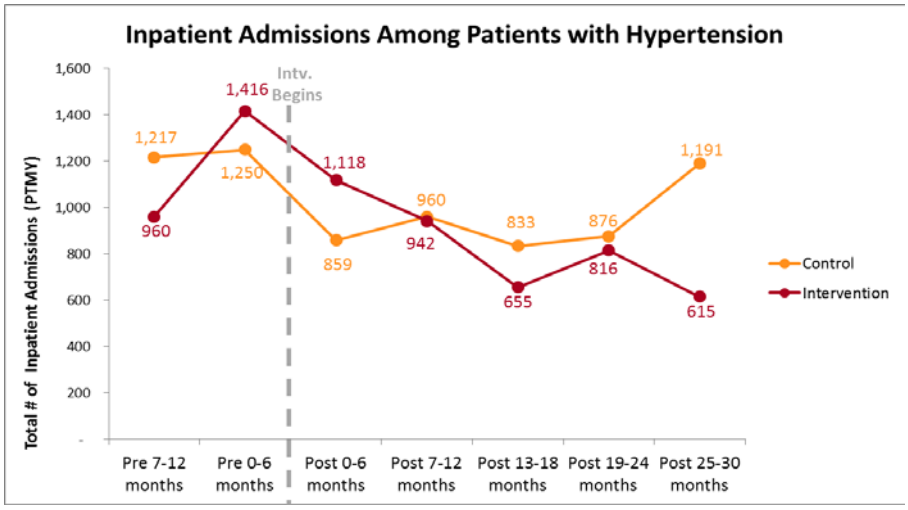
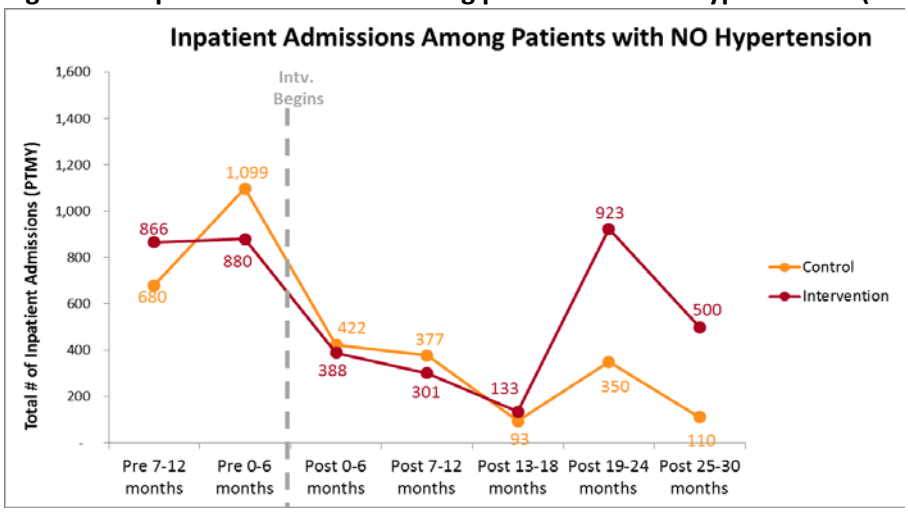


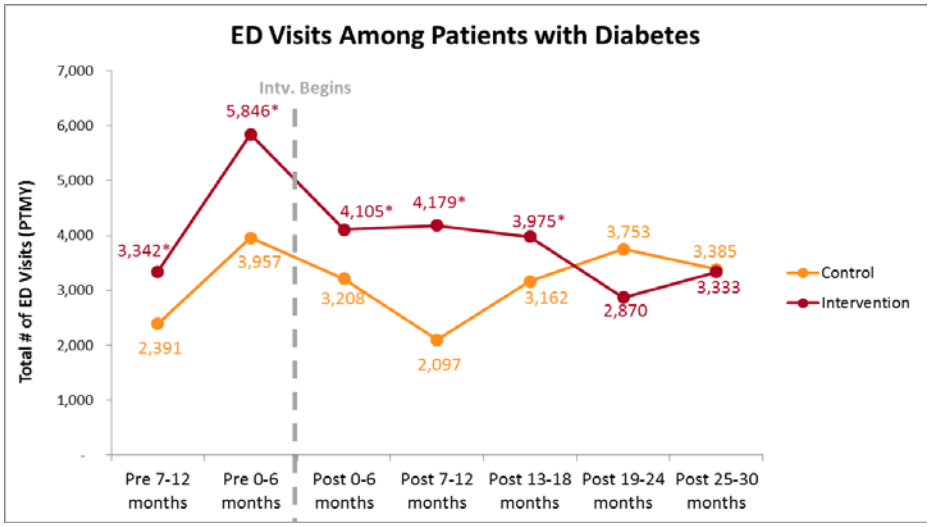
Figure 33. Inpatient admissions among patients without hypertension (68 interventions & 99 controls)



Diabetes. ED visits PTMY decreased for all intervention patients irrespective of diabetes diagnosis, however, over a two-year period, patients without diabetes may have experienced a steeper decline in utilization suggesting a slight differential program impact based on the absence of diabetes.

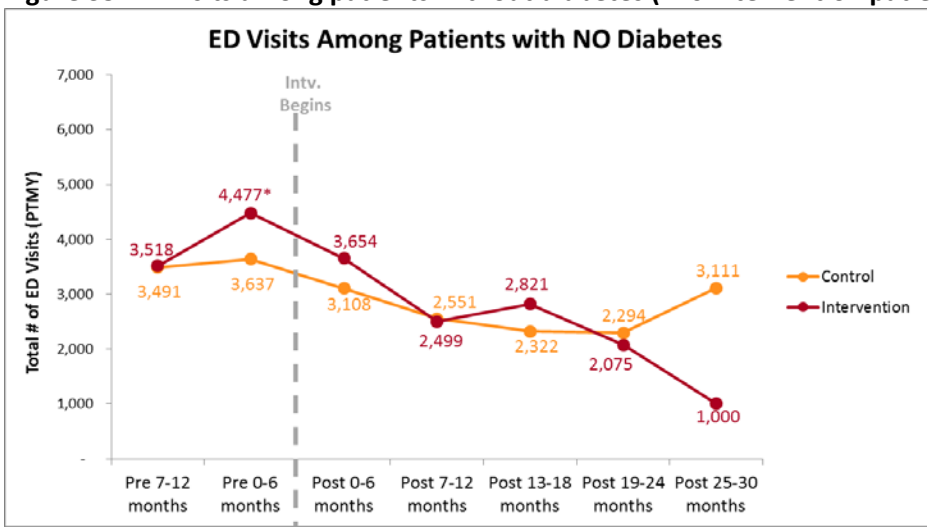
ED Visits. Among patients with diabetes, ED visits PTMY for intervention patients was higher than the control groups for the first year post intervention and then crossed over at the post 19-24 month period. Among patients **without** diabetes, ED visit utilization PTMY (Figure 35) tracked more closely for controls and interventions. Over the two-year period, ED visit utilization declined faster for intervention patients **without** versus **with** diabetes. By the 19-24 month period, however, ED visits PTMY for all intervention patients, irrespective of diabetes diagnosis, were lower than that of controls. But in the post 25-30 month period, intervention patients **with** diabetes saw a slight increase in ED visits PTMY, while patients **without** diabetes had a further drop in ED visits PTMY. The data suggest that the intervention may take longer to impact patients with diabetes.

Figure 34. ED visits among patients with diabetes (142 intervention patients and 104 controls)



*p<.05

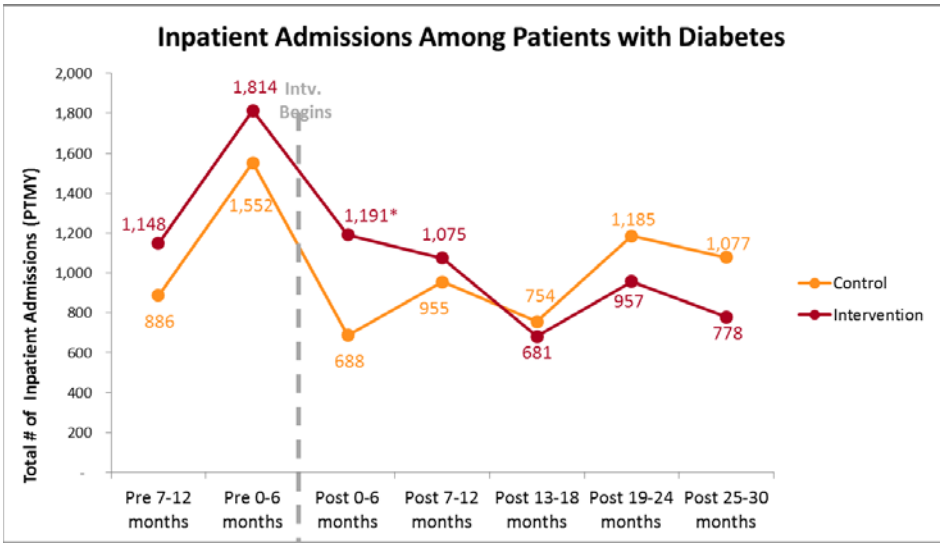
Figure 35. ED visits among patients without diabetes (120 intervention patients and 153 controls)



Diabetes. After a year in the program, intervention patients with diabetes had consistently lower utilization levels than controls, while utilization fluctuated for intervention patients without diabetes suggesting that there may be a slight differential program impact based on prevalence of diabetes.

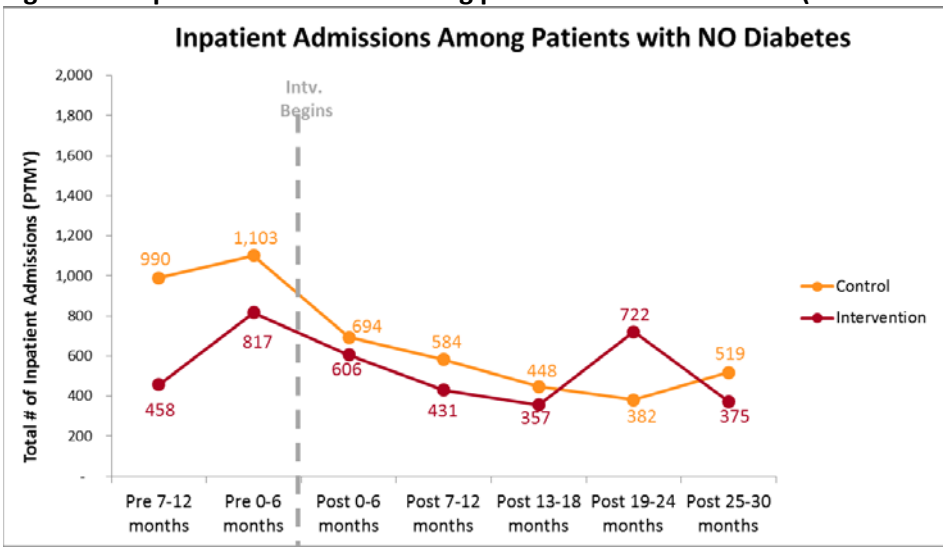
Inpatient Admissions. As seen in Figure 36, among patients with diabetes, inpatient utilization trends for control and intervention groups followed similar patterns of increase and decrease, but the intervention group ultimately had lower utilization than the control group beginning in the post 13-18 month period, which remained lower for the rest of the study duration. Among patients **without** diabetes (Figure 37), the intervention group consistently had lower utilization than the control group, except for a spike in the post 19-24 month period followed by a return to lower levels of utilization.

Figure 36. Inpatient admissions among patients with diabetes (142 interventions & 104 controls)



*p<.05

Figure 37. Inpatient admissions among patients without diabetes (120 interventions & 153 controls)



References

1. Harvey, S., et al., *Using clinical databases to evaluate healthcare interventions.* Int J Technol Assess Health Care, 2010. **26**(1): p. 86-94.
2. Valkenet, K., et al., *Effect of inspiratory muscle training before cardiac surgery in routine care.* Phys Ther, 2013. **93**(5): p. 611-9.
3. Pergolizzi, J.V., Jr., et al., *The prevalence of opioid-related major potential drug-drug interactions and their impact on health care costs in chronic pain patients.* J Manag Care Spec Pharm, 2014. **20**(5): p. 467-76.
4. Lin, T.T., et al., *Class effect of beta-blockers in survivors of ST-elevation myocardial infarction: A nationwide cohort study using an insurance claims database.* Sci Rep, 2015. **5**: p. 13692.
5. Sari, N. and M. Osman, *The effects of patient education programs on medication use among asthma and COPD patients: a propensity score matching with a difference-in-difference regression approach.* BMC Health Serv Res, 2015. **15**: p. 332.
6. Rosenthal, M.B., et al., *Effect of a multipayer patient-centered medical home on health care utilization and quality: the Rhode Island chronic care sustainability initiative pilot program.* JAMA Intern Med, 2013. **173**(20): p. 1907-13.
7. Rosenthal, M.B., et al., *Impact of the Rochester Medical Home Initiative on Primary Care Practices, Quality, Utilization, and Costs.* Med Care, 2015. **53**(11): p. 967-73.
8. Friedberg M.W., Rosenthal M.B., Werner, R. et al. *Effects of a Medical Home and Shared Savings Intervention on Quality and Utilization of Care*. JAMA Intern Med, 2015. **175**(8): 1362-1368.
9. Phillips, Robert L., et al. "Cost, utilization, and quality of care: an evaluation of Illinois' Medicaid primary care case management program." *The Annals of Family Medicine* 12.5 (2014): 408-417.
10. Frequent Users of Health Services Initiative. "Summary report of evaluation findings: A dollars and sense strategy to reducing frequent use of hospital services." *Oakland, CA: The California Endowment and the California Health Care Foundation* (2008).
11. Kyriacou, Demetrios N. "The Enduring Evolution of the P Value." *JAMA*315.11 (2016): 1113-1115.
12. CareOregon. "CareOregon and Community Clinics Partner to Improve Care for High Risk/High Cost Patients." The Health Resilience Program(TM). Lecture.
13. McWilliams, J. Michael, et al. "Early Performance of Accountable Care Organizations in Medicare." *New England Journal of Medicine* (2016).
14. Vickers, Andrew J., and Douglas G. Altman. "Analysing controlled trials with baseline and follow up measurements." *Bmj* 323.7321 (2001): 1123-1124.