Summer 7-15-2021

Multiple Baseline Interrupted Time Series: Describing Changes in New Mexico Medicaid Behavioral Health Home Patients' Care

Jessica Reno
University of New Mexico

Jessica Reno
Candidate

Mathematics and Statistics
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Yiliang Zhu, PhD, Chairperson

Caroline Bonham, MD

Erik Erhardt, PhD

MULTIPLE BASELINE INTERRUPTED TIME SERIES: DESCRIBING CHANGES IN NEW MEXICO MEDICAID BEHAVIORAL HEALTH HOME PATIENTS' CARE

by

JESSICA RENO

B.S. MOLECULAR BIOLOGY, TEXAS TECH UNIVERSITY, 2009
M.P.H. EPIDEMIOLOGY, EMORY UNIVERSITY, 2011

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Statistics

The University of New Mexico
Albuquerque, New Mexico

July 2021

ACKNOWLEDGEMENTS

The data for this study were provided by the New Mexico Human Services Department's (HHS) Medical Assistance Division. The author would like to acknowledge Sally Wait, Martha Payne, Steven Flint, Manuel Martinez, and Jason Martinez of the HHS Behavioral Health Services Division for planning and executing the data pull, and MaryAnn Shaening for evaluation guidance.

Caroline Bonham and Deborah Altschul have guided the evaluation and interpretation of results since 2015. Their dedication to improving behavioral healthcare in New Mexico has resulted in many innovative research and evaluation projects.

Multiple Baseline Interrupted Time Series: Describing Changes in New Mexico Medicaid Behavioral Health Home Patients' Care

by

Jessica Reno

B.S. Molecular Biology, Texas Tech University, 2009
M.P.H. Epidemiology, Emory University, 2011
M.S. Statistics, University of New Mexico, 2021

ABSTRACT

In 2016, the CareLink New Mexico behavioral health homes program began enrolling Medicaid recipients with the goal of increasing care coordination, improving access to services, and decreasing long-term costs of care for adults with serious mental illness (SMI) and children with severe emotional disturbance (SED). To evaluate these aims, a retrospective interrupted time series study using Medicaid claims data was designed. First, a comparable subset of non-enrolled individuals was selected from the pool of Medicaid recipients with SMI or SED using propensity score matching. Then, segmented regression was applied to three outcomes: total Medicaid charges, number of outpatient behavioral health claims, and incurring emergency care claims. Finally, difference-in-difference contrasts were estimated to compare the enrolled individuals' outcomes to their own baseline and to the trajectory of non-enrolled individuals. Enrollment resulted in decreased rate of increase in costs, decreased behavioral health claims, and decreased probability of emergency health care for enrollees.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Purpose

The purpose of this analysis is to utilize a multiple-baseline interrupted time series study design to quantify the impact of the CareLink New Mexico (CLNM) Behavioral Health Home program on the healthcare costs and utilization of Medicaid recipients with serious mental illness (SMI) or severe emotional disturbance (SED). The hypotheses addressed in the analysis are:

1. Program enrollment increases outpatient behavioral healthcare utilization;
2. Program enrollment decreases emergency healthcare utilization; and
3. Program enrollment decreases total Medicaid charges.

The challenges of this analysis are that the study was retrospective, only administrative data are available, not all information used to screen Medicaid recipients for potential enrollees was available, and there were many changes in the New Mexico Medicaid program leading up to program implementation. In this study, each of these challenges is addressed methodologically.

2.1 Health Homes

In 2016, the New Mexico Human Services Department (HSD) began implementing the CLNM behavioral health home program. Designated CLNM provider agencies deliver services to adults with SMI or children with SED in vulnerable populations of Medicaid recipients to enhance the integration and coordination of primary, acute, behavioral, social, and long-term services and supports such as housing, transportation, and employment. In 2018, the federal government passed legislation that allows health homes to expand enrollment to individuals with primary substance use disorders (SUD)[1]. However, New Mexico has not yet expanded enrollment to these individuals. Vulnerable populations include Medicaid recipients with chronic physical comorbidities and other behavioral health (BH) needs. The provider agencies are enabled to engage with patients through more direct relationships and intensive care coordination, which leads to comprehensive needs assessments and plans of care. The designated agencies also improve relationships between primary and specialty providers in order to improve the integration of care for patient service plans. The goals of CLNM are to promote acute and long-term health, prevent risk behaviors, enhance member engagement and self-efficacy, improve quality of life for individuals with mental health disorders, and reduce avoidable utilization of emergency department, inpatient, and residential services. Enrollment began on April 1, 2016, in Curry and San Juan counties, and expanded to Bernalillo, De Baca, Grant, Hidalgo, Lea, Quay, Roosevelt, and Sandoval counties on April 1, 2018. The goal of this study is to evaluate the impact of CLNM enrollment on emergency and outpatient behavioral healthcare and Medicaid charges. Additionally, impact on those with SUDs and be evaluated in order to inform the target population of CLNM going forward.

The study period for this evaluation begins January 1, 2014, and ends March 31, 2019. The year leading up to the study period, 2013, was eventful and tumultuous for New Mexico's Medicaid recipients and providers. BH Medicaid reimbursements in the state of New Mexico were frozen, leading to the closure of several BH clinics across the state and the loss of care for many clients, Medicaid and otherwise. Meanwhile, the Affordable Care Act began enrolling at the end of 2013, and New Mexico expanded Medicaid services to all individuals with incomes less than 138% of the federal poverty level at the beginning of 2014 [2]. Prior to expansion, only children and their mothers, pregnant women, the elderly, and the disabled with low income could enroll in Medicaid services. In the three years following, the total number of Medicaid recipients using BH services increased by over 31%, and payments from the state to providers for behavioral healthcare increased 47% [2]. Centennial Care, the state's modernized Medicaid program, was also established at the beginning of 2014. Two years later, on April 1, 2016, HSD began enrolling Medicaid recipients into CLNM [3].

According to HSD, among individuals with chronic physical conditions, those with mental health comorbidities have health care costs on average 60-75% higher than those without [2]. Individuals with SMI have a higher prevalence of physical comorbidities and a lower life expectancy by 25 years because they are more likely to engage in unhealthy behaviors and less likely to receive preventive care due to providers' preoccupation with their mental health symptoms and, as a result, require more acute healthcare services and experience poorer health outcomes [4]. One study of a behavioral health home program for Medicaid recipients found that enrollment led to reductions in total healthcare costs, reductions in emergency visits, and increases in outpatient visits compared to baseline and non-enrolled individuals [5]. In Maryland, enrollment in the behavioral health home program reduced the probability of emergency visits by 12% per year [6]. In Massachusetts, the total number of emergency visits decreased after enrollment, while visits among non-enrolled individuals increased during the same time period [7]. Similar results were found in studies of programs that integrated physical and mental health care, although they were not classified as health home programs [8].

2.2 Multiple Baseline Interrupted Time Series

The impact of CLNM on claims patterns cannot be adequately evaluated by comparing the average outcomes before and after enrollment because this analysis would not account for underlying trends in the outcomes during the study period. In this case, medical costs have been increasing in general for the past several years, including the study time period, so any decrease in cost may be overwhelmed by the general increase. Contrary to randomized controlled trials in other states, Romaire et al. found that healthcare utilization did not change and costs increased in their pre-post analysis of behavioral health homes in Maine. However, the trends over time are not described other than the average outcomes in the year prior and the year after enrollment [9]. A report on the first 13 health home programs in eleven states noted that the ability of health homes to gain and maintain enrollee engagement is a key factor in health home performance, indicating the importance of accounting for trends over time [10]. Interrupted time series (ITS) methods are more informative ways to study the impact of population-based programs and policies on outcomes over time because they provide a counterfactual trend to which one can compare the trend during the follow-up period. ITS differs from an analysis of variance (ANOVA) in that it quantifies and assesses within-group changes, comparing a group's own baseline to its repeated measures and summarizing the change over time[11]. Using ITS instead of ANOVA avoids some well-described biases of ANOVA that stem from ignoring the underlying trend and not accounting for the large variability in the outcome measurement and provides a more generalizable result [12–14]. Using the pre-intervention trend to calculate the counterfactual trend during the intervention period reduces the risk of bias in drawing conclusions about the intervention effect [15]. For example, if the outcome had been improving in the pre-intervention phase, the counterfactual trend can be used to measure the difference between where the outcome would have been on the same trajectory and where it actually was after the intervention. The alternative option is reversing the treatment (unenrolling the participants from the program) to observe the effect, which is typically not logistically possible, involves accounting for carryover effects, or is unethical, as in the case of healthcare policies and programs [16,17].

While the most effective way to reduce bias in a study is to randomize the participants, this was not possible for an evaluation of CLNM because enrollment began before the study, and participants were enrolled based on a set of criteria. Therefore, this study is a quasi-experiment, which means that there is an increased probability that unmeasured characteristics that may influence the outcomes are not distributed randomly among the intervention and comparison groups, introducing a systematic bias into the intervention effects[18]. ITS designs can be effectively applied to quasi-experiments, as well as randomized trials, and are a recommended approach when randomized controlled trials are not possible, such as the case of CLNM[19]. They are particularly useful in studies attempting to distinguish policy effects from time trends or differences among communities[20]. Additionally, recent evidence has suggested that re-analysis of randomized controlled trials using ITS may provide similar results[11]. This is excellent news for healthcare policy research since programs must often be evaluated in a short amount of time to support advocacy and decision-making.

However, longitudinal studies of average outcomes are often limited by their inability to assess the impact of events concurrent with the intervention on the outcomes of interest. This is an important consideration in policy research since multiple programs are often implemented around the same time in the same or overlapping populations to address a problem or take advantage of current funding availability. Additionally, variation in individual effects is lost or obscured in the average effects. A strategy to address both of these issues is to conduct multiple-baseline interrupted time-series (MB-ITS) analysis in multiple population units, each of which receives the intervention at a different point in time [18]. In the case of CLNM, each individual is enrolled at a different time, resulting in a non-concurrent multiple-baseline design, a term coined by Baer et al. in 1968[21]. The advantage of this design is the ability to demonstrate that a change in the outcome occurs when, and only when, the intervention is implemented[22]. Measuring the intervention effects at the individual level rather than strictly at the group level provides more information about the intervention effects[23]. MB-ITS allows for this design by including a random effect for each individual, instead of aggregating outcomes. MB-ITS requires that there be a clear delineation between the time periods before and after the intervention, such as the date of enrollment in CLNM[24]. It also requires several time points of historical data prior to the period of intervention that are comparable to the post-intervention data. Medicaid data meet these requirements since it was consistently and continuously recorded throughout the baseline and follow-up periods of the study. These aspects make the MB-ITS design a good fit for the CLNM evaluation study, because the goal of the program is to identify individuals whose healthcare utilization and costs could be dramatically improved with preventive and ongoing care for their diagnoses, resulting in a targeted effect of individuals, rather than an overall population effect. Additionally, reimbursement rates change frequently, and controlling for these changes over time by observing individuals with different intervention start dates improves the internal validity of the study and therefore increases the strength of the evidence of the effect of the intervention.

Before continuing, the terminology of the method being used should be addressed. While MB-ITS is the current name used for this method, it is misleading in two ways. First, "multiple baselines" refers to individual-specific baselines, and not multiple baselines per individual. In this study, each individual entered the study at a different time, based on their history of eligibility for Medicaid, and each case entered the intervention at a different time, based on their enrollment date. This is important to note because there are study designs in which individuals return to baseline, and sometimes back to the intervention again (reversal designs)[25]. Second, "time series" refers to a longitudinal study, and not to a time series model that incorporates seasonal variation and addresses autocorrelation issues.

The segmented regression model is used for ITS analysis, which is a widely accessible method to those familiar with regression that can be applied to linear or non-linear outcome trajectories[26–30]. When non-concurrent multiple baselines are included, a multilevel model is specified by incorporating a random effect for population units to account for the dependence of their measurements over time[31]. In the case of this analysis, a multilevel model is used to allow for a random effect of unique individuals with dependent measurements over time. Covariates will be added to the model to account for differences between individuals. Healthcare utilization generally decreases after childhood and dramatically increases in older age, while males are less likely to seek out preventive healthcare and more likely to utilize emergency care. The number of physical comorbidities, the presence of BH medication, and the type of behavioral disorders all impact the need for healthcare. The number of inpatient healthcare claims and total claims are indicators of the need for ongoing healthcare. The availability of healthcare varies across the state, so the residence of the CLNM member in a certain provider's county group impacts their healthcare utilization as well as that of their matched individual. The length of time eligible for study inclusion is a proxy for the length of time a person has been eligible for Medicaid, which impacts their utilization of healthcare. Monthly unemployment rates have an impact on Medicaid enrollment because one of the requirements of Medicaid eligibility is being under a certain income level based on the size of the household and the federal poverty guidelines. Finally, calendar dates are incorporated as covariates to control for temporal trends in Medicaid reimbursement[32].The assumptions of the model start with the assumptions for the type of regression being used. For linear models, these are: 1) each observation of the outcome can be expressed as a linear function of the independent variables, 2) error terms are normally distributed with an expectation of zero, 3) the covariance of error terms is zero, and 4) the independent variables are independent of the error terms. In time series models, the assumption that the covariance of error terms is zero will often be violated to some degree due to autocorrelation, or correlation of error terms with time. However, the actual structure of the covariance is very difficult, or impossible, to estimate in relatively short time periods. This study includes fewer than fifty measurements for each person, and therefore the potential autocorrelation is not possible to accurately estimate and the assumption that autocorrelation is the same across individuals will be made out of necessity. It should also be noted that the effects of autocorrelation on Type I error rates decrease with the number of time points [33]. For multilevel regression models used to assess MB-ITS, the analyst additionally assumes that the time series are independent of each other, meaning that one participant's behaviors do not impact each other [31]. We also must consider sample size, which, when small and non-randomized, can result in biased fixed-effect estimates. Fortunately, the number of enrolled and non-enrolled individuals in this study is large and inclusive, and therefore statistical inference methods are not expected to be impacted.

2.3 Propensity Score Matching

For ITS segmented regression models, the analyst also assumes that without the intervention, the pre-intervention trend would extend into the post-intervention period. One way to assess this assumption is to include a comparison population that does not experience the intervention[15]. When this is done, the analysis can include comparisons of the post-intervention trend to the counterfactual, as well as to the post-intervention comparison trend. Recent evidence demonstrates that ITS analyses that include comparisons might produce estimates very close to randomized controlled trials[34]. Including comparisons addresses historical bias directly by revealing any changes in the outcomes due to non-intervention events during the study period. Another advantage of the Medicaid claims data is the access to comparisons who were not enrolled in the program that can be matched to the cases.

Due to the retrospective nature of the study, the Medicaid recipients were not randomly enrolled in the CLNM program. Instead, Medicaid recipients with SMI or SED were screened using a list of questions including functional impairment, a requirement for assistance with activities of daily living, and cognitive deficits, which determined their eligibility. The answers to these screening questions were not available for this study. Therefore it was not possible to match them to comparison individuals directly on their likelihood of being enrolled. It is fair to assume that most of the potential comparisons for this study would not have met the criteria to be enrolled in CLNM, and therefore confounding characteristics may introduce bias into the analysis of changes in outcomes.

Matching Medicaid recipients according to potential confounders allows similar groups to be identified and reduces possible systematic variability between them, and therefore reduces treatment selection bias[18]. Potential confounders are characteristics thought to impact both the outcome and the probability of enrollment. Propensity scores are one way to calculate the odds of enrollment based on these characteristics and match enrollees to comparable individuals.Therefore, propensity scores were used to match enrolled individuals one-to-one to non-enrolled Medicaid recipients with a similar probability of enrollment based on available data[35]. A propensity score is defined as the conditional probability of exposure to treatment, given the observed covariates, and is estimated by regressing the treatment assignment on observed baseline characteristics using a logistic regression model[36]. The calculated score then represents the likelihood that a person would have been treated considering their characteristics, and the conditional distribution of confounders given the propensity score is similar for intervention and comparison participants[37]. The assumption is that enrollees and non-enrollees matched by score would have similar outcomes throughout the study period if no intervention were implemented. This is theorized to be true if 1) all potential confounders are included in the propensity score estimation model, 2) there are no individuals with a perfect propensity score (which ranges from 0 to 1), and 3) propensity scores are accurately estimated. The propensity scores are conventionally calculated using binary logistic regression to estimate the odds of being enrolled in the intervention given all potential confounders measured in the dataset. Confounders are related to treatment assignment and the outcome. However, confounders influenced by the intervention should not be used [38].

There are four approaches to using propensity scores in an analysis [36]. The first method is matching cases to comparisons by propensity score. Matching by propensity score creates an approximate balance in confounders between intervention and comparison participants, decreasing treatment selection bias. This method is employed when there are many more comparison providers than intervention providers. It is useful when there are many possible confounders, which makes stratified matching difficult. This method is also useful when treatment selection is based on factors that are not directly measured in the dataset. The second method is stratification, where participants are divided into strata by propensity score before performing pooled analysis. This method is used when the sample size is small and the analyst wants to include all participants. The third method is incorporating an inverse-probability-of-treatment weighted estimator into regression. The fourth method is incorporating the propensity score into the regression as a covariate. Because the propensity score provides information for multiple covariates, the model can be more parsimonious. In this study, the matching method was used to limit the number of comparison participants who had a very small probability of being enrolled. It is necessary here to assume that the measured confounders approximate comparability based on any unmeasured confounders, such as the screening questionnaire results [39].

Using a matched subset and segmented regression, this analysis will be conducted as a difference-in-difference study. The difference between the enrolled individuals' baseline and follow-up trends will be compared to the difference between the non-enrolled individuals' baseline and follow-up trends to determine the difference-in-difference. The treatment selection bias inherent in these data will be addressed using this quasiexperimental method and provide strong evidence of changes in outcomes for program coordinators.

3 Methods

3.1 Data Processing

Raw text files were generated by New Mexico HSD’s Medical Assistance Division by querying Medicaid claims and associated line items that were incurred between January 1, 2014, and March 30, 2019, among Medicaid recipients with existing behavioral disorders. The raw files were processed using SAS software, Version 9 of the SAS System for Windows (Copyright © 2013 SAS Institute Inc.). The analysis was conducted using R software, Version 4.0.5[40].

The claim line items were first categorized into 14 types of healthcare services. Figure 1 outlines the algorithm used to categorize line items using the type of claim, type of provider, provider’s specialty, and place of service. Pharmaceutical claims were categorized as BH medications or other medications using HSD-defined groups of drug therapeutic class codes for anti-anxiety medications and anti-depressants, mood-stabilizing anticonvulsants, minor tranquilizers, adrenergics, medications for attention deficit hyperactivity disorder, narcotic and alcohol antagonists, anti-mania medications, major tranquilizers and antipsychotics, and monoamine oxidase inhibitors.

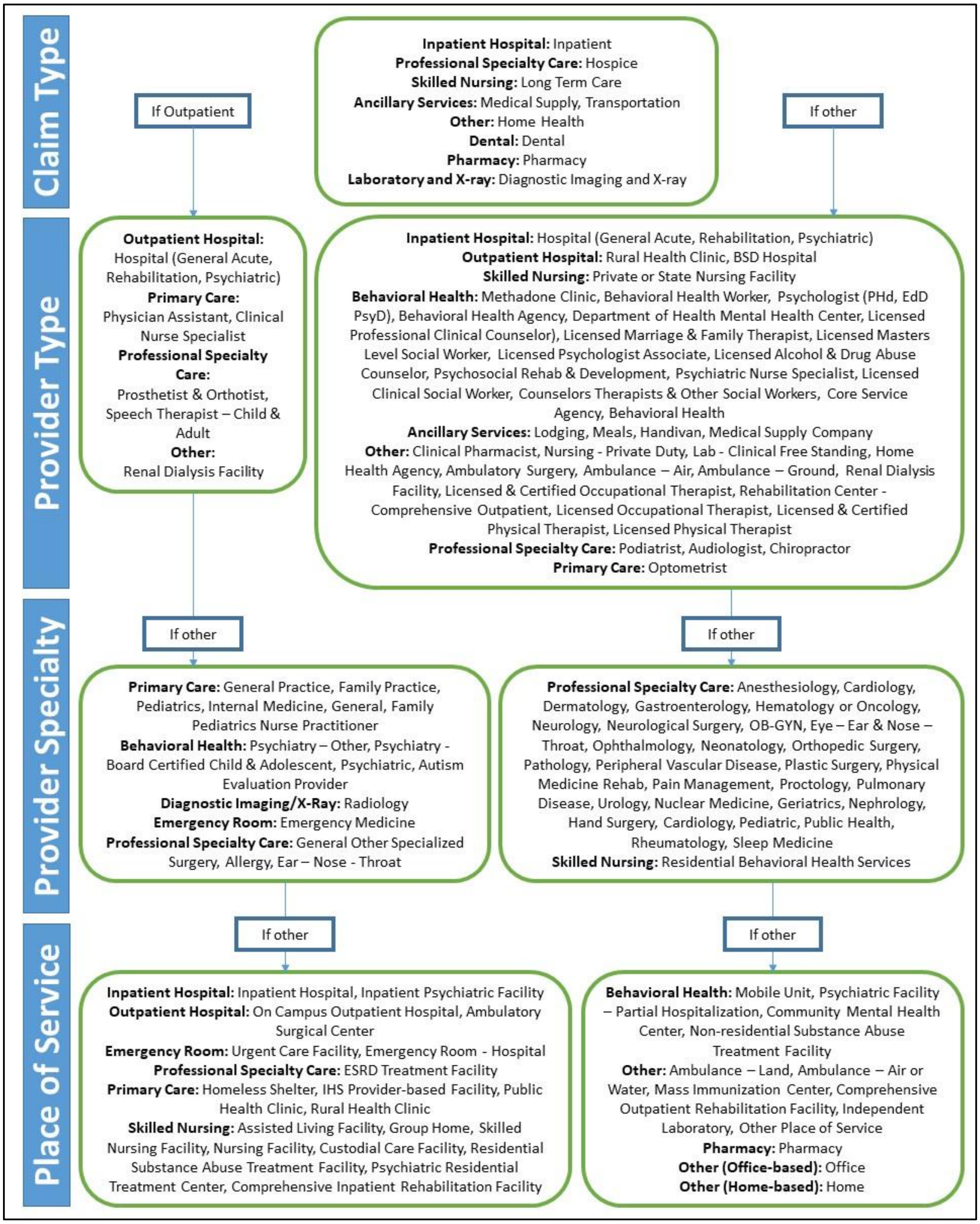


Figure 1. Algorithm for categorizing Medicaid claim line items into types of health care services. Beginning at the top, the claim type was used to categorize claims initially. If in "Outpatient" or an unlisted claim type, provider type was used. If still uncategorized, provider specialty was used. If still uncategorized, place of service was used. Finally, all uncategorized claims were included in "Other." Outpatient claim types followed a different algorithm than other claim types.

All claims for an individual client were searched for diagnoses in order to create an indicator of type of behavioral disorder. Behavioral disorders were identified using HSD-defined groups of International Statistical Classification of Diseases and Related Health Problems (ICD-9-CM, ICD-10-CM) and included mental disorders and substance use disorders (Table 1). Chronic physical comorbidities were identified using the Elixhauser Index-defined[41] groups of ICD-9-CM and ICD-10-CM and included congestive heart failure, cardiac arrhythmia, vascular disease, pulmonary circulation disorders, peripheral vascular disorders, hypertension without complications, hypertension with complications, paralysis, other neurological disorders, chronic pulmonary disease, diabetes without complications, diabetes with complications, hypothyroidism, renal failure, liver disease, peptic ulcer disease excluding bleeding, HIV/AIDS, lymphoma, metastatic cancer, solid tumor without metastasis, rheumatoid arthritis, coagulopathy, obesity, weight loss, fluid and electrolyte disorders, blood loss anemia, and deficiency anemia. The Elixhauser Index was abbreviated to exclude behavioral disorders because they were pre-defined using HSD definitions. Therefore, the index ranges from 0 to 27. One field for race and ethnicity was included in the Medicaid claims dataset. However, less than one percent of Medicaid recipients were “Hispanic,” which is unreasonable for the population of New Mexico. Therefore, the race and ethnicity field was determined to be unreliable and was not included in the analysis.

Table 1. Diagnoses defined as serious mental illness, severe emotional disturbance, and substance use disorder by the New Mexico Human Services Department.

Serious Mental Illness or Severe Emotional Disturbance	Substance Use Disorder
Depressive disorders	Opioid use disorder
Anxiety disorders	Alcohol use disorder
Trauma- and stressor-related disorders	Amphetamine use disorder
Neurodevelopment disorder	Cannabis use disorder
Bipolar and related disorders	Cocaine use disorder
Schizophrenia spectrum and other psychotic disorders	Other stimulant use disorder
Disruptive and impulse control and conduct disorders	Hallucinogen use disorder
Persistent depressive disorders	Other substance use disorder
Obsessive-compulsive related disorders	
Personality disorders	
Feeding and eating disorders	
Somatic symptom and related disorders	
Dissociative disorders	
Cyclothymic disorder	

The geographic location of the client's residence was not available in the dataset. Instead, the county of service was available for each claim. The most common county of service was used as a proxy for the client’s county of residence. Age and gender were also included in the Medicaid claims. When there were discrepancies in this information from claim to claim, the average age was calculated, and the most common gender was used. The charges were summed by client and quarter (3-month periods) and adjusted for inflation by applying the monthly consumer price index for all urban consumers[32]. When there were no claims during a given quarter for a client, but the client did have claims in before and after that quarter, the outcomes were set to zero. An individual’s inclusion in the study period started with their first Medicaid claim during the study period, and ended with their last claim during the study period.

MAD provided a list of the Medicaid recipients enrolled in the CLNM program, which began on April 1, 2016. This list was merged to the client data by Medicaid ID and included the start date(s) and end date(s) of enrollment for each client. However, the individual’s primary CLNM provider was not included in the list. Therefore, billing provider IDs were used to determine the provider of each CLNM member's Medicaid claims, and the most common provider of service was used as a proxy for the member's provider. There were nine CLNM providers in the state; two began in April 2016, one began in July 2018, and the rest began in April 2018.

Some claims were excluded from the analysis. First, some claim data had been shifted in the raw dataset creation, such that the value for one field was under the heading for another field. When possible, these claims were cleaned and shifted appropriately. When this was not possible, the claim was excluded. Second, some claims did not have dates of service and were excluded. Altogether, 1,605,612 of 52,268,726 claims (3.1%) were excluded. Third, individuals with no BH diagnosis during the study period were excluded entirely. Ultimately, there were 386,425 Medicaid recipients included in the analysis.

3.*2 Propensity Scores for Matching Enrolled and Non-enrolled Individuals*

Age, gender, county of residence, number of quarters included in the study, type(s) of behavioral disorders, and number of physical comorbidities were considered potential confounders due to their known relationship to healthcare utilization and treatment selection. A generalized linear model was fit with a binary logit-link to the outcome of CLNM enrollment using the the client-level dataset. Initially, all potential interactions between covariates were also added. Non-significant terms were only removed if the common support, meaning the scores of the comparison group covered the distribution of the enrolled group providing comparable matches, did not noticeably decrease. Otherwise, all terms were left in the model. Adequacy of the propensity score estimation was assessed by observing the distributions of the propensity scores in the two treatment groups, and the distribution of the covariates by propensity score. Enrolled individuals were matched one-to-one to a non-enrolled individual using the nearst propensity score. The final matched pairs were determined using nearest neighbor matching. The date of the enrolled individual in the matched pair was also as the post-intervention start date for the comparison individual, and the pre-intervention period included all claims prior to that start date. The calendar date was retained in order to control for temporal trends in outcomes.

3.3 Models

ITS models for each hypothesis listed below were specified prior to the analysis (Table 2). These models capture the hypothesized relationship of each outcome with enrollment in CLNM and time. Although the target population is individuals with SMI or SED, total healthcare expenditures are higher for individuals with behavioral disorders than those without, primarily due to higher physical healthcare expenditures[42]. One of the main goals of behavioral health homes is to coordinate care in order to prevent the need for these expenditures[43]. Therefore, total charges were used as the first outcome. An indicator of any emergency care claim was used instead of the number of claims because the average interarrival time for emergency claims was 112 days, which is longer than the unit of time for the study, which was a quarter (three months, or 90 days).

Table 2. Outcome measures, number of events, and spacing of events over time for each hypothesis.

Outcome Measure	**Hypothesis**	**Individuals without the Outcome Event**	**Days between Outcome Events (Mean (Median) [IQR])**
Total Charges from All Claims	Charges will decrease after enrollment.	620 (0.2%)	11 (4) [2-9]
Number of Outpatient Behavioral Healthcare Claims	Number of outpatient behavioral healthcare visits will increase after enrollment.	134,457 (34.8%)	21 (7) [4-14]
Any Emergency Health Care Claim	Probability of emergency health care need will decrease after enrollment.	168,288 (43.5%)	112 (42) [11-128]

To begin specifying the model, we first acknowledge the expected outcome *E(Y)* at time *t* is estimated by either the intervention trend or the comparison trend based on individual *i*'s enrollment status:

$$E(Y_{it}|Status) = E(Y_{it}|Intervention)\ if\ status = Intervention \tag{1}$$

or

$$= E(Y_{it}|Comparison)\ if\ status = Comparison$$

In this particular analysis, the intervention trend $E(Y_{it}|Intervention)$ is estimated by the measurements taken from individuals enrolled in the CLNM program, and the comparison trend $E(Y_{it}|Comparison)$ is estimated by the measurements taken from non-enrolled individuals. Both groups have measurements during the baseline and follow-up periods. The baseline period begins January 1, 2014 and ends at the time of the matched pair's CLNM enrollment date. The follow-up period begins at the time of the matched pair's CLNM enrollment date and ends on March 30, 2019. The model will estimate the changes from baseline trend to follow-up trend within each group, and make it possible to quantify the difference in the within-group change between the intervention and comparison group (difference-in-difference). For client i, the baseline period spans the time from the beginning of their baseline measures to the date they were enrolled in CLNM, τ_{i0} (Figure 2). Following τ_{i0} is a latency period of time (τ_{i1} - τ_{i0}) before the outcome is expected to change due to the intervention. The follow-up period spans the time from the end of the latency period, τ_{i1}, to their last measurement in the study.

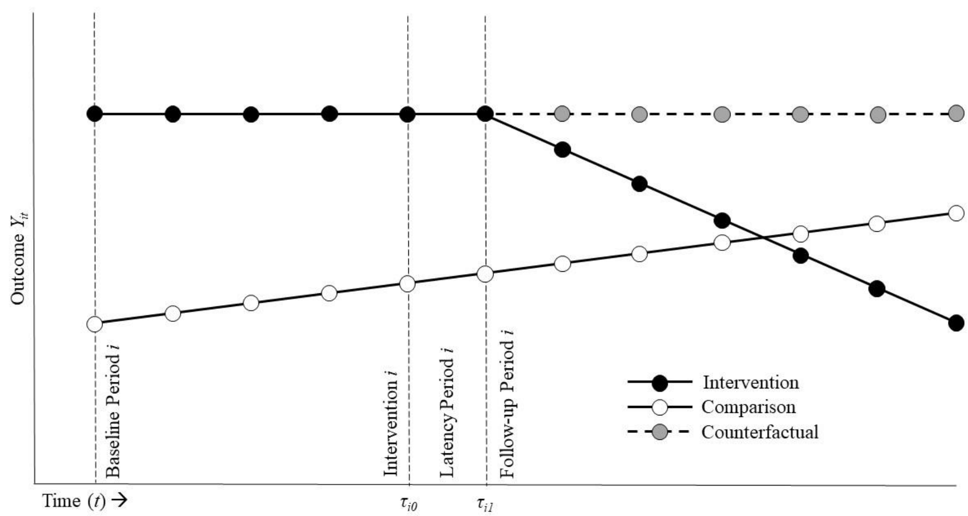


Figure 2. Illustration of a hypothetical outcome Y for client i modeled using segmented regression.

The intervention trend is estimated by the enrolled individuals' measures during the pre- and post-intervention periods,

$$E(Y_{it}|Status = Intervention) = f_{10}(t - \tau_{i0}) \times I(t \leq \tau_{i0}) + f_{11}(t - \tau_{i1}) \times I(t \geq \tau_{i1}) \quad (2)$$

where $f_{10}(t)$ represents the baseline trend's function, and $f_{11}(t)$ represents the follow-up trend's function. The intervention baseline trend estimated from the model can be extrapolated to estimate the counterfactual during the follow-up time period, if the enrolled individuals had not been enrolled in CLNM.

$$E(Y_{it}|Status = Counterfactual) = f_{10}(t - \tau_{i1}) \times I(t \geq \tau_{i1}) \quad (3)$$

Similarly, the comparison trend is estimated by the unenrolled individuals' measures during the baseline and follow-up periods, where the baseline (*$f_{20}(t)$)* and follow-up (*$f_{21}(t)$)* trends are assumed to have different functions than for the intervention trend:

$$E(Y_{it}|Status = Comparison) = f_{20}(t - \tau_{i0}) \times I(t \leq \tau_{i0}) + f_{21}(t - \tau_{i1}) \times I(t \geq \tau_{i1}) \quad (4)$$

Next, the baseline and follow-up trends for the intervention and comparison groups are specified. While this is an iterative step during the model building process, this section shows the illustrative case of a linear trend for each period and different intercepts and slopes for each trend. Additional covariates and polynomial terms may be added to these functions to improve the model fit:

$$f_{10}(t - \tau_{i0}) = \beta_{100} + \beta_{101} \times (t - \tau_{i0}) \qquad t \leq \tau_{i0} \quad (5)$$

$$f_{11}(t - \tau_{i1}) = \beta_{110} + \beta_{111} \times (t - \tau_{i1}) \qquad t \geq \tau_{i1} \quad (6)$$

$$f_{20}(t - \tau_{i0}) = \beta_{200} + \beta_{201} \times (t - \tau_{i0}) \qquad t \leq \tau_{i0} \quad (7)$$

$$f_{21}(t - \tau_{i1}) = \beta_{210} + \beta_{211} \times (t - \tau_{i1}) \qquad t \geq \tau_{i1} \quad (8)$$

where

β_{200} is the intercept of the pre-intervention trend among non-enrolled individuals,

β_{100} is the intercept of the pre-intervention trend among enrolled individuals,

β_{210} is the intercept of the post-intervention trend among un-enrolled individuals,

β_{110} is the intercept of the post-intervention trend among enrolled individuals,

β_{201} is the slope of the pre-intervention trend among un-enrolled individuals,

β_{101} is the slope of the pre-intervention trend among enrolled individuals,

β_{211} is the slope of the post-intervention trend among un-enrolled individuals,

β_{111} is the slope of the post-intervention trend among enrolled individuals, and t_i is the time relative to τ_{i0}.

Finally, using equation (1), we specify the first level of the full ad hoc model:

$$
\begin{aligned}
Y_{it} &= E(Y_{it}|Status) + e_{it} \qquad (9)\\
&= [f_{10}(t-\tau_{i0}) \times I(t \le \tau_{i0}) + f_{11}(t-\tau_{i1}) \times I(t \ge \tau_{i1})]\\
&\times I(Status = Intervention)\\
&+ [f_{20}(t-\tau_{i0}) \times I(t \le \tau_{i0})\\
&+ f_{21}(t-\tau_{i1}) \times I(t \ge \tau_{i1})] \times \big(1 - I(Status = Intervention)\big)\\
&+ e_{it}\\
&= [(\beta_{100} + \beta_{101} \times (t-\tau_{i0})) \times I(t \le \tau_{i0})\\
&+ (\beta_{110} + \beta_{111} \times (t-\tau_{i1})\ ) \times I(t \ge \tau_{i1})] \times I(Status = Intervention)\\
&+ \big[\big(\beta_{200} + \beta_{201} \times (t-\tau_{i0})\big) \times I(t \le \tau_{i0})\\
&+ \big(\beta_{210} + \beta_{211} \times (t-\tau_{i1})\big) \times I(t \ge \tau_{i1})\big]\\
&\times \big(1 - I(Status = Intervention)\big)\\
&+ e_{it}
\end{aligned}
$$

The second level of the model, which allows the regression coefficients to vary randomly across participants, is specified as:

$$\beta_{kmni} = \beta_{kmn} + \gamma_{kmni} \qquad (10)$$

where k = 1,2; m = 0,1; n = 0,1; and γ_{kmni} are the random effects associated with each individual for the specific regression coefficient, respectively. Specifically, we assume {γ_{kmni}, I = 1,…,N_{status} } follow a normal distribution with mean 0 and variance σ_{kmn}.

All potential covariates and interactions with ITS parameters specified in Equation (9) and type of BH disorder (mental disorder only, substance use disorder only, or both) will be incorporated into the first level of the models. In the case of a linear trend, the trajectories for intervention and comparison groups (i=1,2) at baseline or follow-up (j=0,1) can be expressed as

$$f_{ij}\left(t, X_t b_{ij}\right) = \beta_{ij0} + \beta_{ij1} \times t + X_t b_{ij} \tag{11}$$

where X_t is a N_i x p covariate matrix, including potentially time-varying covariates as discussed below, and b_{ij} is the p-vector of regression coefficients.

The base model will include only the ITS parameters specified in Equation (9). Total charges will be modeled using linear regression and a binary indicator of any emergency care claim will be modeled using logistic regression. The number of BH claims is a count outcome, so Poisson regression will be applied and assessed for overdispersion. If overdispersion is detected negative binomial regression will be assessed for goodness-of-fit, since claims data typically exhibits zero inflation. The Akaike Information Criterion (AIC) will be used to identify the most appropriate regression method. Once this step is completed, all potential covariates and interactions with ITS parameters specified in Equation (9) and type of BH disorder (mental disorder only, substance use disorder only, or both) will be incorporated into the first level of the models. Basic splines will be incorporated for continuous covariates, and AIC will be used to determine the number of degrees of the polynomial. Once the splines are determined, non-significant terms will be dropped in a backwards stepwise fashion using analysis of variance until only significant terms remain. Random intercepts for matched pair and client will be incorporated one at a time and assessed for fit using the likelihood ratio test or test of deviance.

Potential covariates are gender, age, number of physical comorbidities, BH medication, behavioral disorders, number of inpatient healthcare claims, number of total claims, provider county group, eligible time in the study, unemployment rate, and calendar date.

Difference-in-difference tests determine whether the difference that enrolled individuals experienced was significantly different than that experienced by the non-enrolled individuals. In this case, we are most interested in the differences between the counterfactual and observed, compared by enrollment. The emmeans package in R was used to calculate estimated marginal means using final model results and test contrasts.

4 Analysis

4.1 Individual Characteristics

Table 3 describes the characteristics of all Medicaid recipients with SMI or SED in New Mexico between January 1, 2014, and March 31, 2019, by enrollment in CLNM. CLNM enrolled a larger proportion of Medicaid recipients who were eligible for the entire study period (44% vs. 21%) and a larger proportion with both mental and substance use disorders (44% vs. 18%) than the non-enrolled Medicaid recipient population. Of course, enrolled individuals were more likely to live in CLNM provider county groups than non-enrolled individuals (93% vs. 49%). Providers in Curry and San Juan counties began serving CLNM enrollees in 2016, and the other providers began in 2018. As a result, CLNM enrollees disproportionally represent these 2 counties. However, only 2 of New Mexico's 33 counties are not represented by at least one CLNM enrollee: Catron and Los Alamos. Together, these two counties make up less than one percent of New Mexico's Medicaid client population with BH disorders.

Table 3. Demographics of New Mexico Medicaid recipients with behavioral disorders by enrollment in CareLink New Mexico Program, January 2014 – March 2018.

Characteristic	**Not Enrolled**	**Enrolled**
Total	382,789	3,636
Age		
0-5 years	17,868 (4.7%)	94 (2.6%)
5-18 years	91,331 (23.9%)	777 (21.4%)
18-65 years	250,397 (65.4%)	2,695 (74.1%)
65+ years	23,192 (6.1%)	70 (1.9%)
Unknown	1 (0.0%)	0 (0.0%)
Gender		
Unknown	7 (0.0%)	0 (0.0%)
Female	215,527 (56.3%)	2,047 (56.3%)
Male	167,255 (43.7%)	1,589 (43.7%)
CLNM Enrollment		
0-6 months	NA	1,289 (35.5%)
6-12 months	NA	1,118 (30.7%)
1+ years	NA	1,229 (33.8%)
Study Eligibility		
Up to 1 year	120,028 (31.4%)	360 (9.9%)
2 years	89,419 (23.4%)	424 (11.7%)
3 years	55,145 (14.4%)	642 (17.7%)
4 years	37,298 (9.7%)	609 (16.7%)
5+ years	80,899 (21.1%)	1,601 (44.0%)
Behavioral Disorder(s)		
Mental Disorder	295,841 (77.3%)	2,015 (55.4%)
Substance Use Disorder	17,499 (4.6%)	19 (0.5%)
Both	69,449 (18.1%)	1,602 (44.1%)
Chronic Physical Conditions		
None	133,601 (34.9%)	745 (20.5%)
One	80,654 (21.1%)	640 (17.6%)
More Than One	168,534 (44.0%)	2,251 (61.9%)
Provider County Group		
Bernalillo	117,424 (30.7%)	653 (18.0%)
Curry	7,101 (1.9%)	1,189 (32.7%)
Hidalgo and Grant	6,645 (1.7%)	136 (3.7%)
Lea	10,418 (2.7%)	434 (11.9%)
Other	193,558 (50.6%)	264 (7.3%)
Quay, De Beca, and Roosevelt	5,358 (1.4%)	267 (7.3%)
Sandoval	20,224 (5.3%)	189 (5.2%)
San Juan	22,061 (5.8%)	504 (13.9%)

The average age of enrolled and non-enrolled individuals was similar (33 vs. 32 years, respectively, Figure 3), and the enrolled group consisted of a larger proportion of adults under the age of 65 years than the non-enrolled group.

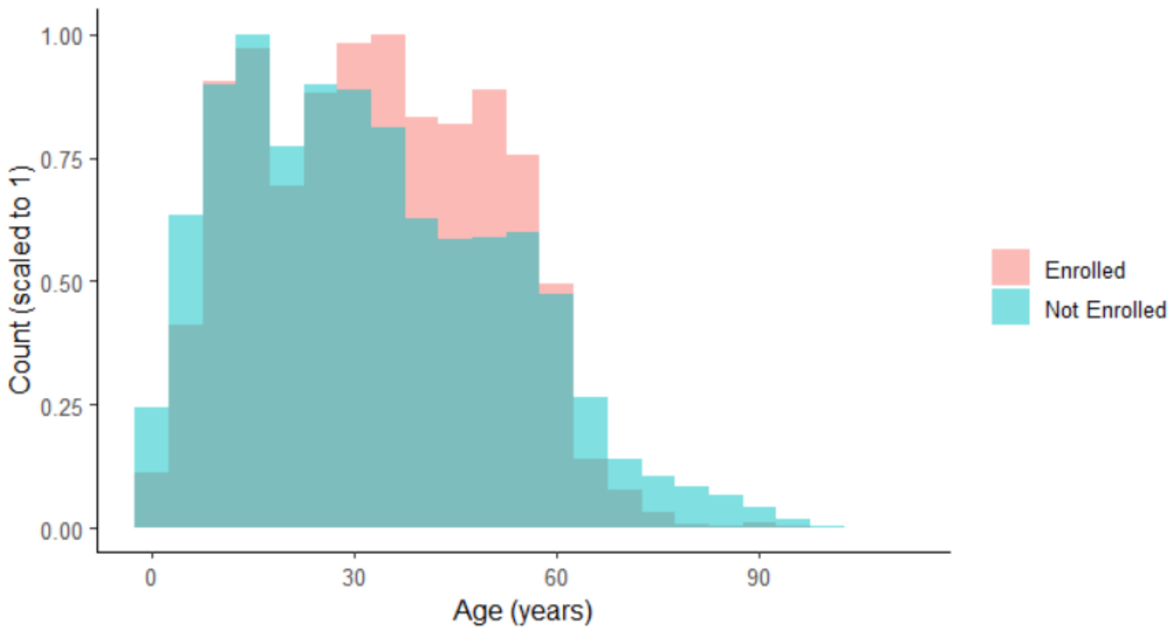


Figure 3. Distribution of age among New Mexico Medicaid recipients with behavioral disorders by enrollment in CareLink New Mexico program.

The number of quarters eligible for study inclusion is an indicator of Medicaid eligibility since eligibility data were not available for many individuals. Most enrolled individuals were eligible for at least a year of the study period (median = 14), but most non-enrolled individuals were not eligible for a full year (median = 8).

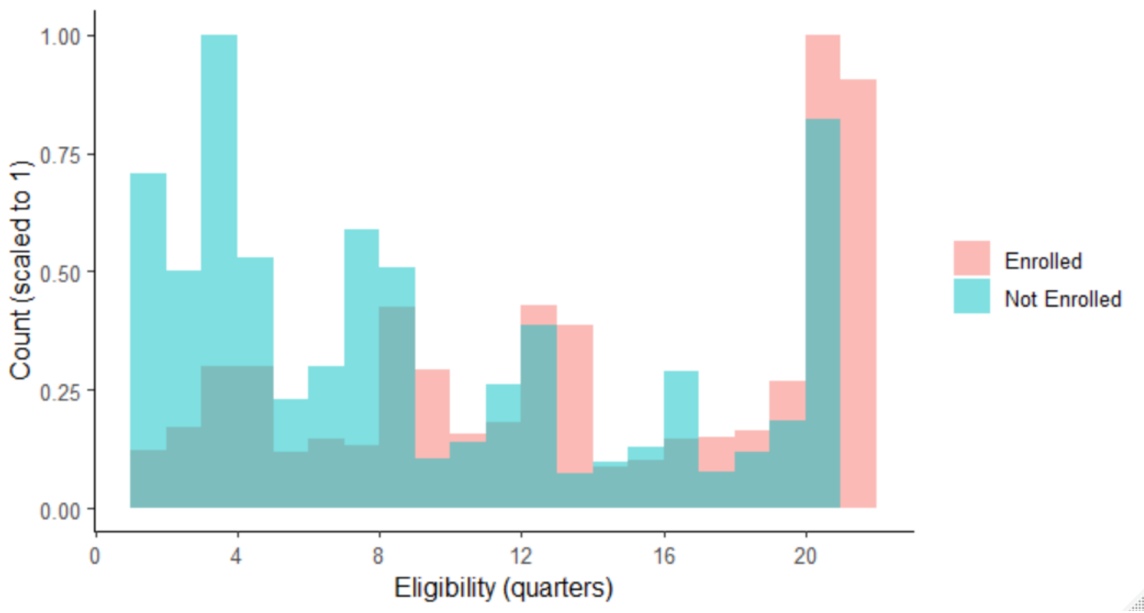


Figure 4. Distribution of eligibility in quarters of New Mexico Medicaid recipients with behavioral disorders by enrollment in CareLink New Mexico program.

Due to the screening criteria for CLNM, enrolled individuals had more physical health issues than non-enrolled individuals. More than sixty percent of enrolled individuals had multiple chronic physical conditions, while only 44% of non-enrolled individuals had more than one (Figure 5).

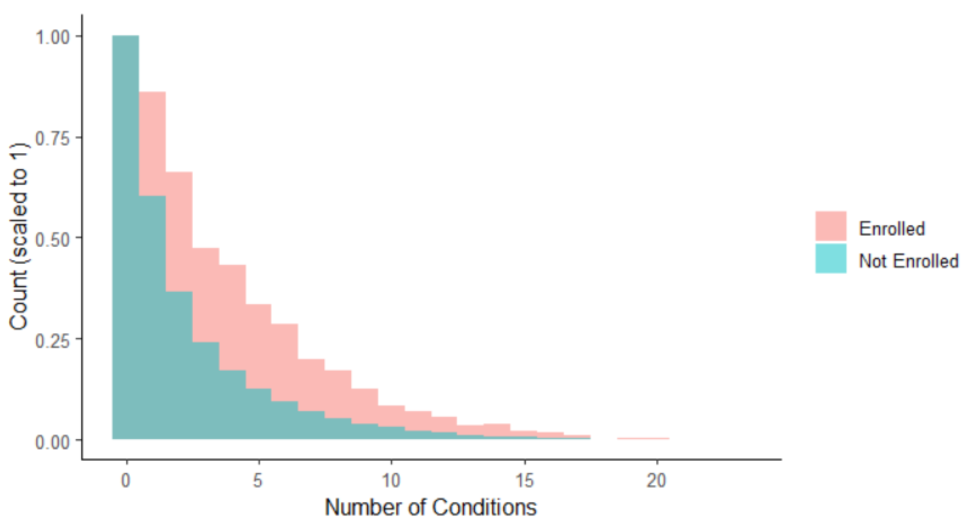


Figure 5. Distribution of the number of physical comorbidities of New Mexico Medicaid recipients with behavioral disorders by enrollment in CareLink New Mexico program.

As expected in an observational study, the enrolled individuals are quite different from the non-enrolled individuals. In particular, they had been eligible for Medicaid for a longer period of time, were more likely to have multiple physical and behavioral disorders, and generally lived in areas served by CLNM providers. Since there are thousands of enrolled individuals approximately one hundred times the number of non-enrolled individuals, we only need a subset of comparable non-enrolled individuals in order to carry out the analysis. Therefore, we will use propensity score matching to select this subset.

4.2 Propensity Score Matching

4.2.1 Descriptive Analysis of Potential Covariates

Only one individual was missing an age and seven were missing gender, for a total of 7 unique individuals missing any potential covariates, and none were enrolled in CLNM. Therefore, multiple imputation was foregone, and these seven individuals were excluded from the regression analysis. There were 629 individuals whose county of residence could not be determined. Because these individuals' county of residence could not be determined due to a lower number of claims and treatment in multiple counties, these individuals were placed in their own county category of "Unknown."

Older individuals had higher charges and fewer BH claims (Figure 6). Young adults were most likely to have emergency medical claims.

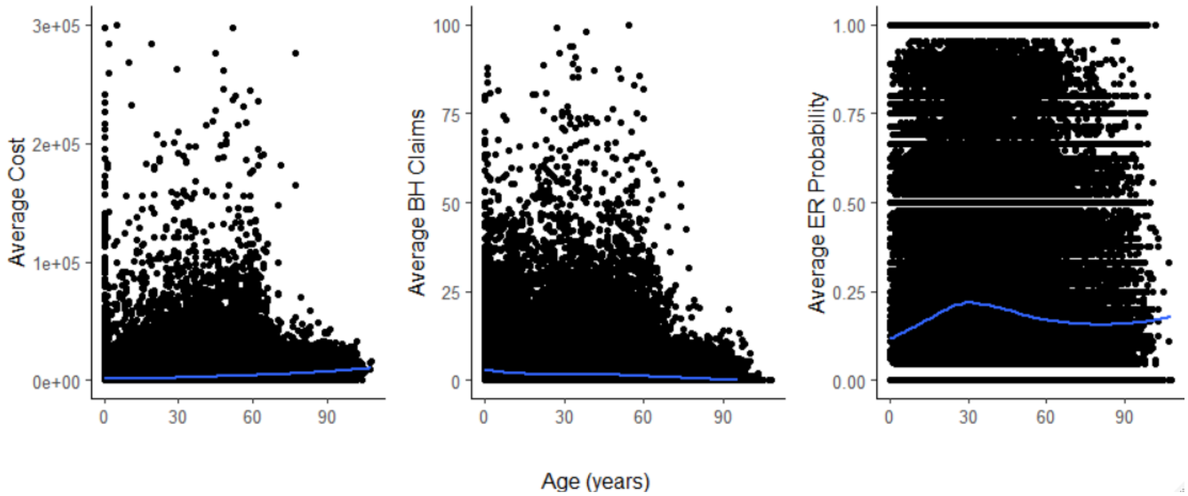


Figure 6. Age of New Mexico Medicaid recipients with behavioral health disorders by CareLink New Mexico program enrollment and outcomes. Smoothing lines created using loess methods.

Male and female individuals have similar average total charges ($2,949 and $2,462, respectively) and similar average number of emergency claims (0.3 each). However, males have a higher average quarterly number of BH claims than females (2.1 and 1.4, respectively). There were high rates of enrollment in Curry, Lea, Roosevelt, and San Juan counties due to the location of early CLNM providers. However, Lea has a very low number of BH and emergency care claims per quarter (0.9 and 1.0, respectively) compared to counties like Bernalillo (2.1 and 0.3, respectively) and Santa Fe (2.4 and 0.4, respectively), potentially due to the availability of care in metropolitan and rural areas. The average quarterly charges did not vary widely by county of residence.

The average charges and probability of utilizing emergency care decreased with the number of contributed months (Figure 7). However, the average number of quarterly BH claims increased.

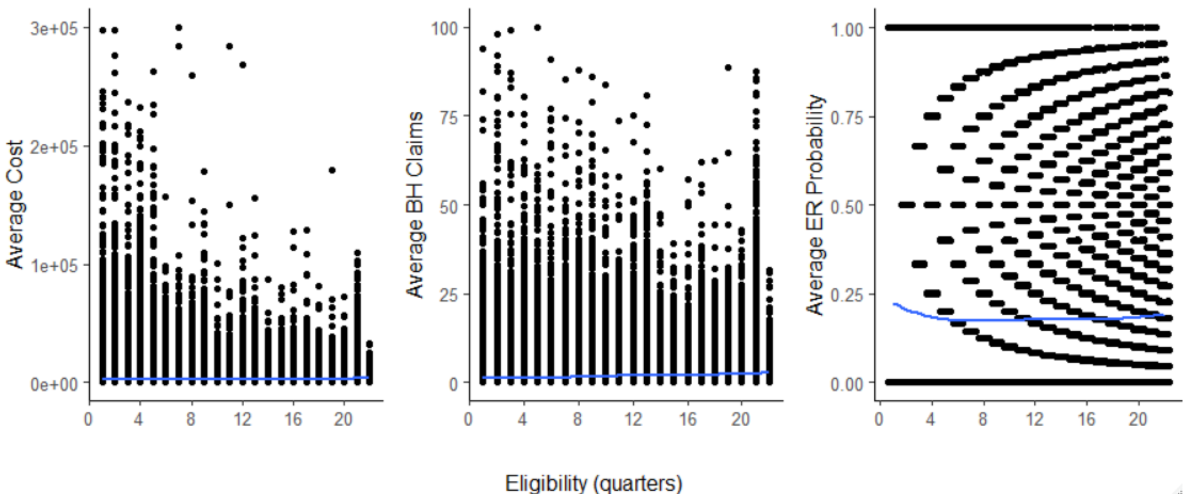


Figure 7. Number of quarters eligible for study inclusion among New Mexico Medicaid recipients with behavioral health disorders by CareLink New Mexico enrollment and outcomes. Smoothing lines created using loess methods.

Medicaid recipients with both SUDs and mental disorders had higher charges, a greater number of BH claims, and a higher probability of emergency care utilization (2.3%, $3,796, 2.8, and 0.5 respectively) than those with only mental disorders (0.7%, $2,333, 1.4, and 0.3 respectively) or those with only substance use disorders (0.1%, $3,943, 2.7, and 0.3 respectively, Figure 8).

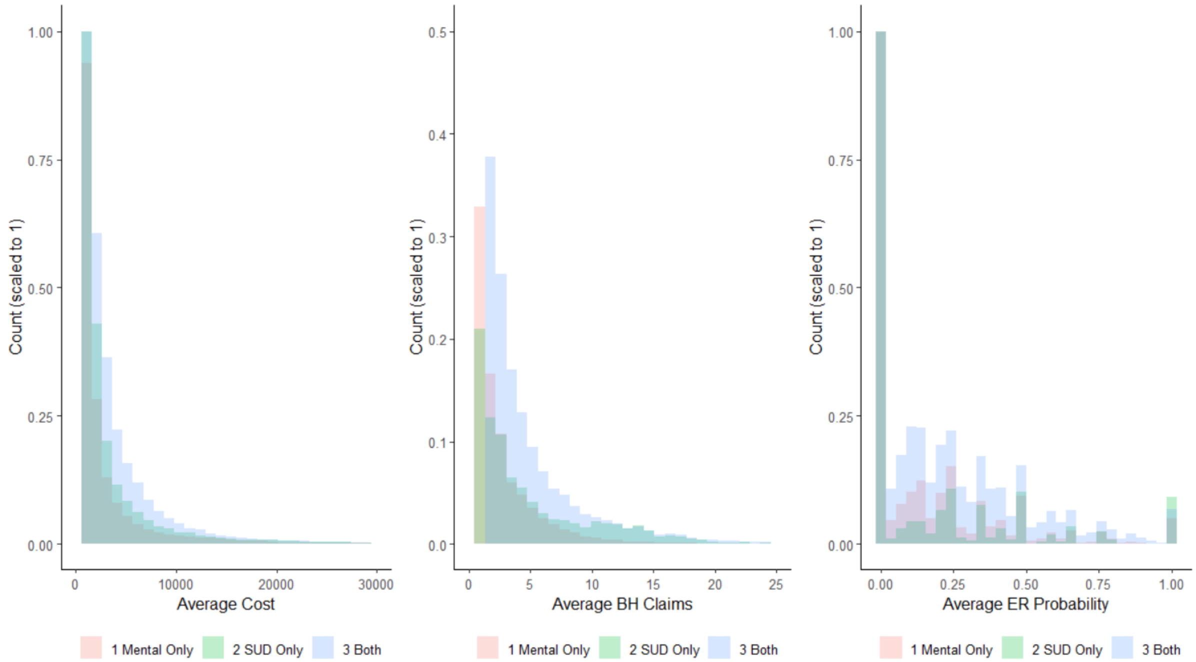


Figure 8. Behavioral health disorder(s) of New Mexico Medicaid recipients with behavioral health disorders by CareLink New Mexico program enrollment and outcomes.

The distributions of the number of chronic physical conditions were similar among enrollees and non-enrollees. However, the average quarterly charges increased with the number of conditions, particularly from 0 to 6 conditions (Figure 9). Conversely, the average quarterly number of BH claims decreased with the number of conditions. The average quarterly probability of emergency care utilization increased steadily with the number of conditions.

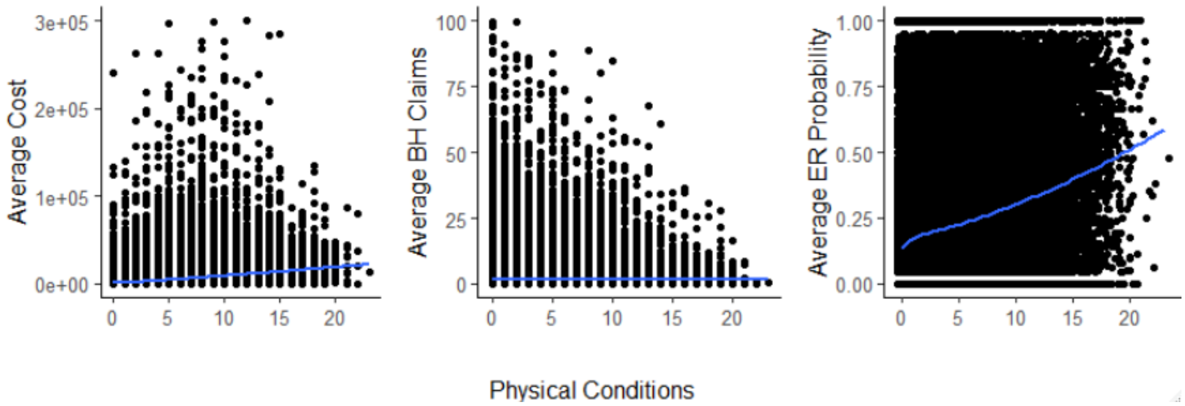


Figure 9. Number of physical conditions among New Mexico Medicaid recipients with behavioral health disorders by CareLink New Mexico program enrollment and outcomes. Smoothing lines created using loess methods.

4.2.2 Propensity Score Estimation

All interactions with county were removed from the propensity score model, except for that with age, as these terms led to non-convergence of the model and outcomes of 0 or 1 for some individuals. Interactions of gender with number of physical conditions and number of contributed quarters were also removed, which did not reduce common support. The probabilities from the final model were merged to the client-level dataset as propensity scores ranging from very close to zero to 0.6376 (Figure 10).

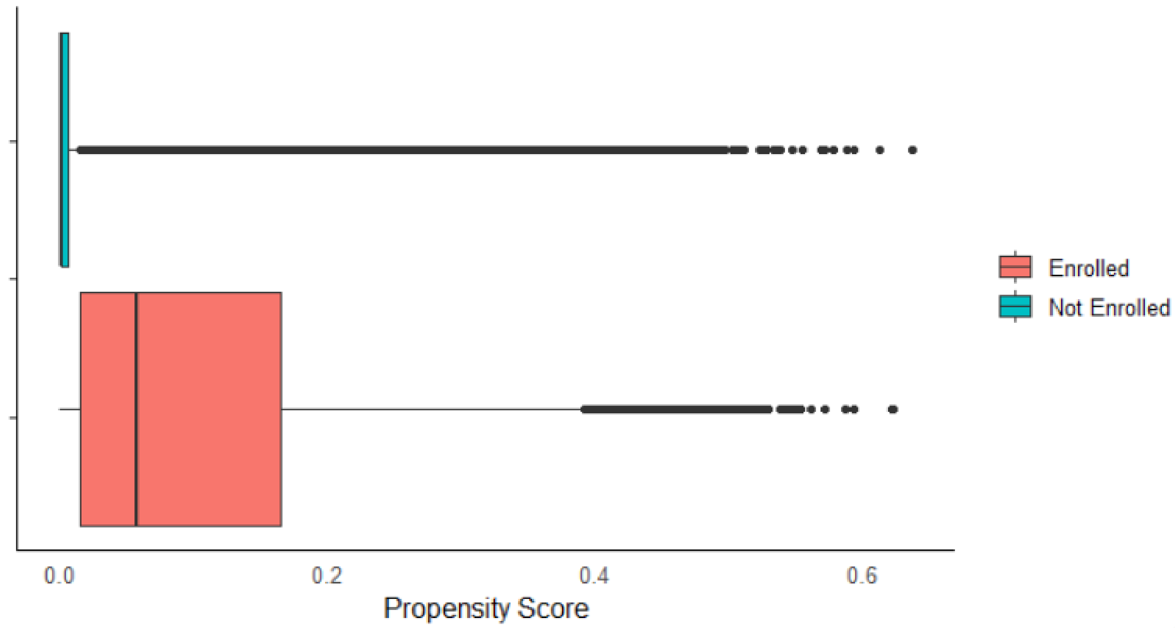


Figure 10. Distribution of propensity scores among New Mexico Medicaid recipients with behavioral health disorders by enrollment in the CareLink New Mexico program.

Given the common support of propensity scores between enrolled and non-enrolled individuals, the matching method used was the nearest matching propensity score for each enrollee. Therefore, each enrollee had one matched non-enrollee, and the final number of individuals included in the models was 7,272 (3,636 enrollees and 3,636 non-enrollees) with a total of 1,955,980 claims. The means of all covariates included in the propensity score estimation model were similar for enrollees and matched non-enrollees by propensity score (Figure 11). Comparing enrollees and matched non-enrollees, the means of age (33.1 each), number of contributed quarters (14.3 and 13.8, respectively), and number of physical conditions (3.4 each) were similar.

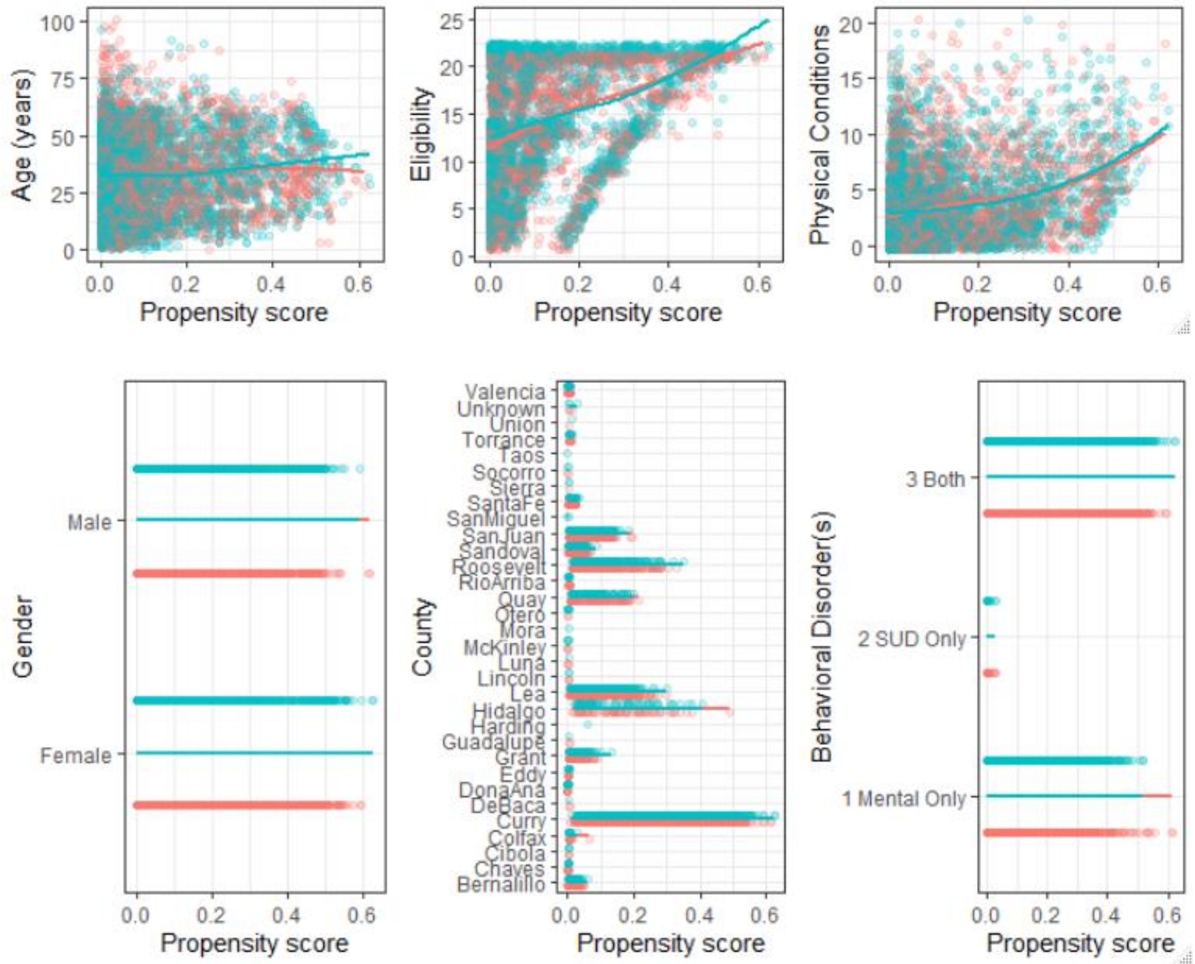


Figure 11. Mean of covariates of New Mexico Medicaid recipients with behavioral health disorders who were enrolled (green) and their matched non-enrolled recipients (red) in the CareLink New Mexico program, by propensity score used to match them in pairs. Loess smoothing was used to estimate the mean of numeric covariates.

In the final matched subset, the demographics of individuals are more similar between enrolled and non-enrolled individuals than in the full sample. In particular, the length of Medicaid eligibility and the frequency of multiple physical and behavioral disorders are more similar (Table 4). In addition, there are no longer any individuals from Catron or Los Alamos counties included, since these counties had no individuals enrolled in CLNM. Propensity scores among matched pairs differed by up to 0.02, and 75% were different by less than 0.00002.

Table 4. Demographics of a matched subset of New Mexico Medicaid recipients with behavioral disorders by enrollment in CareLink New Mexico program, 2014-2018.

Characteristic	**Not Enrolled**	**Enrolled**
Total	3,636	3,636
Age		
0-5 years	90 (2.5%)	94 (2.6%)
5-18 years	759 (20.9%)	777 (21.4%)
18-65 years	2,644 (72.7%)	2,695 (74.1%)
65+ years	143 (3.9%)	70 (1.9%)
Gender		
Unknown	0 (0.0%)	0 (0.0%)
Female	1,999 (55.0%)	2,047 (56.3%)
Male	1,637 (45.0%)	1,589 (43.7%)
HH Enrollment		
0-6 months	NA	1,289 (35.5%)
6-12 months	NA	1,118 (30.7%)
1+ years	NA	1,229 (33.8%)
Study Eligibility		
Up to 1 year	427 (11.7%)	360 (9.9%)
2 years	470 (12.9%)	424 (11.7%)
3 years	574 (15.8%)	642 (17.7%)
4 years	524 (14.4%)	609 (16.7%)
5+ years	1,641 (45.1%)	1,601 (44.0%)
Behavioral Disorder(s)		
Mental Disorder	2,000 (55.0%)	2,015 (55.4%)
Substance Use Disorder	20 (0.6%)	19 (0.5%)
Both	1,616 (44.4%)	1,602 (44.1%)
Chronic Physical Conditions		
None	729 (0.0%)	745 (0.0%)
One	656 (20.0%)	640 (20.5%)
More Than One	2,251 (18.0%)	2,251 (17.6%)
Provider County Group		
Bernalillo	614 (16.9%)	653 (18.0%)
Curry	1,176 (32.3%)	1,189 (32.7%)
Hidalgo and Grant	135 (3.7%)	136 (3.7%)
Lea	468 (12.9%)	434 (11.9%)
Other	257 (7.1%)	264 (7.3%)
Quay, De Beca, and Roosevelt	286 (7.9%)	267 (7.3%)
Sandoval	191 (5.3%)	189 (5.2%)
San Juan	509 (14.0%)	504 (13.9%)

4.3 Quarterly Measures

4.3.1 By Demographics

Using the subset of enrolled individuals and matched comparison individuals, we now explore the associations of healthcare charges and utilization with demographics during the baseline period. The outcome measures change considerably with age (Figure 12). Notably, there are changes among 18-year-olds and adults over 65 years. This coincides with dramatic changes in the number of Medicaid recipients at these ages based on eligibility criteria. At age 65, most New Mexicans are eligible to receive Medicare. This makes Medicaid the payer of last resort for the 65+ years age group. Aside from these changes, it appears that charges increase with age, as do the number of claims for inpatient care and the prescription of behavioral medications (until 65 years of age). The need for emergency health care decreases with age after young adulthood.

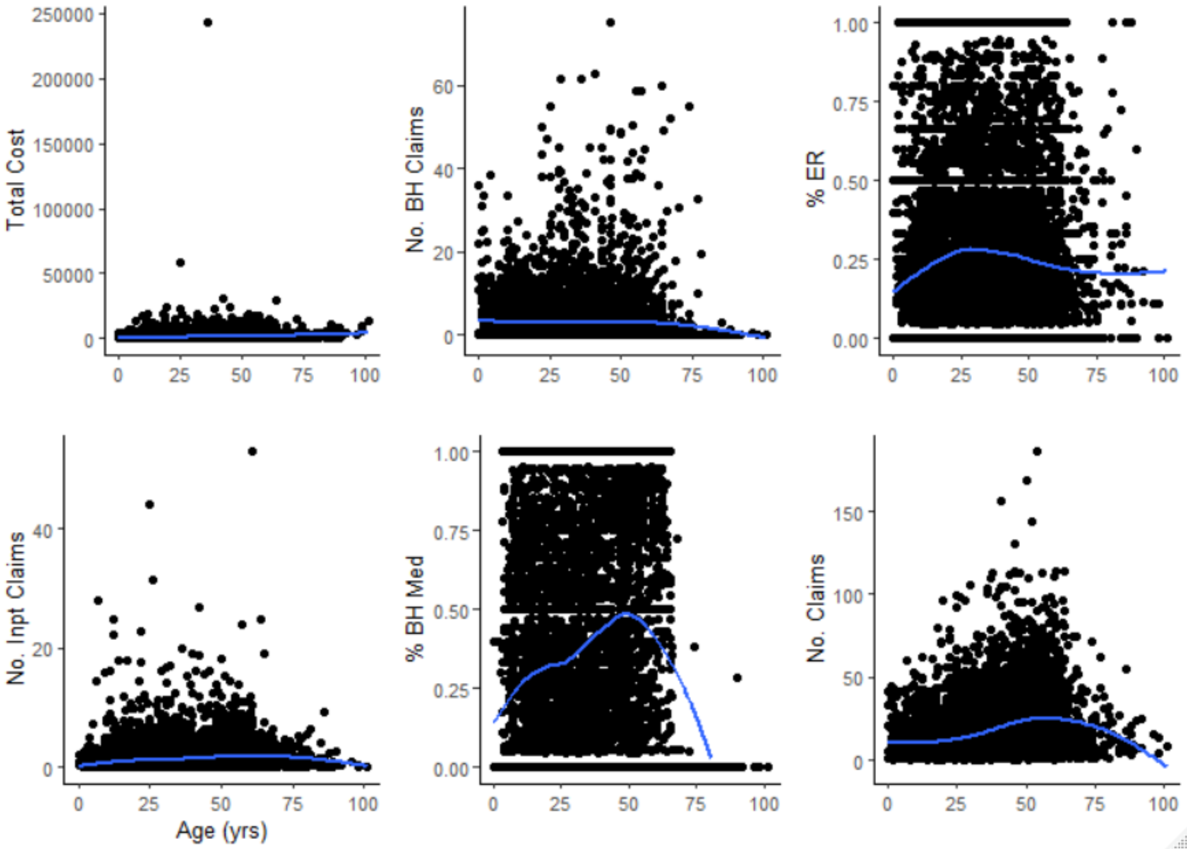


Figure 12. Scatterplots of average outcome measures by age among matched subset of Medicaid recipients with behavioral disorders prior to the CareLink New Mexico program.

The number of quarters each individual was eligible for inclusion in this study was not strongly correlated with charges, probability of emergency care claims, or the number of inpatient claims (Figure 13). The number of BH claims, probability of BH medication prescription, and number of claims increased with inclusion time.

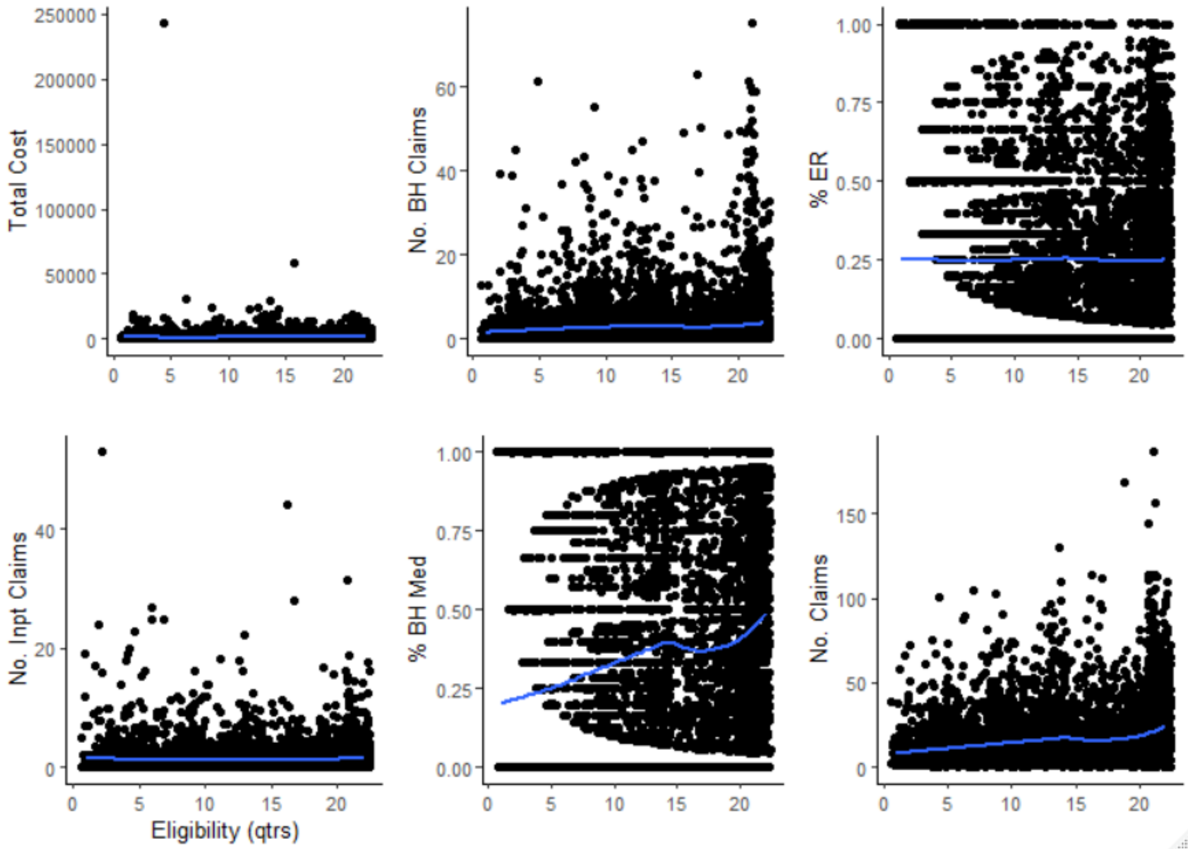


Figure 13. *Scatterplots of average outcome measures by the length of time eligible for study inclusion among a matched subset of Medicaid recipients with behavioral disorders prior to the CareLink New Mexico program.*

The number of physical comorbidities each individual had was positively correlated with charges, the probability of emergency care claims, the number of inpatient claims, and the number of total claims (Figure 14). The number of BH claims and the probability of being prescribed a BH medication increased until there were ten comorbidities present, and then decreased.

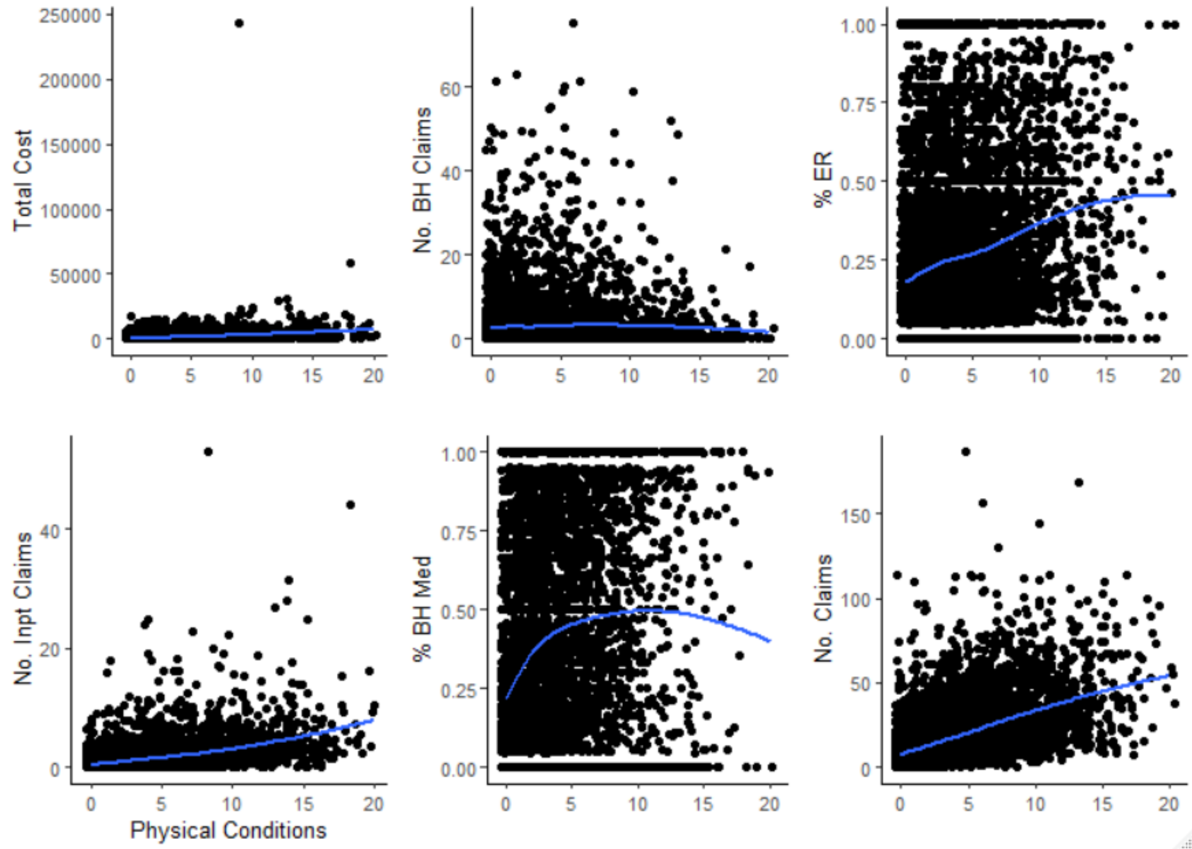


Figure 14. Scatterplots of average outcome measures by the number of chronic physical conditions among a matched subset of Medicaid recipients with behavioral disorders prior to the CareLink New Mexico program.

As unemployment rates increased, the total charges, number of BH claims, probability of behavioral medication prescription, and total number of claims increased as well (Figure 15). However, the probability of utilizing emergency health care was lower when unemployment rates were high.

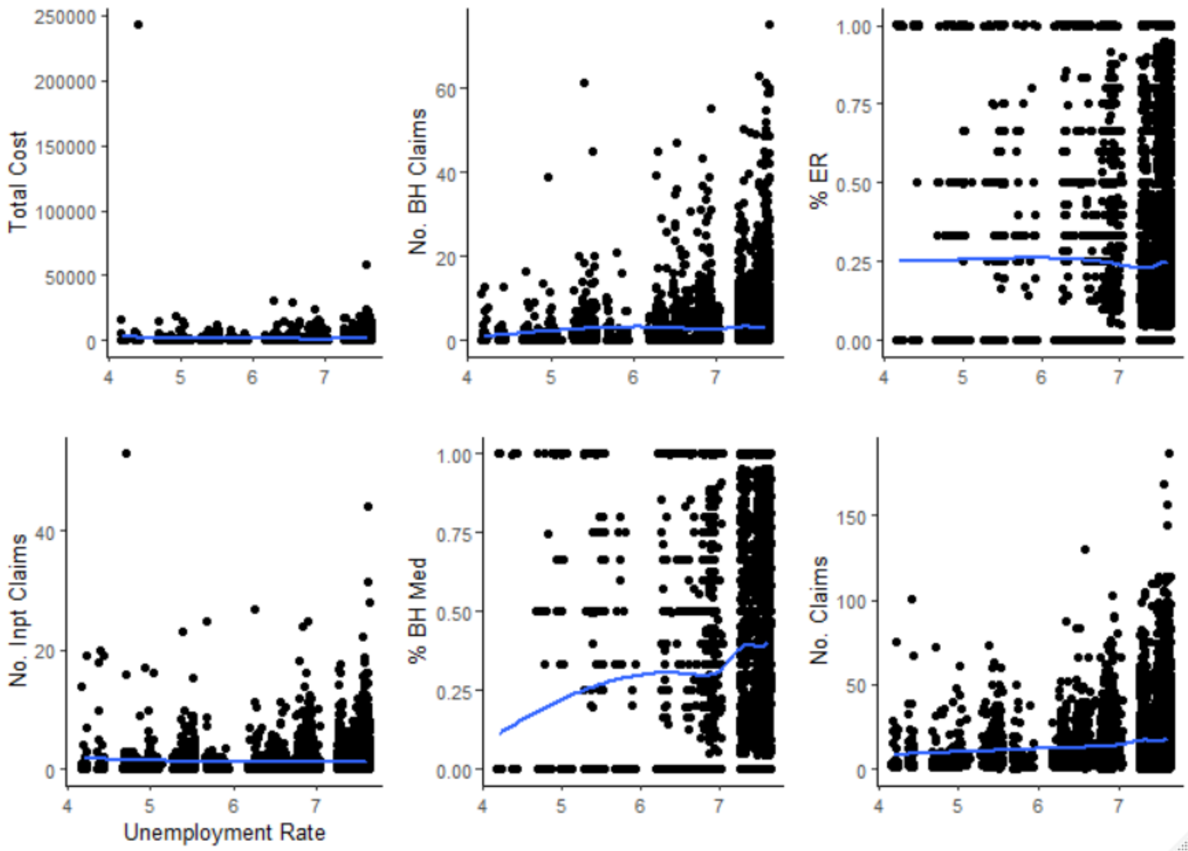


Figure 15. *Scatterplots of average outcome measures by statewide unemployment rate among a matched subset of Medicaid recipients with behavioral disorders prior to the CareLink New Mexico program.*

As calendar time passed, the total charges, number of BH claims, number of inpatient claims, probability of BH medication prescription, and total number of claims increased, but the probability of emergency care claims decreased (Figure 16).

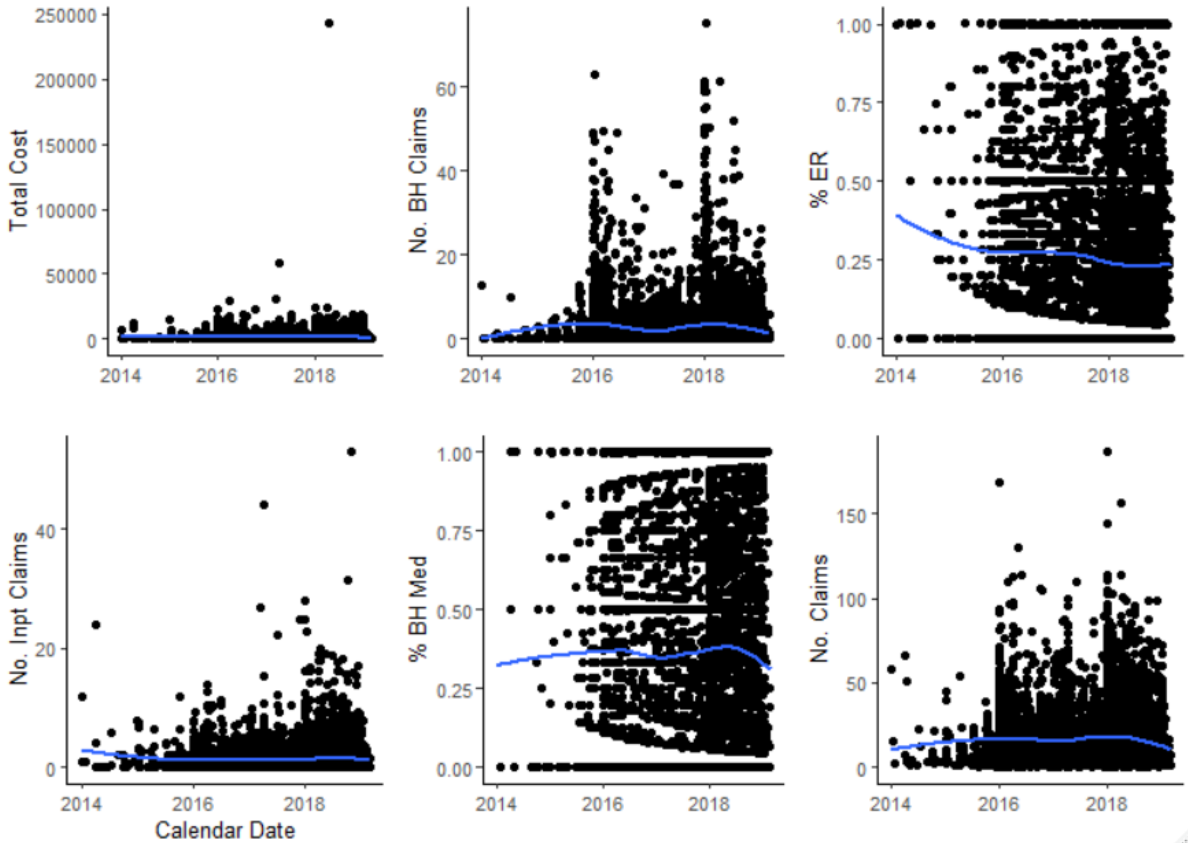


Figure 16. Scatterplots of average outcome measures by calendar date of claims among a matched subset of Medicaid recipients with behavioral disorders prior to the CareLink New Mexico program.

Individuals enrolled in CLNM had a higher average total charges than non-enrolled individuals, a higher number of BH claims, and a higher percentage of behavioral medications. Total charges were similar by gender, although male individuals had more BH claims per quarter than female individuals (Table 5). Those living in Grant and Hidalgo Counties had the highest charges, and those living in Curry County had the lowest number of BH claims. Lea county had an extremely high average probability of emergency care utilization. This county is situated in the far southeast corner of the state and is located over the Permian Basin where a large amount of crude oil is produced, so the population is different from the rest of the state, as oil workers come from other parts of the country when production is high. Finally, individuals with both mental and substance use disorders had the highest average of all three outcomes.

Table 5. Quarterly outcome measures of New Mexico Medicaid recipients with behavioral disorders by demographics and clinical characteristics prior to the CareLink New Mexico program.

Characteristic	**Total Charges**	**Behavioral Health Claims**	**Emergency Claims**
	Mean (Median) [Interquartile Range]		
CLNM Enrollment			
No	$1,270.40 ($496.30) [$200.60-$1,303.30]	1.8 (0.4) [0.0-2.0]	0.2 (0.2) [0.0-0.4]
Yes	$1,466.42 ($0,719.56) [$0,297.59-$1,733.09]	4.0 (1.6) [0.4-4.5]	0.3 (0.2) [0.0-0.4]
Gender			
Female	$1,318.40 ($640.60) [$263.00-$1,529.10]	2.6 (0.8) [0.0-2.8]	0.3 (0.2) [0.0-0.4]
Male	$1,433.20 ($540.60) [$214.50-$1,528.20]	3.3 (1.1) [0.1-3.7]	0.2 (0.1) [0.0-0.4]
Behavioral Disorder(s)			
Mental Disorder	$1,366.00 ($524.10) [$209.70-$1,357.80]	2.8 (0.7) [0.0-3.0]	0.4 (0.1) [0.0-0.5]
Substance Use Disorder	$718.60 ($0,341.90) [$178.00-$630.00]	2.0 (0.8) [0.0-1.4]	0.5 (0.0) [0.0-0.6]
Both	$1,482.44 ($0,736.78) [$304.31-$1,768.08]	3.3 (1.0) [0.1-3.6]	0.7 (0.3) [0.0-0.8]
Provider County Group			
Bernalillo	$1,429.60 ($713.50) [$278.80-$1,767.60]	3.8 (1.6) [0.5-4.7]	0.2 (0.1) [0.0-0.3]
Curry	$1,374.70 ($529.70) [$208.40-$1,478.70]	2.1 (0.6) [0.0-2.6]	0.2 (0.2) [0.0-0.4]
Hidalgo and Grant	$2,717.30 ($855.90) [$370.90-$1,853.50]	2.8 (1.0) [0.1-4.1]	0.1 (0.0) [0.0-0.1]
Lea	$988.73 ($434.58) [$191.11-$1,093.22]	2.9 (0.6) [0.0-2.4]	0.4 (0.4) [0.1-0.7]
Other	$1,355.07 ($584.58) [$263.72-$1,596.26]	3.0 (1.0) [0.3-3.2]	0.2 (0.2) [0.0-0.4]
Quay, De Baca, and Roosevelt	$1,380.08 ($667.17) [$299.70-$1,681.03]	3.3 (1.1) [0.1-3.9]	0.2 (0.1) [0.0-0.3]
Sandoval	$1,151.28 ($595.79) [$260.58-$1,435.26]	2.4 (1.2) [0.3-3.1]	0.2 (0.1) [0.0-0.3]
San Juan	$1,342.90 ($677.60) [$288.70-$1,559.80]	3.5 (0.7) [0.0-3.1]	0.3 (0.2) [0.0-0.4]

4.3.2 Total Charges

Medicaid charges ranged from $0 (which was common due to the inclusion of quarters when no claims were submitted) to $242,842, with an average of $1,501 and a median of $465. Figure 17 displays the quantiles of total charges per quarter by CLNM enrollment. There is more variability in charges among CLNM enrollees and a floor effect at zero.

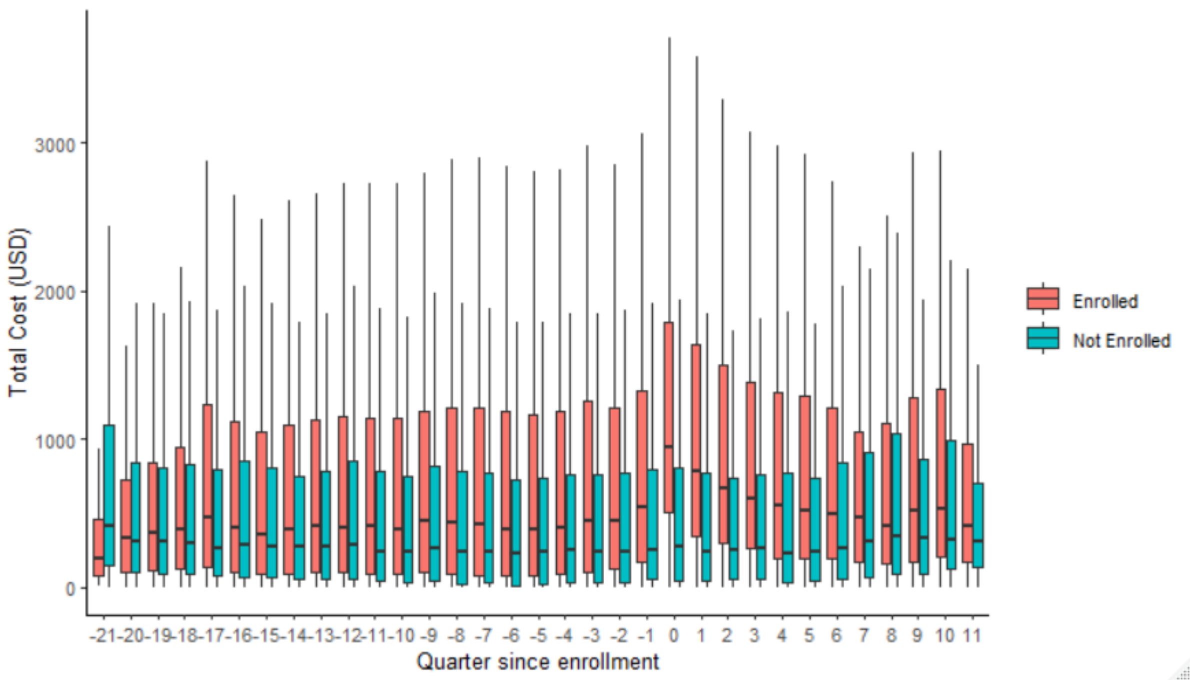


Figure 17. Quantiles of quarterly medical charges among a matched subset of New Mexico Medicaid recipients with behavioral disorders by enrollment in the CareLink New Mexico program. Quarter "0" is the first quarter of enrollment.

Figure 18 displays spaghetti plots of the trends in quarterly total charges for enrolled and non-enrolled individuals. There appears to be a general decrease in charges in the follow-up period among the enrolled individuals that does not occur among the non-enrolled individuals.

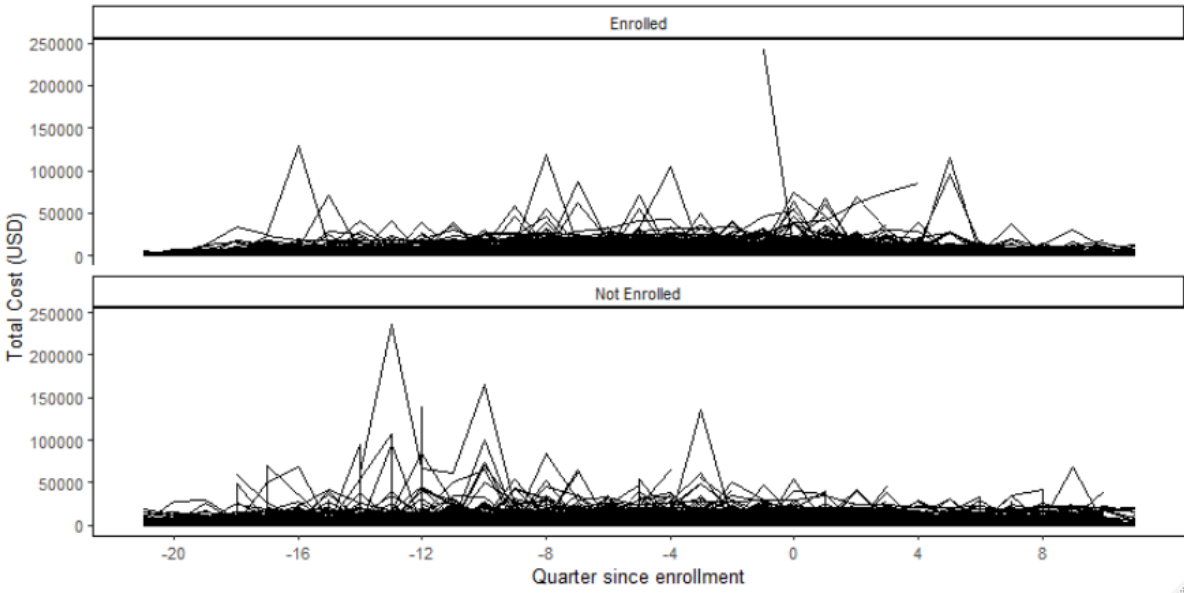


Figure 18. Quarterly Medicaid charges among a matched subset of New Mexico Medicaid recipients with behavioral disorders by enrollment in the CareLink New Mexico program. Quarter "0" is the first quarter of enrollment.

Total charges were averaged over each quarter for descriptive purposes. The average charges of non-enrolled individuals steadily increased from $1,488 nine quarters prior to CLNM enrollment to $1,570 eleven quarters after enrollment began (Figure 19). During the same time period, enrolled individuals' chargest decreased from $1,544 to $1,244, with a distinct increase at enrollment to $2,140 during the first quarter of enrollment. Notably, the total charges among enrolled individuals was increasing at a faster rate than that of non-enrolled individuals prior to enrollment.

Figure 19 demonstrates the importance of accounting for underlying trends in the analysis. The average of these quarterly charges among enrolled individuals was $1,373 before enrollment and $1,617 after enrollment due to the large investment in services during the first year, and particularly the first quarter. The average charges among non-enrolled individuals were $1,430 before April 1, 2016, and $1,383 afterward. Using a pre-post analysis, one would infer that enrollment in the HH program increases healthcare expenditures, when Figure 19 describes a situation where an initial investment in needed services is followed by a decrease in quarterly charges to below the pre-enrollment charges, and below costs of comparison individuals.

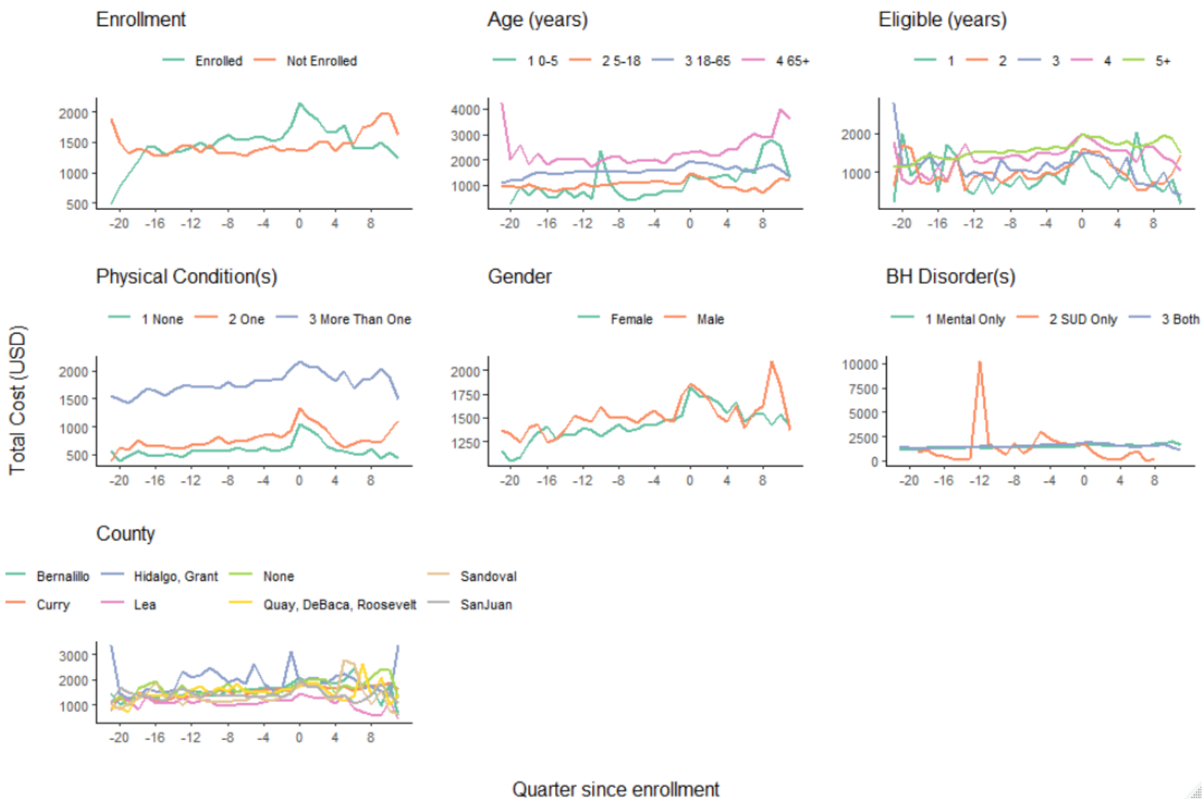


Figure 19. Average quarterly Medicaid charges among a matched subset of New Mexico Medicaid recipients with behavioral disorders by individual characteristics. Quarter "0" is the first quarter of enrollment in CareLink New Mexico.

Charges increased dramatically for Medicaid recipients 65 years of age and older since the beginning of the study period. The trend in charges also differed by the number of chronic physical conditions with which individuals had been diagnosed. In particular, individuals with more than one chronic physical condition had higher charges than others and a steady increase in charges throughout the study period, from $1,793 nine quarters prior to the beginning of CLNM enrollment to $1,920 eleven quarters afterward. Since charges are largely driven by the number of claims, this will be a covariate considered for inclusion in the modeling phase of the analysis. It is clear that those with the highest quartile of claims also have increased charges over time.

Charges can also be driven by the economic environment as it changes over time. As a proxy for this effect, we plotted charges by unemployment rates. The monthly unemployment rate ranged from 4.2% in 2018 to 7.6% in 2014 during the study period. There was no obvious association between quartile of unemployment rate and charges.

4.3.3 Behavioral Health Care

The number of quarterly BH claims ranged from 0 to 280, with an average of 3.5 and a median of 0. Figure 20 displays the quantiles of the number of outpatient visits for BH claims per quarter by CLNM enrollment. There is more variability in the number of BH claims among enrolled individuals and a floor effect at zero. The median number of claims is zero for non-enrolled individuals during all quarters, but the median number of visits increases from zero to more than zero after enrollment among enrolled individuals.

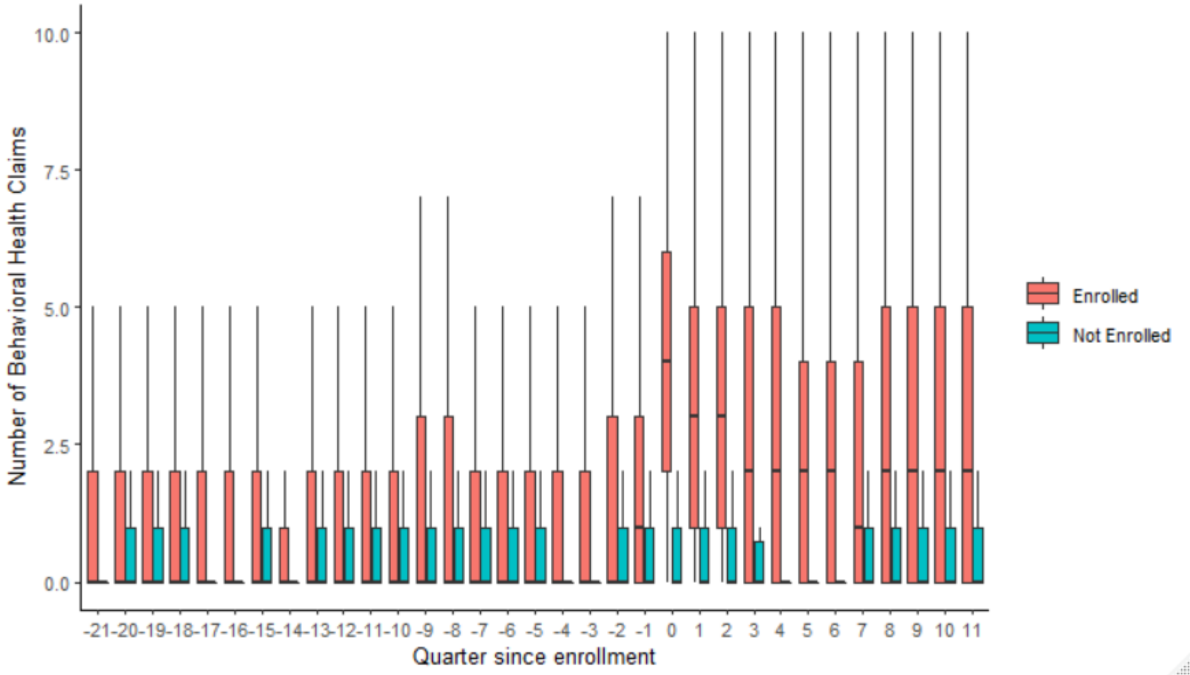


Figure 20. Quantiles of quarterly numbers of behavioral health care claims among a matched subset of New Mexico Medicaid recipients with behavioral disorders by individual characteristics. Quarter "0" is the first quarter of enrollment.

Figure 21 displays spaghetti plots of the trends in the quarterly number of BH claims for enrolled and non-enrolled individuals. There appears to be an immediate increase in claims in the follow-up period among the enrolled individuals that is not sustained and no change among non-enrolled individuals.

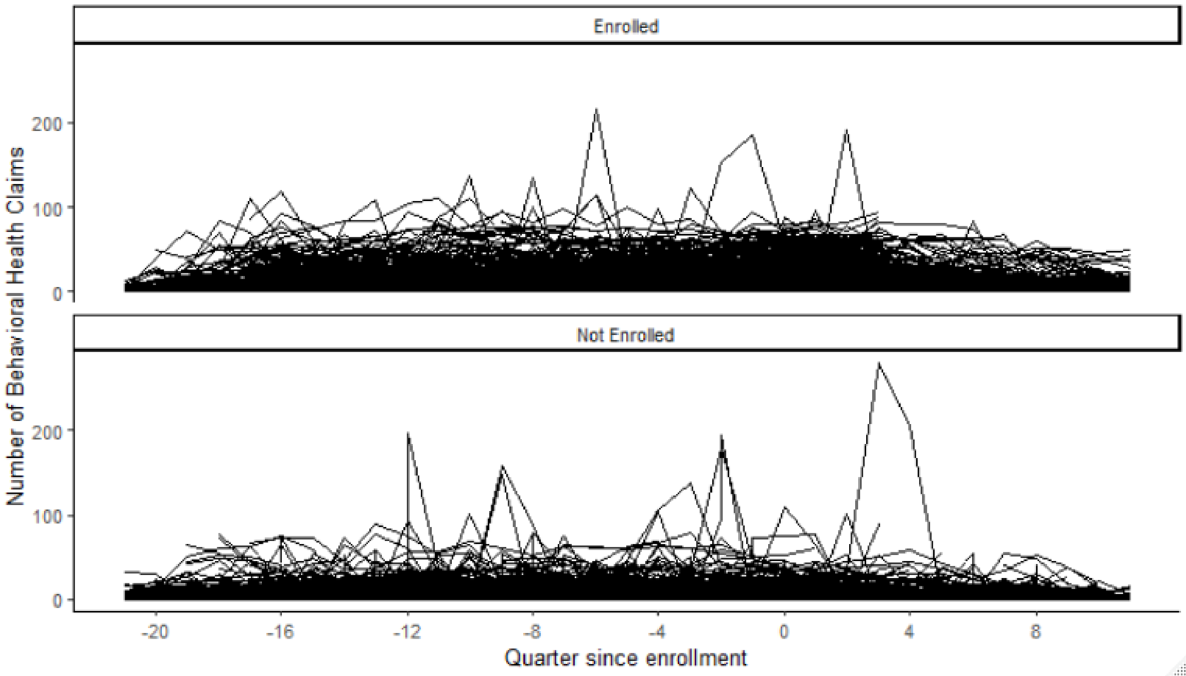


Figure 21. *Quarterly Medicaid outpatient visits for behavioral health care among a matched subset of New Mexico Medicaid recipients with behavioral health disorders by enrollment in the CareLink New Mexico program. Quarter "0" is the first quarter of enrollment.*

Individuals' numbers of outpatient visits for behavioral healthcare were averaged over each quarter for descriptive purposes. While this measure gradually increased by a small amount for individuals not enrolled in CLNM (from 1.9 nine quarters prior to enrollment to 2.2 eleven months afterward), enrolled individuals experienced a striking temporary increase within the first quarter of enrollment compared to two quarters prior (from 4.4 to 9.8) and a sustained level change to at least 4.7 for the rest of the study period.

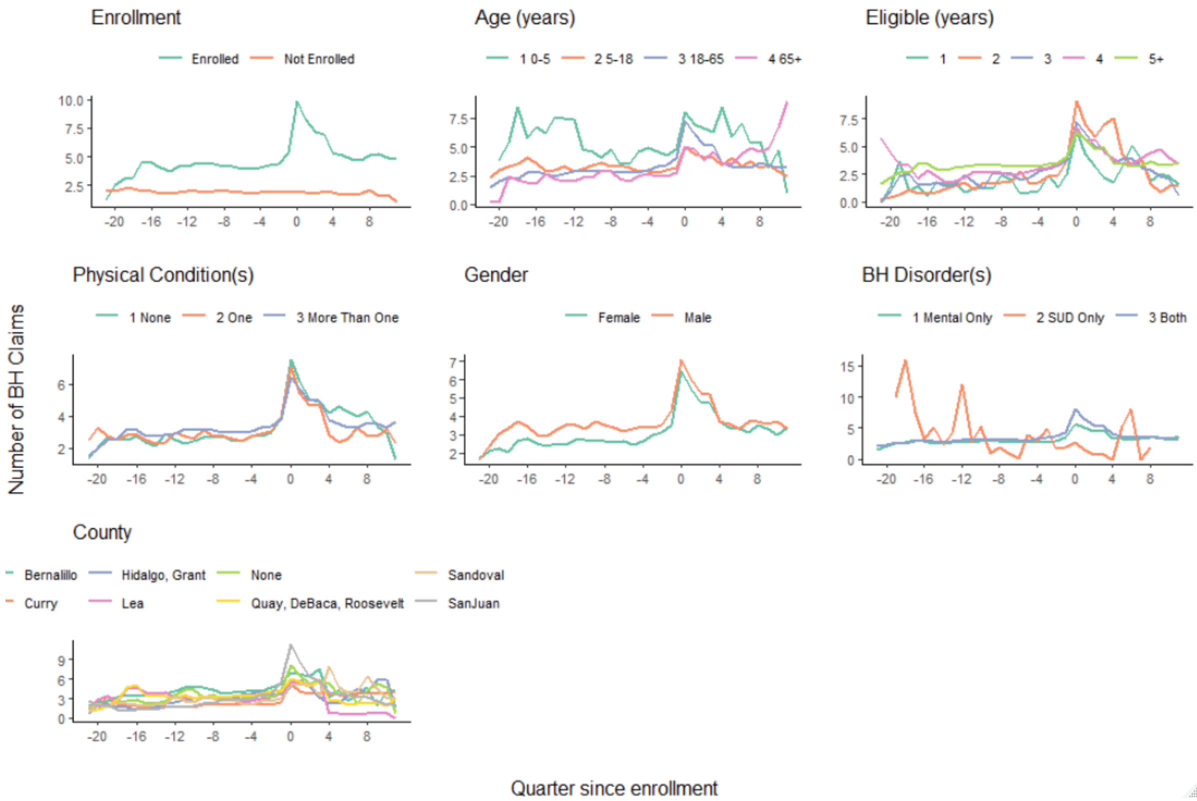


Figure 22. Quarterly average outpatient behavioral health claims among a matched subset of New Mexico Medicaid recipients with behavioral disorders by individual characteristics. Quarter "0" is the first quarter of enrollment in CareLink New Mexico.

4.3.5 Emergency Health Care

The proportion of Medicaid recipients with emergency care claims each quarter ranged from 17.3% in the first quarter of the pre-enrollment period to 27.6% in the last quarter of the pre-enrollment period for CLNM enrollees, and from 22.7% in the first quarter of the post-enrollment period to 31.7% in the last quarter of the post-enrollment period (Figure 23). There appears to be an immediate decrease in claims in the follow-up period among the enrolled individuals and no change among the non-enrolled individuals. Individuals' average number of emergency health claims was averaged over each quarter. Enrolled individuals experienced a decrease in emergency health care immediately following enrollment (0.6 to 0.4), which is not sustained for the rest of the study period.

Figure 23. Quarterly average emergency health claims among a matched subset of New Mexico Medicaid individuals with behavioral disorders by enrollment in the CareLink New Mexico program. Quarter "0" is the beginning of enrollment.

4.4 Regression Analysis

4.4.1 Total Charges

4.4.1.1 Unadjusted Model

An unadjusted linear ITS model of the logarithmic total charges (plus $0.01) was fit due to the right-skewed distribution of raw charges (Figure 24). The figure demonstrates the zero-inflated nature of the total charges, which will be discussed during the model diagnostic descriptions. The base model included fixed effects for the parameters specified in 3.3 Models, which were specified in such a way that the change in slopes from the pre- to post-enrollment periods could be quantified and tested among enrolled and non-enrolled individuals. A random intercept for matched pair was included, as specified by the study design. Adding covariates and interaction terms between behavioral disorder and the ITS terms improved the AIC from 553,760 to 517,467. Incorporating splines for continuous covariates further improved it to 452,713. Finally, removing non-significant terms in a stepwise backwards fashion resulted in an AIC of 452,704.

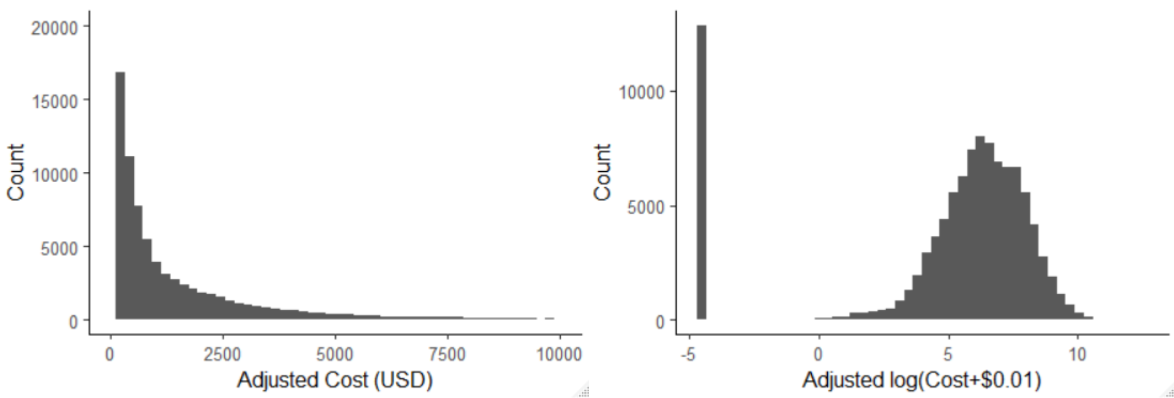


Figure 24. Histograms of (a) raw charges and (b) logarithmic charges (plus $0.01) among the matched subset of Medicaid recipients.

4.4.1.2 Random Effects

A third level was added to the adjusted model in the form of a random intercept for individual within matched cluster, resulting in a decrease in AIC to 447,854. However, a likelihood ratio test did not confirm that the three-level model fit better than the two-level model ($p=1$). Furthermore, the variance of the cluster was zero and the variance of the individual was 0.8462. Therefore, the random intercept for cluster was removed, reducing the AIC to 447,852. The likelihood ratio test confirmed that the random effect of individual was a better fit than the fixed effect model ($p<0.05$). At this point, a few interaction terms were no longer significant in the model, and they were removed one at a time as previously described, resulting in the final model (Appendix B. Model Results).

4.4.1.3 Diagnostics

The final model demonstrates a lack of fit for measurements of zero charges (Figure 25). A separate cluster of residuals was revealed when plotted against fitted values, and these were identified as the residuals of the measurements of zero charges. Additionally, there were two extreme outliers with residual values of -22 and -49. These two residuals come from two observations from the same individual during the pre-enrollment period. The individual was enrolled in the CLNM program, lived in Bernalillo County, had a mental disorder and no substance use disorder, and was 52 years of age. This individual had the two highest numbers of total quarterly claims during the pre-enrollment period (272 and 263), which resulted in the two largest absolute values of residuals.

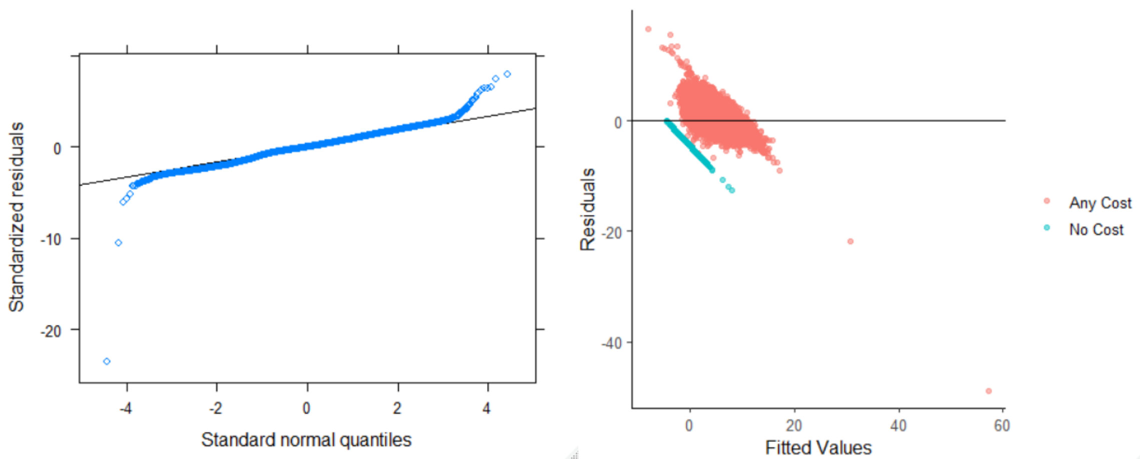


Figure 25. (a) Q-Q Plot of residuals and (b) residuals plotted against fitted values from adjusted mixed effects interrupted time series model of the logarithm of the total charges (plus $0.01) among the matched subset of Medicaid recipients.

The final model demonstrates a linear relationship between residuals and time, age, eligibility, physical comorbidities, unemployment rate, and calendar date (Figure 26). The variables containing information about the number of claims have outliers that impact the linear relationship with the outcome.

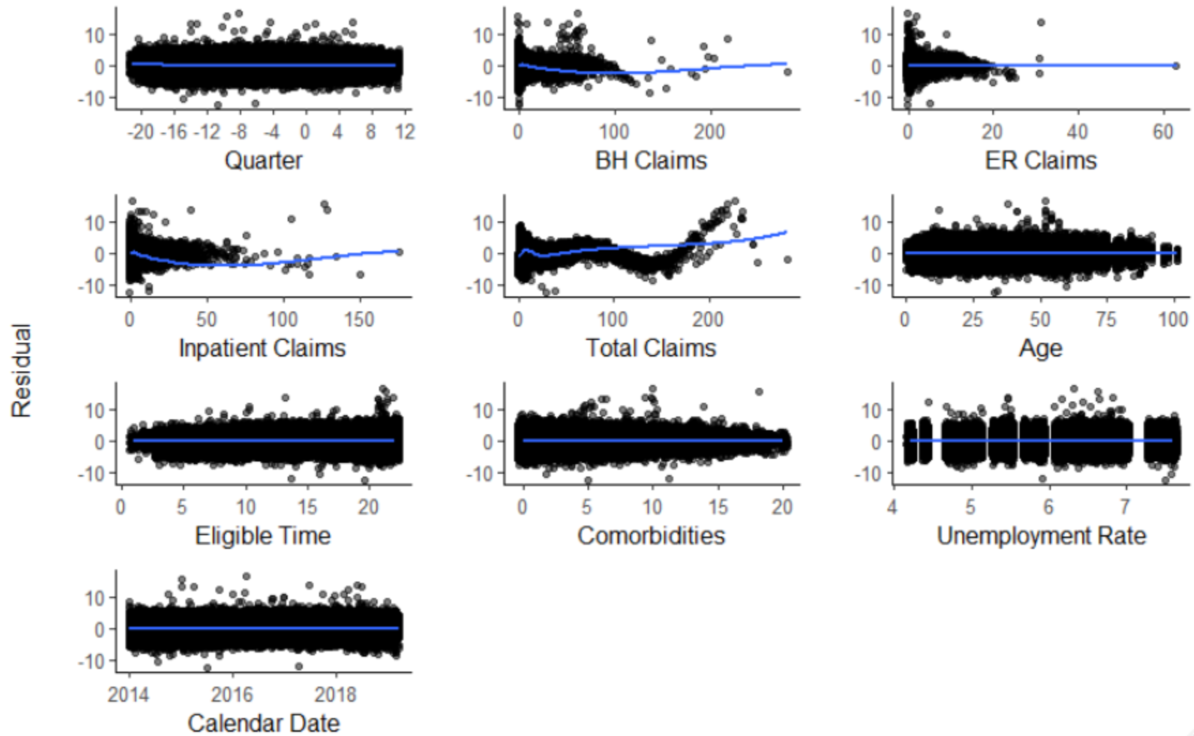


Figure 26. Pearson residuals by covariates from adjusted mixed effects interrupted time series model of the logarithm of perturbed total charges among the matched subset of Medicaid recipients. Two extreme outliers are excluded from these plots.

4.4.1.4 Final Model

The coefficients of the main effects of interest in the final ITS model are in Table 6. Final effects correspond with the predicted values in the following figures after adding slope terms together. The baseline slope of the total charges among enrolled individuals was 0.02. In the post-enrollment period, this slope decreased by 0.01, resulting in a slope of 0.01, and this decrease in slope was not statistically significant (p=0.0989). For non-enrolled individuals, the baseline slope of the total charges was 0.01, and increased significantly to 0.03 in the post-enrollment period (p-value<0.0100). There were no statistically significant differences by behavioral disorders in the post-enrollment period. The complete table of model results, including covariates, is in Appendix B.

Table 6. Covariate coefficients of final interrupted time series model of the natural logarithm of charges (plus $0.01) among the matched subset of Medicaid recipients.

Segment	Term	Coefficient	p-value	Final Effect
Overall				
Enrollees at Baseline	Intercept	3.28	<0.0001	3.28
	Slope	0.02	0.0002	0.02
Non-Enrollees at Baseline	Intercept	3.00	<0.0001	3.00
	Slope	0.01	0.0344	0.01
Enrollees at Follow-up	Intercept	3.44	<0.0001	3.44
	Slope	-0.01	0.0989	0.01
Non-Enrollees at Follow-up	Intercept	3.00	<0.0001	3.00
	Slope	0.02	0.0100	0.03
Effect of Disorder				
Enrollees at Baseline	Intercept: Mental Disorder Only	Ref	Ref	Ref
	Intercept: Substance Use Disorder Only	0.21	0.4208	3.49
	Intercept: Both Disorders	0.02	0.5837	3.30

Figure 27 demonstrates the predicted charges for individuals with no inpatient, emergency, or BH claims, no BH medication, residence in Bernalillo county, mental health disorders only, and average values of all other covariates. The results indicate that both enrolled and non-enrolled individuals experienced increases in total charges for the entire study period, although the rate of change decreased for enrolled individuals and increased for non-enrolled individuals, as previously described. Counterfactual trends are included in the figure, demonstrating that enrolled individuals will soon enter a period of cost savings compared to their baseline trend (starting three quarters after the end of the study period, or three and a half years after enrollment), while non-enrolled individuals’ charges continue to increase. Furthermore, enrolled individuals’ charges will dip below those of non-enrolled individuals seven quarters after the end of the study period, or 4.75 years after enrollment.

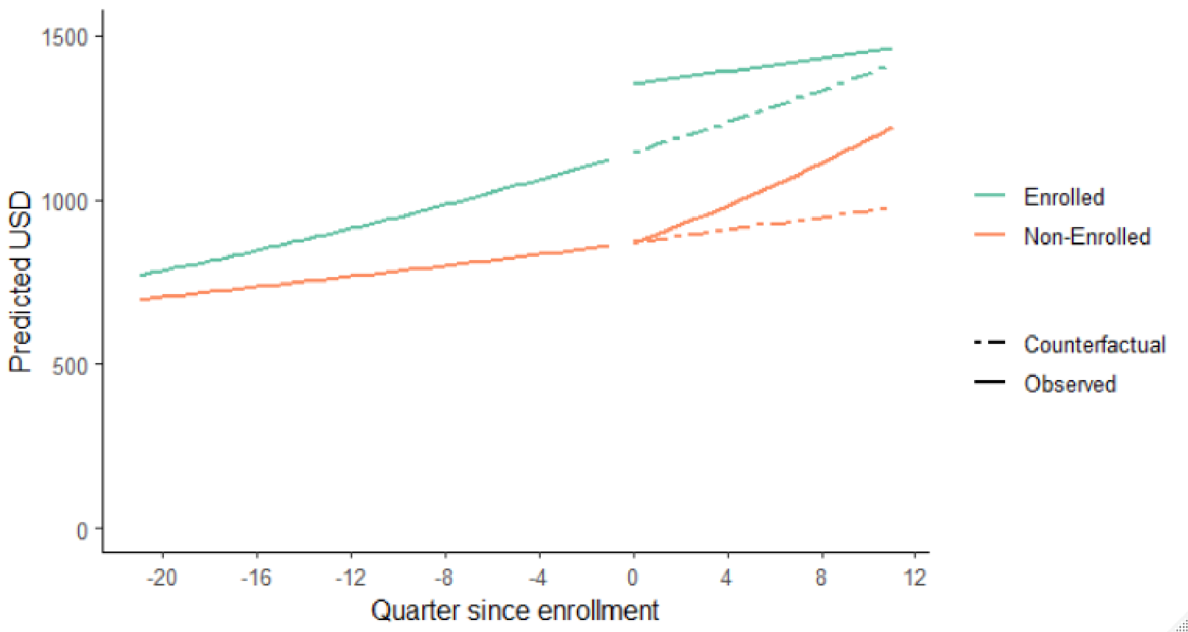


Figure 27. Predicted values by enrollment and time from final interrupted time series model of the natural logarithm of charges (plus $0.01) among the matched subset of Medicaid recipients.

The segments of the regression model differ by type of behavioral disorder. Those with a SUD had higher average quarterly charges in all segments, but a much higher average in the pre-enrollment period for enrolled individuals.

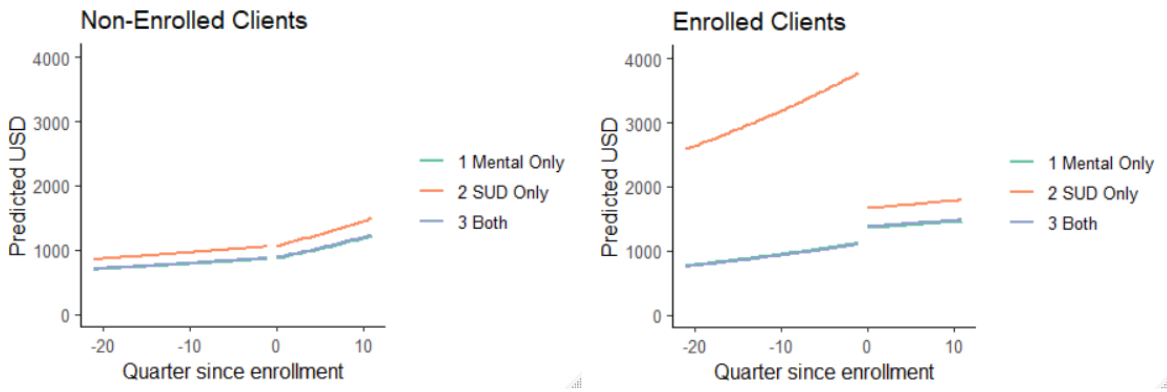


Figure 28. Predicted values by enrollment, time, and behavioral disorder category from final interrupted time series model of the natural logarithm of charges (plus $0.01) among the matched subset of Medicaid recipients.

Using the model results, we find that the difference between the observed and counterfactual values at the end of the study period for those with only mental disorders is 0.03 (95% CI [-0.17 – 0.24]) for enrolled individuals, and 0.22 (95% CI [0.01 – 0.43]) for non-enrolled individuals. The difference in the differences is 0.03 – 0.22= -0.19 (95% CI [-0.40, 0.02], p-value = 0.0816). For those with only SUDs, the difference in the differences is -0.97 (95% CI [0.11 - -2.05]) – 0.22 (95% CI [0.43 – 0.01] = -1.19 (95% CI [-2.03 – -0.36], p-value = 0.0050).

4.4.2 Behavioral Health Care

4.4.2.1 Unadjusted Model

An unadjusted fixed effects Poisson ITS model of the number of outpatient BH claims was count nature of the outcome measure. The base model included fixed effects for the parameters specified in 3.3 Models, which were specified in such a way that the change in slopes from the pre- to post-enrollment periods could be quantified and tested among enrolled and non-enrolled individuals. A Pearson's Chi-Squared overdispersion test determined that there was overdispersion in the model (p-value <0.001). Next, an unadjusted fixed effects negative binomial model was fit to the data, and the AIC decreased from 1,068,333 to 401,247.

Due to the large proportion of records with zero BH claims (57,339, 56%, Figure 29), a zero-inflated negative binomial model was fit to the data, resulting in a slightly higher AIC (401,249). Since the zero-inflation intercept was not significantly different than 1 (p-value = 0.167), zero-inflation was not incorporated into the base model.

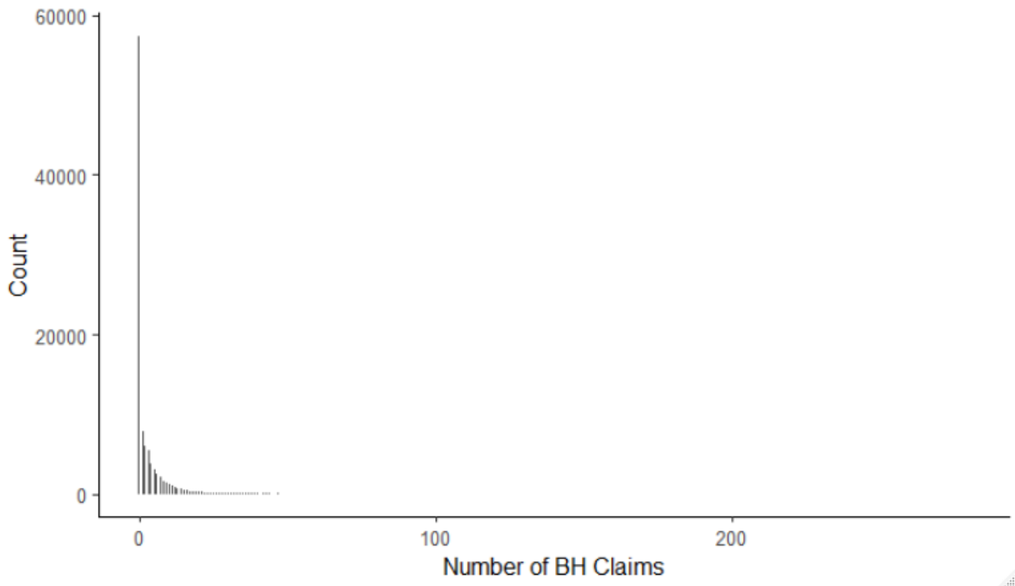


Figure 29. Histogram of the number of quarterly behavioral health claims among the matched subset of Medicaid recipients.

Adding covariates and interaction terms between behavioral disorder and the ITS terms improved the AIC to 370,708. Indicators of mental disorders and SUDs were included in place of the categorical term to resolve singularity in the model. Incorporating splines for continuous covariates further improved it to 370,210, and incorporating interactions between covariates and ITS terms further improved it to 368,175. Removing non-significant terms in a stepwise backwards fashion resulted in an AIC of 368,257.

4.4.2.2 Random Effects

Random intercepts for client and matched pair were added one at a time and then together as second and third levels. The glmmTMB package was used to fit mixed effects negative binomial models using Template Model Builder. The AIC value was lowest for the model with a random effect for individual (336,517), followed by the three-level model (337,039) and the model with a random effect for matched pair (353,855). The three-level model's random intercept for individual had a variance of 0.99, while the random intercept for matched pair had a variance of 0.40. The two-level model with random intercept for individual had a variance of 1.12 associated with the random intercept. Furthermore, a deviance test for goodness-of-fit did not show that the three-level model fit better than the two-level model with random intercept for individual (p-value = 1.0000). The final model was a two-level model with random intercept for individual.

4.4.2.3 Diagnostics

The final model demonstrates a pattern of positive residuals for lower fitted values, and negative residuals for higher fitted values of the number of BH claims (Figure 30). There is also a clear distinction between records with no BH claims and one or more BH claims.

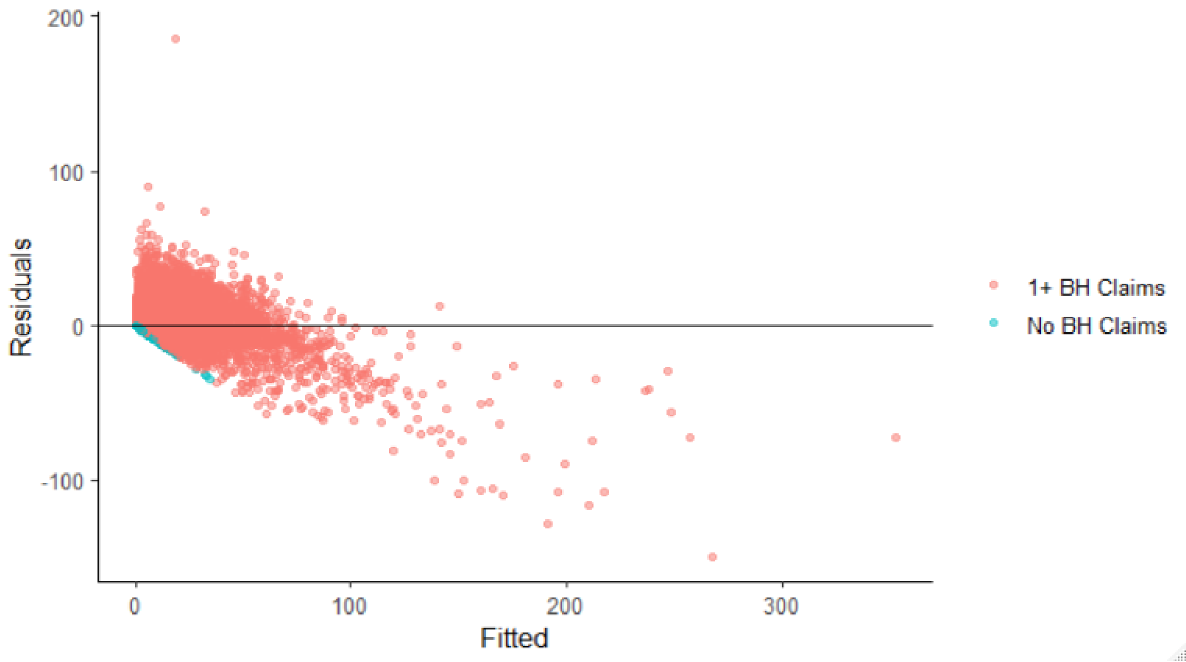


Figure 30. *Residuals plotted against fitted values from the adjusted mixed effects interrupted time series negative binomial model of the number of behavioral health claims among the matched subset of Medicaid recipients.*

A disproportionate number of negative residuals appear in the records with a higher number of total claims (Figure 31).

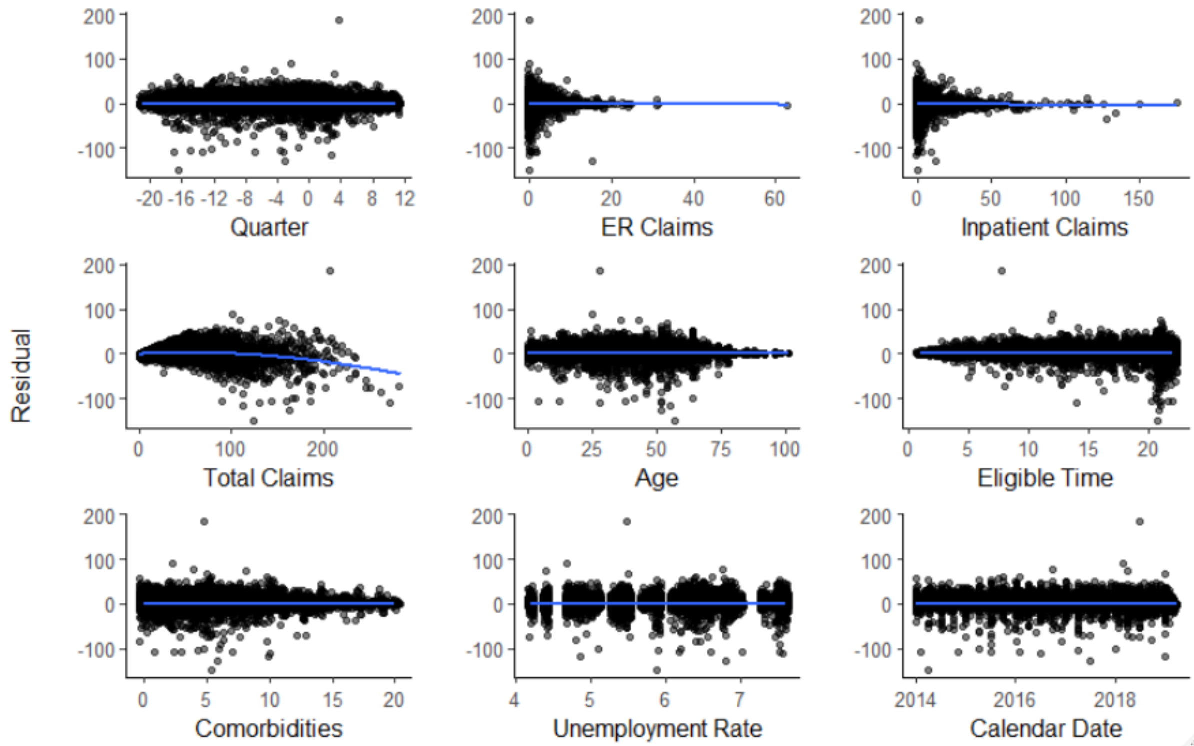


Figure 31. Pearson residuals by covariates from the adjusted mixed effects interrupted time series negative binomial model of the number of behavioral health claims among the matched subset of Medicaid recipients.

4.4.2.4 Final Model

The coefficients of the main effects of interest in the final ITS model are in Table 7. Final effects correspond with the predicted values in the following figures after adding slope terms together. The baseline slope of total charges among enrolled individuals was 0.02. In the post-enrollment period, this slope decreased by 0.01, resulting in a slope of 0.01, and this decrease in slope was not statistically significant (p=0.0989). For non-enrolled individuals, the baseline slope of the total charges was 0.01, and increased significantly to 0.03 in the post-enrollment period (0.03, p<0.0100). Non-enrolled individuals with a substance use disorder only had a significantly higher intercept in the pre-enrollment period compared to those with mental disorders only (1.01, p=0.0148). There were no statistically significant differences by behavioral disorders in the post-enrollment period. The complete table of model results, including covariates, is in Appendix B.

Table 7. Covariate coefficients of final interrupted time series negative binomial model of the number of behavioral health claims among the matched subset of Medicaid recipients.

Segment	Term	Coefficient	p-value	Final Effect
Overall				
Enrollees at Baseline	Intercept	0.16	0.2802	0.16
	Slope	0.01	0.0459	0.01
Non-Enrollees at Baseline	Intercept	-0.12	0.4878	-0.12
	Slope	0.01	0.7867	0.01
Enrollees at Follow-up	Intercept	0.40	0.0110	0.40
	Slope	-0.08	<0.0001	-0.07
Non-Enrollees at Follow-up	Intercept	-0.23	0.2418	-0.23
	Slope	-0.03	0.0066	-0.02
Effect of Mental Disorder				
Non-Enrollees at Follow-up	Slope	0.03	0.3973	0.01
Effect of Substance Use Disorder				
Enrollees at Baseline	Intercept	0.12	0.0076	0.28
	Slope	-0.01	0.0016	0.00
Non-Enrollees at Baseline	Intercept	0.08	0.1192	-0.04
	Slope	-0.01	0.0044	0.00
Enrollees at Follow-up	Intercept	0.21	<0.0001	0.61
	Slope	-0.04	<0.0001	-0.12
Non-Enrollees at Follow-up	Intercept	-0.02	0.7544	-0.25
	Slope	0.04	0.0028	0.01

Figure 32 demonstrates the predicted number of BH claims for individuals with no inpatient or emergency claims, no BH medication, only mental disorders, and average values of all other covariates. The results indicate that both enrolled and non-enrolled individuals experienced an increase in the number of BH claims during the pre-enrollment period, but during the post-enrollment period, BH claims continued to increase for non-enrolled individuals and decreased for enrolled individuals.

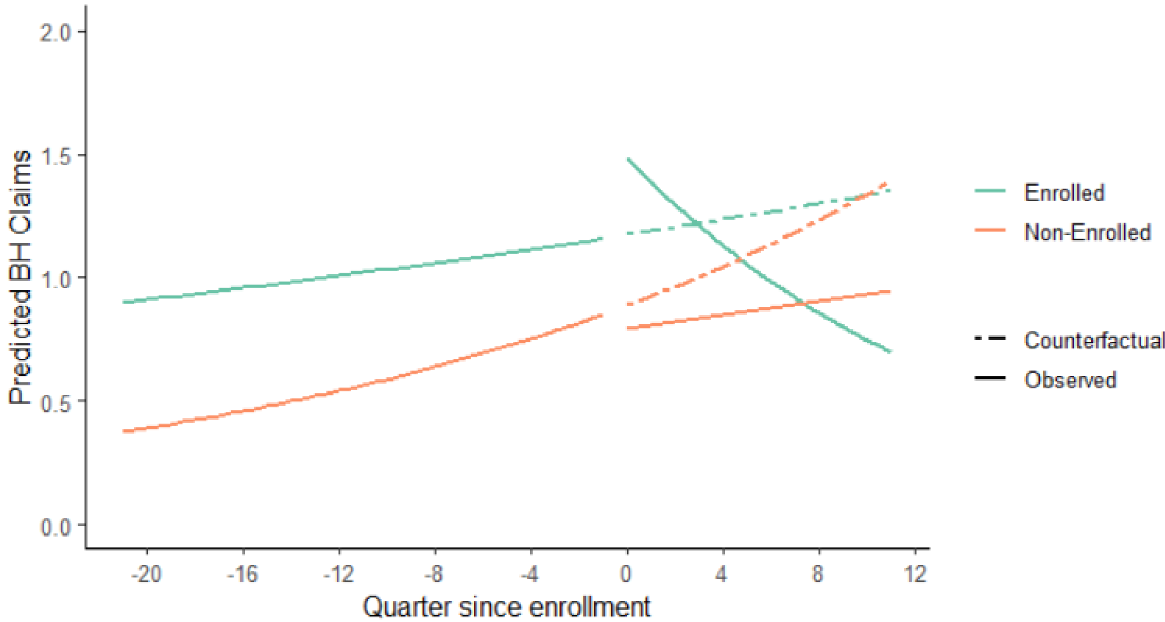


Figure 32. Predicted values by enrollment and time from the final interrupted time series negative binomial model of the number of behavioral health claims among the matched subset of Medicaid recipients.

The segments of the regression model are similar for those with both a mental disorder and a SUD (Figure 33).

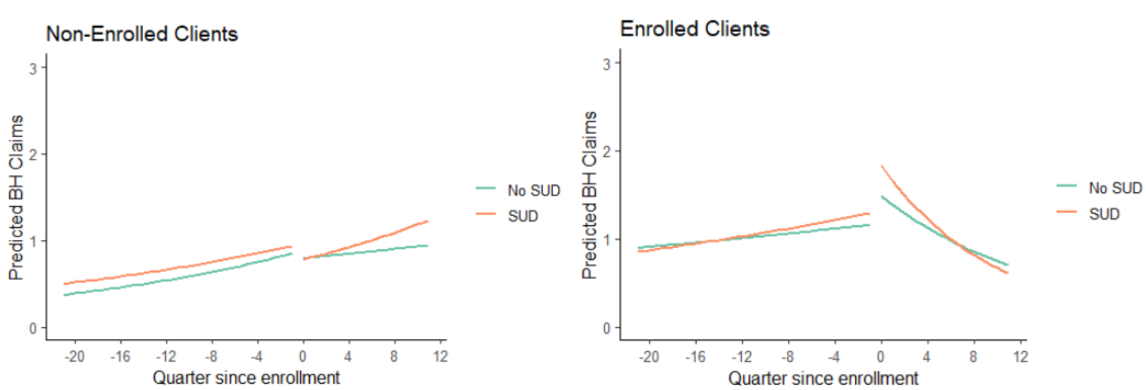


Figure 33. Predicted values by enrollment, time, and behavioral disorder category from the final interrupted time series negative binomial model of the number of behavioral health claims among the matched subset of Medicaid recipients.

Using the model results, we find that the difference between the observed and counterfactual values at the end of the study period for those with only mental disorders is -

0.66 (95% CI [-0.46 – -0.86]) for enrolled individuals, and -0.39 (95% CI [-0.07 – -0.70]) for non-enrolled individuals. The difference in the differences is -0.66 – -0.39= -0.27 (95% CI [-0.55 – 0.00], p-value = 0.0537). For those with only SUDs, the difference in the differences is -1.01 (95% CI [-0.80 - -1.21]) – -0.10 (95% CI [0.21 – -0.40] = -0.91 (95% CI [-1.18 – -0.64], p-value<0.0001).

4.4.3 Emergency Health Care

4.4.3.1 Unadjusted Model

The base fixed effects binomial ITS model of the probability of emergency healthcare visits included fixed effects for the parameters specified in 3.3 Models, which were specified in such a way that the change in slopes from the pre- to post-enrollment periods could be quantified and tested among enrolled and non-enrolled individuals. Adding covariates and interaction terms between behavioral disorder and the ITS terms improved the AIC from 112,761 to 106,019. Indicators of mental disorders and SUDs were included in place of the categorical term to resolve singularity in the model. Incorporating splines for continuous covariates further improved it to 102,362. Removing non-significant terms in a stepwise backwards fashion resulted in an AIC of 102,355.

4.4.3.2 Random Effects

Random intercepts for client and matched pair were added one at a time and then together as second and third levels. The AIC value was lowest for the two-level model including a random intercept for individual (96,532), followed by the three-level model (96,534) and the two-level model including a random intercept for matched pair (101,799). The three-level model’s random intercept for individual had a variance of 1.68, while the random intercept for matched pair had a variance of 0.01. The two-level model with random intercept for individual had a variance of 1.69 associated with the random intercept. Furthermore, a deviance test for goodness-of-fit did not show that the three-level model fit better than the two-level model with random intercept for individual (p-value = 1.0000). The final model was a two-level model with random intercept for individual.

4.4.3.3 Diagnostics

Inspection of the model residuals demonstrates a pattern in the residuals with an increasing number of inpatient claims. This potential covariate was removed during the model building process due to resulting fitting probabilities of zero and one. Many emergency healthcare visits result in admissions to inpatient care, which results in a relationship between these two variables. Since the inpatient claims often result from emergency claims, the decision was made not to include the number of inpatient claims as a predictor of the probability of emergency care.

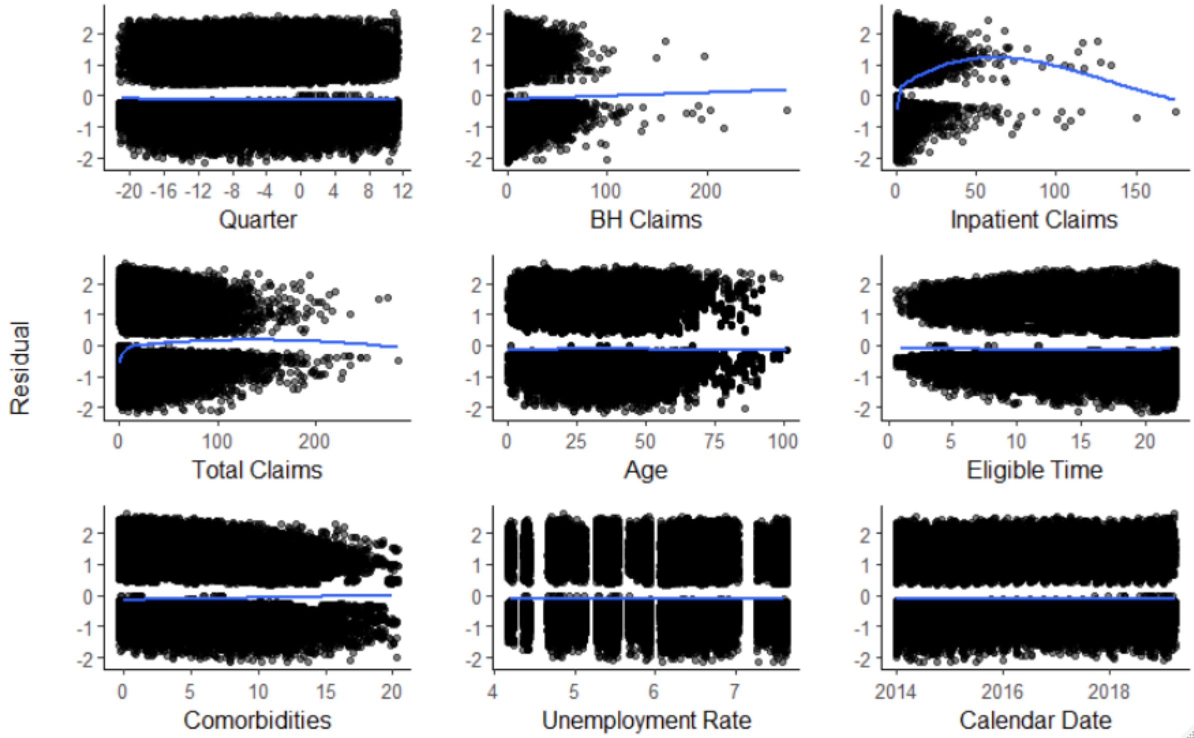


Figure 34. Pearson residuals by potential covariates from adjusted binomial mixed effects interrupted time series model of the odds of emergency healthcare among the matched subset of Medicaid recipients.

4.4.3.4 Final Model

The coefficients of the main effects of interest in the final ITS model are in Table 8. Final effects correspond with the predicted values in the following figures after adding slope terms together. The baseline slope of the probability of emergency care among enrolled individuals was 0.05. In the post-enrollment period, this slope decreased by 0.07, resulting in a slope of -0.02, and this decrease in slope was statistically significant (p=<0.0001). For non-enrolled individuals, the baseline slope was 0.03, which did not change in the post-enrollment period (p=0.8370). Enrolled and non-enrolled individuals with a substance use disorder had significantly higher intercepts in the pre-enrollment period (p=0.0018 and 0.0001, respectively) and the post-enrollment period (p=0.0080 and 0.0077, respectively), but the slopes were not statistically different. The complete table of model results, including covariates, is in Appendix B.

Table 8. Covariate coefficients of the final interrupted time series binomial model of the probability of emergency healthcare claims among the matched subset of Medicaid recipients.

Segment	Term	Coefficient	p-value	Final Effect
Enrollees at Baseline	Intercept	-1.44	<0.0001	-1.44
	Slope	0.05	<0.0001	0.05
Non-Enrollees at Baseline	Intercept	-1.60	<0.0001	-1.60
	Slope	0.03	<0.0001	0.03
Enrollees at Follow-up	Intercept	-1.60	<0.0001	-1.60
	Slope	-0.07	<0.0001	-0.02
Non-Enrollees at Follow-up	Intercept	-13.21	0.3815	-13.21
	Slope	-0.00	0.8370	0.03
Effect of Mental Disorder				
Non-Enrollees at Follow-up	Intercept	11.58	0.4427	-1.63
Effect of Substance Use Disorder				
Enrollees at Baseline	Intercept	0.22	0.0018	-1.22
	Slope	-0.01	0.1501	0.04
Non-Enrollees at Baseline	Intercept	0.29	0.0001	-1.31
	Slope	-0.01	0.1681	0.02
Enrollees at Follow-up	Intercept	0.21	0.0080	-1.39
	Slope	0.03	0.0712	0.00
Non-Enrollees at Follow-up	Intercept	0.25	0.0077	-12.96
	Slope	0.03	0.1747	0.05

Figure 35 demonstrates the predicted odds of emergency health care for individuals with no BH medication, only mental disorders, and average values of all other covariates. The results indicate that both enrolled and non-enrolled individuals experienced increases in the odds of emergency care during the pre-enrollment period, but enrolled individuals' slope became negative in the post-enrollment period while non-enrolled individuals' odds of emergency care continued to increase. Counterfactual trends are included in the figure, demonstrating that enrolled individuals had a higher odds of emergency care than non-enrolled individuals at the time of enrollment, but a much lower odds at the end of the study period

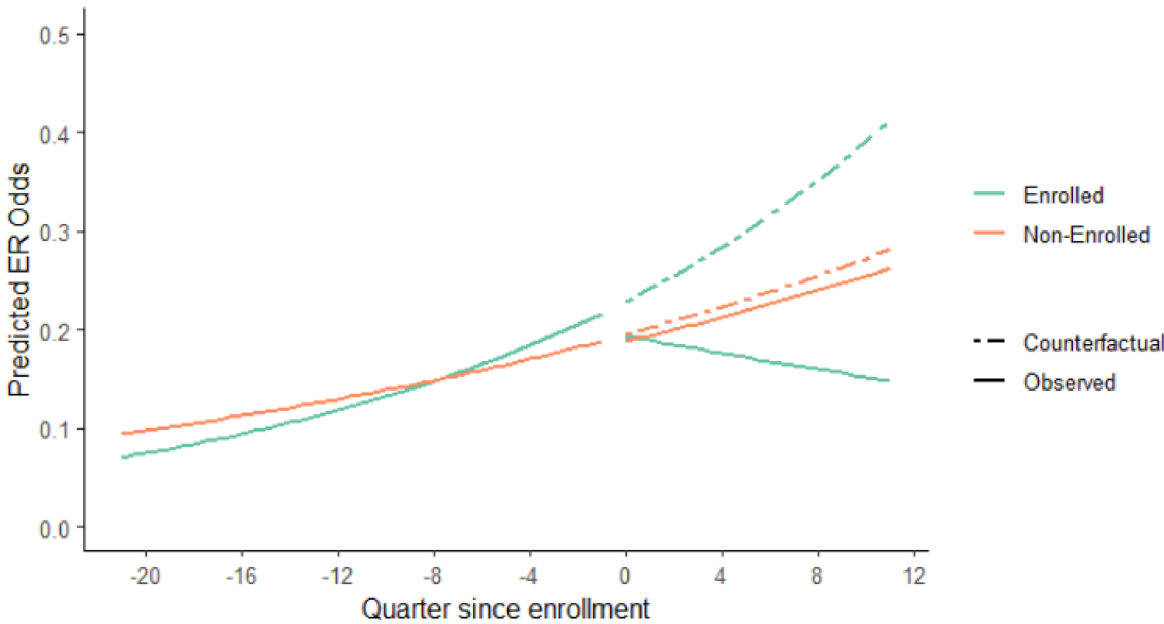
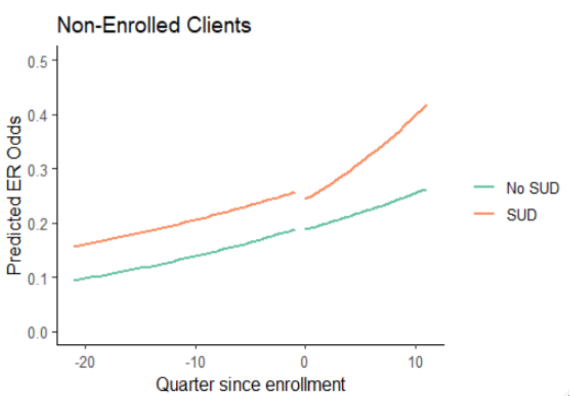


Figure 35. Predicted values by enrollment and time from the final interrupted time series binomial model of the probability of emergency claims among the matched subset of Medicaid recipients.

The segments of the regression model differ by type of behavioral disorder. Those with a SUD had a higher odds of emergency care throughout the study period.



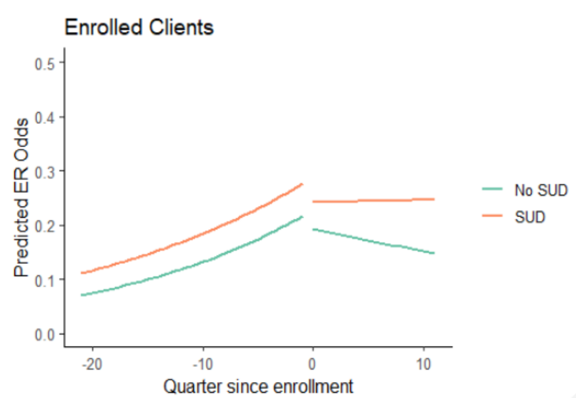


Figure 36. Predicted values by enrollment, time, and substance use disorder from the final interrupted time series binomial model of the odds of emergency health claims among the matched subset of Medicaid recipients.

Using the model results, we find that the difference between the observed and counterfactual values at the end of the study period for those with only mental disorders is -

0.98 (95% CI [-0.60 – -1.37]) for enrolled individuals, and -0.07 (95% CI [0.32 – -0.46]) for non-enrolled individuals. The difference in the differences is -0.98 – -0.07= -0.91 (95% CI [-1.32 - -0.51], p-value = <0.0001). For those with only SUDs, the difference in the differences is -0.63 (95% CI [-0.29 - -0.97]) – -11.40 (95% CI [27.37 – -50.16] = 10.77 (95% CI [-18.8 – 40.3], p-value = 0.4754).

5 Discussion

5.1 Conclusions

The findings of this study indicate that enrollment in the CLNM program resulted in decreased probability of emergency or urgent health care for enrollees during the study period. Additionally, the subset of individuals with SUD and no mental disorder experienced a decrease in charges, and the rest of the enrollees are projected to have lower charges than those of non-enrollees within five years of enrollment. However, among the same subset, there was a decrease in outpatient BH claims for enrolled individuals compared to non-enrolled individuals.

The average quarterly charges for enrolled individuals were projected to be $1,412 at the end of the study period based on their pre-enrollment data, but was estimated to be $1,460. However, the charges were increasing in the pre-enrollment period and decreasing in the post-enrollment period, and projected to be lower than the pre-enrollment prediction within a year of the study end. Furthermore, the average quarterly charges for non-enrolled individuals was increasing at a faster rate in the post-enrollment period than it was in the pre-enrollment period, resulting in an average quarterly charges of $1,220 instead of the projected $977. When these results were compared statistically, the non-enrolled group did not have a statistically larger difference in charges compared to projections (p-value = 0.0816), but the difference was statistically significant for the subgroup of individuals with a SUD and no mental disorder (p-value = 0.0050). This was due primarily to the much larger pre-enrollment average quarterly charges in this group for enrolled individuals.

The average number of BH claims for enrolled individuals was projected to be 1.8 at the end of the study period based on their pre-enrollment data, but was estimated to be 0.4. While these individuals' claims were increasing in the pre-enrollment period, the intercept decreased in the post-enrollment period, and the number of claims began decreasing, while the non-enrolled individuals' claims continued increasing. The difference in the changes between the two groups was not statistically significant (p-value = 0.4143). However, the enrolled individuals with only SUD had on average 30.6 BH claims at the end of the study period, while non-enrolled individuals with only SUD had only 1.6. The difference in the changes between these two groups was statistically significant (p-value <0.0001).

The average odds of having an emergency or urgent care claim for enrolled individuals was projected to be 0.41 at the end of the study period based on their pre-enrollment data, but was estimated to be 0.15. This was largely due to the change in trend from positive in the pre-enrollment period to negative in the post-enrollment period. The average odds for non-enrolled individuals was projected to be 0.28 and estimated to be 0.26 due to a slight decrease in the rate of increase. When these results were compared statistically, the enrolled group did have a statistically larger change in odds compared to projections (p-value<0.0001), but the difference was not statistically significant for the subgroup of individuals with a SUD and no mental disorder (p-value = 0.4754). In both the enrolled and non-enrolled groups, individuals with only SUD continued to have an increase in their odds of emergency claims, although the rate of increase slowed for enrolled individuals and quickened for non-enrolled individuals.

5.2 Limitations

There are several limitations of this study. Some of these limitations are related to the use of Medicaid claims data as the primary data source. There were 3 enrolled individuals whose CLNM provider could not be determined, either because they saw two providers an equal number of times or because they never had a claim filed by a CLNM provider. Unfortunately, there was no list available of the individuals enrolled by each provider. Indian Health Service charges were difficult to categorize into categories of care due to their coding in the claims database. Some BH, emergency, and inpatient claims may have been categorized as "other" charges. Monthly Medicaid eligibility information was unavailable for recipients. Instead, the assumption was made that recipients were eligible for all months between their first and last Medicaid claim in the study period. As a result, the number of time periods without claims may be inflated. There were 2 matched individuals who never had any charges associated with any of their claims. Since many of these individuals have claims (even inpatient claims in some cases), they may be dual-enrolled recipients of Medicaid and Medicare. Medicare claims were unavailable for this analysis. Access to screening services and changes in physical health indicators could not be assessed due to the limitations of claims data. Other studies demonstrate beneficial changes in diabetes screenings, dyslipidemia screenings, blood pressure monitoring, hepatitis and tuberculosis screenings, colon cancer screenings, education in nutrition, weight management, education in exercise, education on smoking, patient satisfaction with their medical care, cholesterol, blood pressure, and weight loss [44]. Screening data also would have informed the matching process further.

The study was also limited in its scope due to the availability of information about the CLNM program. In particular, this analysis would have benefited from a comparison of the reimbursement costs saved to the program costs invested. Unfortunately, the cost of the program itself has never been calculated in New Mexico, so it is not possible to account for the cost in relation to any savings observed. A two-year evaluation of Iowa Medicaid's Integrated Health Home Program found that $37 million were saved, but $47 million were spent in tier payments (payments to the care providers) and outreach expenses [45]. In this evaluation, the average savings per member per month was $110. In a report to the U.S. Congress about 32 Health Home programs implemented in 21 states and the District of Columbia, the authors note that the estimates of cost savings might be low, since the savings examined in the available studies are limited to Medicaid, and do not account for potential savings in other programs such as Medicare that might result from the improved health status of Medicaid enrollees who receive health home services [46]. Additionally, there were no outcome measures to test that were not expected to change for enrollees during the post-enrollment period. Multiple outcomes should be studied when conducting an ITS analysis [47]. Observing similar effects of enrollment on multiple, independent outcomes is strong evidence of a causal relationship between enrollment and changes in the participants, but the evidence can be strengthened by demonstrating that changes were targeted and specific.

Further work on the analysis is warranted. The residuals of the model of charges indicate that a two-stage model (or hurdle model) may better predict trends in charges. The non-zero charges demonstrated a good fit in the linear model, but the charges of zero differed in their residuals. A two-stage model would require the probability of zero charges to be modeled first, and then the charges with non-zero values. Second, an ongoing evaluation of this current program would provide more accurate predictions of trends. With a longer period of time, seasonality and autocorrelation may be accurately assessed and incorporated into the models.

5.3 Strengths

This study reduced bias in measurements of treatment effects in a retrospective, observational study by comparing observed trends among enrollees to their own baseline and a matched group of non-enrollees. This allowed the policy effect to be separated from trends over time and differences in different communities. Additionally, the policy effect was separated from impacts of concurrent events because each individual enrolled at a different time, and the change in outcomes after their initial enrollment was estimated. This was possible by including random effects in the models.

The current CLNM program does not target individuals with substance use disorder (SUD). This study revealed that 45% of enrollees had SUD, and furthermore identified 19 individuals with SUD and no mental disorder enrolled in CLNM based on other criteria, and matched 20 similar individuals to them.

Appendices

Appendix A. R Code

Reference List of Variables:

Cost_adjust: Adjusted total charges

Count_bh: Number of behavioral health claims

Er: Indicator of any emergency claims

New_qtr: Study quarter relative to matched pair's enrollment date

Hhmember: Indicator of enrollment in CLNM program

Period: Indicator of follow-up time period

Disorder: Behavioral disorder category

Elix_ctr: Number of physical conditions centered on mean

Unemp_ctr: Unemployment rate centered on mean

Cal_ctr: Calendar date centered on mean

Qtr_ctr: Number of quarters of study eligibility centered on mean

Count_inpt: Number of inpatient claims

Age_ctr: Age in years centered on mean

Count_er: Number of emergency claims

Ctr_total: Number of total claims

Prov_county: Provider county group

Mental: Indicator of mental disorder

Sud: Indicator of SUD

Final Adjusted Model of Total Charges:

```
lmer(log(cost_adjust+0.01) ~ 0

    #Intervention model at baseline

    +I((new_qtr)*I(hhmember == 1)) + I(I(period == 0)*I(hhmember == 1))

        + disorder*I(I(period == 0)*I(hhmember == 1))

    #Intervention model at follow-up

    +I(period*(new_qtr)*I(hhmember == 1)) + I(I(period == 1)*I(hhmember == 1))

    #Comparison model at baseline

    +I((new_qtr)*I(hhmember == 0)) + I(I(period == 0)*I(hhmember == 0))

    #Comparison model at follow-up

    +I(period*(new_qtr)*I(hhmember == 0)) + I(I(period == 1)*I(hhmember == 0))


    #Covariates and splines

    +bs(elix_ctr, degree = 2)+bs(unemp_ctr, degree = 3)

    +bs(cal_ctr, degree = 4)+bs(qtr_ctr, degree = 5)

    +bs(count_bh, degree = 5)+bs(count_inpt, degree = 5)

    +bs(age_ctr, degree = 5)+bs(count_er, degree = 5)

    +bs(ctr_total, degree = 5)+prov_county+disorder


    +(1 | client_system_id),

    data = quarters, na.action = na.omit, REML = FALSE)
```

Final Adjusted Model of Number of the Behavioral Health Claims:

```
bh11 <- glmmTMB(count_bh ~ 0

        #Intervention model at baseline

        + I((new_qtr)*I(hhmember == 1)) + I(I(period == 0)*I(hhmember == 1))

        + I(sud*(new_qtr)*I(hhmember == 1)) + I(sud*I(period == 0)*I(hhmember == 1))

        + bhmed*I((new_qtr)*I(hhmember == 1))

        + ctr_total*I((new_qtr)*I(hhmember == 1))

        + ctr_total*I(I(period == 0)*I(hhmember == 1))

        + elix_ctr*I((new_qtr)*I(hhmember == 1))

        + qtr_ctr*I((new_qtr)*I(hhmember == 1))

        + qtr_ctr*I(I(period == 0)*I(hhmember == 1))

        + I(count_inpt*(new_qtr)*I(hhmember == 1))

        + I(count_inpt*I(period == 0)*I(hhmember == 1))

        + I(count_er*(new_qtr)*I(hhmember == 1))

        + I(count_er*I(period == 0)*I(hhmember == 1))

        + I(age_ctr*(new_qtr)*I(hhmember == 1))

        + I(age_ctr*I(period == 0)*I(hhmember == 1))

        + I(unemp_ctr*(new_qtr)*I(hhmember == 1))

        + I(unemp_ctr*I(period == 0)*I(hhmember == 1))

        + I(cal_ctr*I(period == 0)*I(hhmember == 1))


        #Intervention model at follow-up

        + I(I(period == 1)*I(hhmember == 1)) + I(sud*I(period == 1)*I(hhmember == 1))
```

+ I(sud*I(period == 0)*I(hhmember == 0))

+ bhmed*I(I(period == 1)*I(hhmember == 1))

+ ctr_total*I(I(period == 1)*I(hhmember == 1))

+ elix_ctr*I(I(period == 1)*I(hhmember == 1))

+ I(count_er*I(period == 1)*I(hhmember == 1))

+ I(age_ctr*I(period == 1)*I(hhmember == 1))

+ I(cal_ctr*I(period == 1)*I(hhmember == 1))

+ I(period*(new_qtr)*I(hhmember == 1))

+ I(sud*period*(new_qtr)*I(hhmember == 1))

+ I(sud*(new_qtr)*I(hhmember == 0))

+ ctr_total*I(period*(new_qtr)*I(hhmember == 1))

+ elix_ctr*I(period*(new_qtr)*I(hhmember == 1))

+ I(count_inpt*period*(new_qtr)*I(hhmember == 1))

+ I(age_ctr*period*(new_qtr)*I(hhmember == 1))

#Comparison model at baseline

+ I(I(period == 0)*I(hhmember == 0)) + I(cal_ctr*I(period == 0)*I(hhmember == 0))

+ I((new_qtr)*I(hhmember == 0))

+ I(mental*(new_qtr)*I(hhmember == 0))

+ I(sud*period*(new_qtr)*I(hhmember == 0))

+ ctr_total*I((new_qtr)*I(hhmember == 0))

+ qtr_ctr*I((new_qtr)*I(hhmember == 0))

```
        + I(count_inpt*(new_qtr)*I(hhmember == 0))
        + I(age_ctr*(new_qtr)*I(hhmember == 0))

#Comparison model at follow-up
+ I(I(period == 1)*I(hhmember == 0)) + I(sud*I(period == 1)*I(hhmember == 0))
+ I(period*(new_qtr)*I(hhmember == 0))
        + ctr_total*I(period*(new_qtr)*I(hhmember == 0))
        + I(count_inpt*period*(new_qtr)*I(hhmember == 0))

#Covariates and splines
+bhmed+ctr_total+elix_ctr+qtr_ctr+bs(count_inpt, degree = 2)
+bs(count_er, degree = 3)+bs(age_ctr, degree = 4)
+bs(unemp_ctr, degree = 4)+bs(cal_ctr, degree = 5)

+(1 | client_system_id),
ziformula=~0, data = quarters, na.action = na.omit, family = "nbinom1")
```

Final Adjusted Model of ER Probability:

```
glmer(er ~ 0
	#Intervention model at baseline
	+ I(I(period == 0)*I(hhmember == 1)) + I(sud*I(period == 0)*I(hhmember == 1))
	+ I((new_qtr)*I(hhmember == 1)) + I(sud*(new_qtr)*I(hhmember == 1))
	#Intervention model at follow-up
	+ I(I(period == 1)*I(hhmember == 1)) + I(sud*I(period == 1)*I(hhmember == 1))
	+ I(period*(new_qtr)*I(hhmember == 1))
		+ I(sud*period*(new_qtr)*I(hhmember == 1))
	#Comparison model at baseline
	+ I(I(period == 0)*I(hhmember == 0)) + I(sud*I(period == 0)*I(hhmember == 0))
	+ I((new_qtr)*I(hhmember == 0)) + I(sud*(new_qtr)*I(hhmember == 0))
	#Comparison model at follow-up
	+ I(I(period == 1)*I(hhmember == 0)) + I(mental*I(period == 1)*I(hhmember == 0))
		+ I(sud*I(period == 1)*I(hhmember == 0))
	 + I(period*(new_qtr)*I(hhmember == 0))
		+ I(sud*period*(new_qtr)*I(hhmember == 0))

	#Covariates and splines
	 +bs(age_ctr, degree = 5)+bs(elix_ctr, degree = 2)+qtr_ctr
	 +bs(cal_ctr, degree = 4) +bhmed

	+(1 | client_system_id), data = quarters, na.action = na.omit, family = "binomial")
```

Appendix B. Model Results

Table B.1. Covariate coefficients of the final interrupted time series model of the natural logarithm of charges (plus $0.01) among the matched subset of Medicaid recipients.

Segment	Term	Coefficient	p-value
Overall			
Enrollees at Baseline	Intercept	3.28	<0.0001
	Slope	0.02	0.0002
Non-Enrollees at Baseline	Intercept	3.00	<0.0001
	Slope	0.01	0.0344
Enrollees at Follow-up	Intercept	3.44	<0.0001
	Slope	-0.01	0.0989
Non-Enrollees at Follow-up	Intercept	3.00	<0.0001
	Slope	0.02	0.0100
Effect of Disorder			
Enrollees at Baseline	Intercept: Mental Disorder Only	Ref	Ref
	Intercept: Substance Use Disorder Only	0.21	0.4208
	Intercept: Both Disorders	0.02	0.5837
Non-Enrollees at Baseline	Intercept: Mental Disorder Only	Ref	Ref
	Intercept: Substance Use Disorder Only	1.01	0.0148
	Intercept: Both Disorders	-0.04	0.3320
Covariates			
Number of Physical Comorbidities	First degree polynomial	-0.55	<0.0001
	Second degree polynomial	0.26	0.0738
Unemployment Rate	First degree polynomial	-0.87	<0.001
	Second degree polynomial	0.02	0.7685
	Third degree polynomial	-0.49	<0.0001
Calendar Date	First degree polynomial	-1.21	<0.0001
	Second degree polynomial	-0.41	0.0044
	Third degree polynomial	-2.68	<0.0001
	Fourth degree polynomial	0.02	0.8765
Length of Study Eligibility	First degree polynomial	-1.78	0.0014
	Second degree polynomial	0.21	0.6275
	Third degree polynomial	-2.28	<0.0001
	Fourth degree polynomial	-1.64	<0.0001
	Fifth degree polynomial	-1.68	<0.0001
Number of Behavioral Health Claims	First degree polynomial	-1.26	<0.0001
	Second degree polynomial	6.83	<0.0001
	Third degree polynomial	-0.34	<0.0001
	Fourth degree polynomial	0.98	<0.0001
	Fifth degree polynomial	-0.96	<0.0001
Number of Inpatient Claims	First degree polynomial	4.65	<0.0001
	Second degree polynomial	-8.53	<0.0001
	Third degree polynomial	0.11	0.0091

	Fourth degree polynomial	-4.57	0.2981
	Fifth degree polynomial	0.11	<0.0001
Age	First degree polynomial	1.09	0.0130
	Second degree polynomial	-3.88	<0.0001
	Third degree polynomial	3.02	0.0030
	Fourth degree polynomial	-4.69	<0.0001
	Fifth degree polynomial	5.22	<0.0001
Number of Emergency Claims	First degree polynomial	5.35	<0.0001
	Second degree polynomial	-0.44	<0.0001
	Third degree polynomial	0.01	<0.0001
	Fourth degree polynomial	-0.02	<0.0001
	Fifth degree polynomial	6.89	0.0015
Number of Total Claims	First degree polynomial	0.42	<0.0001
	Second degree polynomial	-0.90	<0.0001
	Third degree polynomial	0.02	<0.0001
	Fourth degree polynomial	-0.02	<0.0001
	Fifth degree polynomial	0.01	<0.0001
Provider County Group	Bernalillo	-0.19	0.0001
	Curry	-0.28	<0.0001
	Hidalgo, Grant	0.12	0.1308
	Lea	-0.36	<0.0001
	Other	-0.02	0.7745
	Quay, De Baca, Roosevelt	-0.12	0.0461
	Sandoval	0.03	0.6750

Table B.2. Covariate coefficients of the final interrupted time series model of the number of behavioral health claims among the matched subset of Medicaid recipients.

Segment	Term	Coefficient	p-value
Overall			
Enrollees at Baseline	Intercept	0.16	0.2802
	Intercept – Number of Inpatient Claims	-0.01	<0.0001
	Intercept – Number of Emergency Claims	-0.02	0.0101
	Intercept – Age	>-0.01	0.0645
	Intercept – Unemployment Rate	-0.08	<0.0001
	Intercept – Calendar Date	<0.01	0.2147
	Intercept – Behavioral Medication	-0.03	0.2195
	Intercept – Total Claims	-0.01	<0.0001
	Intercept – Eligibility Time	0.01	<0.0001
	Slope	0.01	0.0459
	Slope – Number of Inpatient Claims	<0.01	<0.0001
	Slope – Number of Emergency Claims	<0.01	0.4067
	Slope – Age	<0.01	<0.0001
	Slope – Unemployment Rate	-0.01	<0.0001
	Slope – Behavioral Medication	>-0.01	0.0348
	Slope – Total Claims	>-0.01	<0.0001
	Slope – Physical Conditions	>-0.01	<0.0001
	Slope – Eligibility Time	>-0.01	<0.0001
Non-Enrollees at Baseline	Intercept	-0.12	0.4878
	Intercept – Calendar Date	>-0.01	<0.0001
	Slope	0.01	0.7867
	Slope – Number of Inpatient Claims	0.01	<0.0001
	Slope – Age	<0.01	<0.0001
	Slope – Total Claims	>-0.01	<0.0001
	Slope – Eligibility Time	<0.01	0.0169
Enrollees at Follow-up	Intercept	0.40	0.0110
	Intercept – Number of Emergency Claims	-0.04	<0.0001
	Intercept – Age	>-0.01	0.0114
	Intercept – Calendar Date	<0.01	<0.0001
	Intercept – Total Claims	-0.01	<0.0001
	Intercept – Physical Conditions	0.02	<0.0001
	Slope	-0.08	<0.0001
	Slope – Number of Inpatient Claims	-0.01	<0.0001
	Slope – Age	<0.01	<0.0001
	Slope – Total Claims	<0.01	<0.0001
	Slope – Physical Conditions	0.01	<0.0001
Non-Enrollees at Follow-up	Intercept	-0.23	0.2418
	Slope	-0.03	0.0066
	Slope – Number of Inpatient Claims	-0.03	<0.0001
	Slope – Total Claims	<0.01	<0.0001

Effect of Mental Disorder			
Non-Enrollees at Baseline	Slope	0.03	0.3973
Effect of Substance Use Disorder			
Enrollees at Baseline	Intercept	0.12	0.0076
	Slope	-0.01	0.0016
Non-Enrollees at Baseline	Intercept	0.08	0.1192
	Slope	-0.01	0.0044
Enrollees at Follow-up	Intercept	0.21	<0.0001
	Slope	-0.04	<0.0001
Non-Enrollees at Follow-up	Intercept	-0.02	0.7544
	Slope	0.04	0.0028
Covariates			
Number of Physical Comorbidities		-0.12	<0.0001
Unemployment Rate	First degree polynomial	0.01	0.8630
	Second degree polynomial	0.05	0.4278
	Third degree polynomial	0.04	0.5311
	Fourth degree polynomial	0.05	0.1100
Calendar Date	First degree polynomial	0.02	0.8754
	Second degree polynomial	0.08	0.5850
	Third degree polynomial	0.06	0.7659
	Fourth degree polynomial	0.08	0.6487
	Fifth degree polynomial	-0.10	0.6137
Length of Study Eligibility		0.01	<0.0001
Number of Inpatient Claims	First degree polynomial	-0.09	0.6824
	Second degree polynomial	-0.02	0.9721
Age	First degree polynomial	0.09	0.8048
	Second degree polynomial	-0.02	0.9672
	Third degree polynomial	-0.04	0.9477
	Fourth degree polynomial	-0.03	0.9605
Number of Emergency Claims	First degree polynomial	-0.06	0.7454
	Second degree polynomial	-0.01	0.9892
	Third degree polynomial	-0.00	0.9970
Number of Total Claims		0.04	<0.0001
Behavioral Health Medication		0.26	<0.0001

Table B.3. Covariate coefficients of the final interrupted time series model of the probability of emergency care claims among the matched subset of Medicaid recipients.

Segment	Term	Coefficient	p-value
Overall			
Enrollees at Baseline	Intercept	-1.44	<0.0001
	Slope	0.05	<0.0001
Non-Enrollees at Baseline	Intercept	-1.60	<0.0001
	Slope	0.03	<0.0001
Enrollees at Follow-up	Intercept	-1.60	<0.0001
	Slope	-0.07	<0.0001
Non-Enrollees at Follow-up	Intercept	-13.21	0.3815
	Slope	-0.00	0.8370
Effect of Mental Disorder			
Non-Enrollees at Follow-up	Intercept	11.58	0.4427
Effect of SUD			
Enrollees at Baseline	Intercept	0.22	0.0018
	Slope	-0.01	0.1501
Non-Enrollees at Baseline	Intercept	0.29	0.0001
	Slope	-0.01	0.1681
Enrollees at Follow-up	Intercept	0.21	0.0080
	Slope	0.03	0.0712
Non-Enrollees at Follow-up	Intercept	0.25	0.0077
	Slope	0.03	0.1747
Covariates			
Number of Physical Comorbidities	First degree polynomial	2.30	<0.0001
	Second degree polynomial	2.66	<0.0001
Calendar Date	First degree polynomial	0.64	<0.0001
	Second degree polynomial	-1.98	<0.0001
	Third degree polynomial	0.08	0.6338
	Fourth degree polynomial	-0.93	<0.0001
Length of Study Eligibility		-0.03	<0.0001
Age	First degree polynomial	-1.40	0.2719
	Second degree polynomial	4.46	0.0338
	Third degree polynomial	-9.06	0.0162
	Fourth degree polynomial	5.01	0.1324
	Fifth degree polynomial	-3.43	0.1372
Behavioral Medication		0.59	<0.0001

References

1. Walden G. *H.R.6 - SUPPORT for Patients and Communities Act*.; 2017. Accessed July 2, 2021. https://www.congress.gov/bill/115th-congress/house-bill/6

2. *Behavioral Health Services for Medicaid Expansion Adults*. New Mexico Legislative Finance Committee; 2018.

3. New Mexico Human Services Department. *Provider Policy Manual: CareLink NM Health Homes*.; 2017.

4. Maragakis A, RachBeisel J. Integrated Care and the Behavioral Health Home: A New Program to Help Improve Somatic Health Outcomes for Those With Serious Mental Illness. *J Nerv Ment Dis*. 2015;203(12):891-895. doi:10.1097/NMD.0000000000000392

5. Highland J, Nikolajski C, Kogan J, Ji Y, Kukla M, Schuster J. Impact of Behavioral Health Homes on Cost and Utilization Outcomes. *Psychiatr Serv Wash DC*. Published online April 28, 2020:appips201900141. doi:10.1176/appi.ps.201900141

6. Bandara SN, Kennedy-Hendricks A, Stuart EA, et al. The effects of the Maryland Medicaid Health Home Waiver on Emergency Department and inpatient utilization among individuals with serious mental illness. *Gen Hosp Psychiatry*. 2020;64:99-104. doi:10.1016/j.genhosppsych.2019.12.004

7. Tepper MC, Cohen AM, Progovac AM, et al. Mind the Gap: Developing an Integrated Behavioral Health Home to Address Health Disparities in Serious Mental Illness. *Psychiatr Serv Wash DC*. 2017;68(12):1217-1224. doi:10.1176/appi.ps.201700063

8. Breslau J, Leckman-Westin E, Han B, et al. Impact of a mental health based primary care program on emergency department visits and inpatient stays. *Gen Hosp Psychiatry*. 2018;52:8-13. doi:10.1016/j.genhosppsych.2018.02.008

9. Romaire M, Alterbaum R, Collins A. Medicaid Behavioral Health Homes: Lessons Learned and Early Findings From Maine. *Psychiatr Serv Wash DC*. 2020;71(11):1179-1187. doi:10.1176/appi.ps.201900490

10. Spillman BC, Allen EH. *Evaluation of the Medicaid Health Home Option for Beneficiaries with Chronic Conditions: Evaluation of Outcomes of Selected Health Home Program, Annual Report - Year Five*. Urban Institute; 2017.

11. Fretheim A, Tomic O. Statistical process control and interrupted time series: a golden opportunity for impact evaluation in quality improvement. *BMJ Qual Saf*. 2015;24(12):748-752. doi:10.1136/bmjqs-2014-003756

12. Campbell DT, Ross HL. The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis. *Law Soc Rev*. 1968;3(1):33-53. doi:10.2307/3052794

13. Center DB, Leach RA. The multiple baseline across subjects design: proposed use in research. *J Manipulative Physiol Ther*. 1984;7(4):231-236.

14. Linden A. Use of the Pre-Post Method to Measure Cost Savings in Disease Management. *Dis Manag Health Outcomes*. 2007;15(1):13-18. doi:10.2165/00115677-200715010-00003

15. Kratochwill TR, Levin JR. What time-series designs may have to offer educational researchers. *Contemp Educ Psychol*. 1978;3(4):273-329. doi:10.1016/0361-476X(78)90034-6

16. LaPointe L. Multiple Baseline Designs. *Clin Aphasiology*. Published online 1978. http://aphasiology.pitt.edu/410/1/08-03.pdf

17. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton, Mifflin and Company; 2002:xxi, 623.

18. Hawkins NG, Sanson-Fisher RW, Shakeshaft A, D'Este C, Green LW. The multiple baseline design for evaluating population-based research. *Am J Prev Med*. 2007;33(2):162-168. doi:10.1016/j.amepre.2007.03.020

19. Coly A, Parry G. *Evaluating Complex Health Interventions: A Guide to Rigorous Research Designs*. AcademyHealth; 2017.

20. Basu S, Meghani A, Siddiqi A. Evaluating the Health Impact of Large-Scale Public Policy Changes: Classical and Novel Approaches. *Annu Rev Public Health*. 2017;38:351-370. doi:10.1146/annurev-publhealth-031816-044208

21. Baer DM, Wolf MM, Risley TR. Some current dimensions of applied behavior analysis. *J Appl Behav Anal*. 1968;1(1):91-97. doi:10.1901/jaba.1968.1-91

22. Tanious R, Onghena P. Randomized Single-Case Experimental Designs in Healthcare Research: What, Why, and How? *Healthcare*. 2019;7(4). doi:10.3390/healthcare7040143

23. Ferron JM, Farmer JL, Owens CM. Estimating individual treatment effects from multiple-baseline data: a Monte Carlo study of multilevel-modeling approaches. *Behav Res Methods*. 2010;42(4):930-943. doi:10.3758/BRM.42.4.930

24. Miller CJ, Smith SN, Pugatch M. Experimental and quasi-experimental designs in implementation research. *Psychiatry Res*. 2020;283:112452. doi:10.1016/j.psychres.2019.06.027

25. Dallery J, Raiff BR. Optimizing behavioral health interventions with single-case designs: from development to dissemination. *Transl Behav Med*. 2014;4(3):290-303. doi:10.1007/s13142-014-0258-z

26. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46(1):348-355. doi:10.1093/ije/dyw098

27. Gillings D, Makuc D, Siegel E. Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care. *Am J Public Health*. 1981;71(1):38-46. doi:10.2105/ajph.71.1.38

28. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. *Acad Pediatr*. 2013;13(6 Suppl):S38-44. doi:10.1016/j.acap.2013.08.002

29. Simonton DK. Cross-sectional time-series experiments: Some suggested statistical analyses. *Psychol Bull*. 1977;84(3):489-502. doi:10.1037/0033-2909.84.3.489

30. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther*. 2002;27(4):299-309.

31. Ferron JM, Bell BA, Hess MR, Rendina-Gobioff G, Hibbard ST. Making treatment effect inferences from multiple-baseline data: the utility of multilevel modeling approaches. *Behav Res Methods*. 2009;41(2):372-384. doi:10.3758/BRM.41.2.372

32. U.S. Bureau of Labor Statistics. Consumer Price Index Historical Tables for U.S. City Average. Accessed February 19, 2021. https://www.bls.gov/regions/mid-atlantic/data/consumerpriceindexhistorical_us_table.htm

33. Huitema BE, Mckean JW, Mcknight S. Autocorrelation Effects on Least-Squares Intervention Analysis of Short Time Series. *Educ Psychol Meas*. 1999;59(5):767-786. doi:10.1177/00131649921970134

34. St.Clair T, Cook TD, Hallberg K. Examining the Internal Validity and Statistical Precision of the Comparative Interrupted Time Series Design by Comparison With a Randomized Experiment. *Am J Eval*. 2014;35(3):311-327. doi:10.1177/1098214014527337

35. Kim H. Propensity Score Analysis in Non-Randomized Experimental Designs: An Overview and a Tutorial Using R Software. *New Dir Child Adolesc Dev*. 2019;2019(167):65-89. doi:10.1002/cad.20309

36. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Propensity Score Methods for Bias Reduction in Observational Studies of Treatment Effect. *Rheum Dis Clin North Am*. 2018;44(2):203-213. doi:10.1016/j.rdc.2018.01.002

37. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. doi:10.1093/biomet/70.1.41

38. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. *Eur J Cardio-Thorac Surg Off J Eur Assoc Cardio-Thorac Surg*. 2018;53(6):1112-1117. doi:10.1093/ejcts/ezy167

39. Morgan CJ. Reducing bias using propensity score matching. *J Nucl Cardiol Off Publ Am Soc Nucl Cardiol*. 2018;25(2):404-406. doi:10.1007/s12350-017-1012-y

40. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2017. https://www.R-project.org

41. Manitoba Centre for Health Policy. Concept: Elixhauser Comorbidity Index. Accessed June 28, 2016. http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1436

42. Melek SP, Norris DT, Paulus J. *Economic Impact of Integrated Medical-Behavioral Healthcare: Implications for Psychiatry*. Milliman, Inc.; 2014.

43. Pande RL, Morris M, Peters A, Spettell CM, Feifer R, Gillis W. Leveraging remote behavioral health interventions to improve medical outcomes and reduce costs. *Am J Manag Care*. 2015;21(2):e141-151.

44. Murphy KA, Daumit GL, Stone E, McGinty EE. Physical health outcomes and implementation of behavioural health homes: a comprehensive review. *Int Rev Psychiatry Abingdon Engl*. 2018;30(6):224-241. doi:10.1080/09540261.2018.1555153

45. Momany ET, Damiano PC, Shane DM, Nguyen-Hoang P, Bentler S, Wachsmuth J. *Evaluation of Iowa Medicaid's Integrated Health Home Program*. University of Iowa; 2017:35.

46. *Report to Congress on the Medicaid Health Home State Plan Option*. The U.S. Department of Health and Human Services; 2018.

47. Kazdin AE, Kopel SA. On resolving ambiguities of the multiple-baseline design: Problems and recommendations. *Behav Ther*. 1975;6(5):601-608. doi:10.1016/S0005-7894(75)80181-X