

Abstract

Hypothesis

Method

Results

Discussion

Conclusion

# The Predictiveness of Mental Health Diagnostic Groupings on High Cost Claim Frequency

Code ▼

*Northern Arizona Regional Behavioral Health Alliance  
Flagstaff, AZ*

John Ryan Kivela, MA

2023-02-03

## Abstract

This study is an analysis of adjudicated claims for Alliance ACO Providers in the Northern Region of Arizona. The purpose of this investigation is to identify areas of focus for program development and clinical advancement by identifying predictive variables from the most recent contract year, 2021-2022.

A predictive analysis of high cost claims and ICD10 mental health groupings was conducted showing that *Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders* (F20-F29) and *Mood [affective] disorders* (F30-F39) are strong predictors of high cost claims.

## Hypothesis

Recent feedback from Alliance ACO Providers, and the recent results of the Alliance ACO Progress Report, support a closer analysis of the factors impacting inpatient admissions for mental health conditions. While the Alliance as a whole demonstrated an above average performance for the 6 month baseline period, some individual providers failed to meet the NCQA Mean for Follow-Up After Hospitalization for Mental Illness (FUH)<sup>1</sup>.

This study hypothesizes that different diagnostic groupings of mental health disorders may have a greater impact on high cost claims. These insights will help the Alliance develop data-driven interventions for groups at greater risk of high cost claims.

# Method

## Data Management, transparency, and integrity

The Alliance collects referential data through direct query of curated BCBSAZ hosted servers and databases. As such, the data has, at a minimum, the level of security and validation offered by BCBSAZ services.

Data is stored using Microsoft Office 365 cloud storage. Office 365 uses several encryption protocols, including Transport Layer Security/Secure Sockets Layer (TLS/SSL), Internet Protocol Security (IPSec), and Advanced Encryption Standard (AES). Office 365 data encryption applies to data at rest (data stored in One Drive for Business or SharePoint libraries) and data in transit (mail messages, notes from online meetings, or whenever a device is communicating within or between servers<sup>2</sup>).

### *Description of the data sample*

This sample includes adjudicated claims from BCBSAZ for the period of October 1, 2021 to September 30, 2022. The population includes individuals with an adjudicated claim for a HEDIS My2022 qualifying inpatient code associated with a HEDIS My2022 qualifying mental health diagnosis<sup>3</sup>. The sample is filtered to only include claims for BCBSAZ AHCCCS Complete Care members with residential addresses in the Northern Geographical Service Area. The sample is further differentiated by including only contracted Alliance Providers.

The sample size was 22675.

## Preparation of Data for Analysis

The data was extracted from the BCBSAZ data warehouse on 12/31/22 using SQL Management Studio. The SQL Code is available for review on request. After extraction, the data was reviewed for errors, such as duplications or NULL values. The original data is stored in encrypted folders. Prior to analysis the data was de-identified by converting date of birth to age, converting date of service to a month/year format, and creating an alternative ID in lieu of patient name or BCBSAZ ID.

## Preparing the data in R

Data files were loaded to the model:

- De-identified claims history from the Contract Year 2021 - 2022
- Reference tables for:
  - ICD10 codes
  - HEDIS My 2022 qualifying inpatient codes<sup>4</sup>
  - ICD 10 code groupings for mental health which will be used in a regression analysis<sup>5</sup>.

Reference objects were created and some data was renamed to assist in data frame construction and analysis.

A date frame was created from the claims history, which was joined with ICD10 mental health grouping identifiers and inpatient service codes based on primary diagnosis.

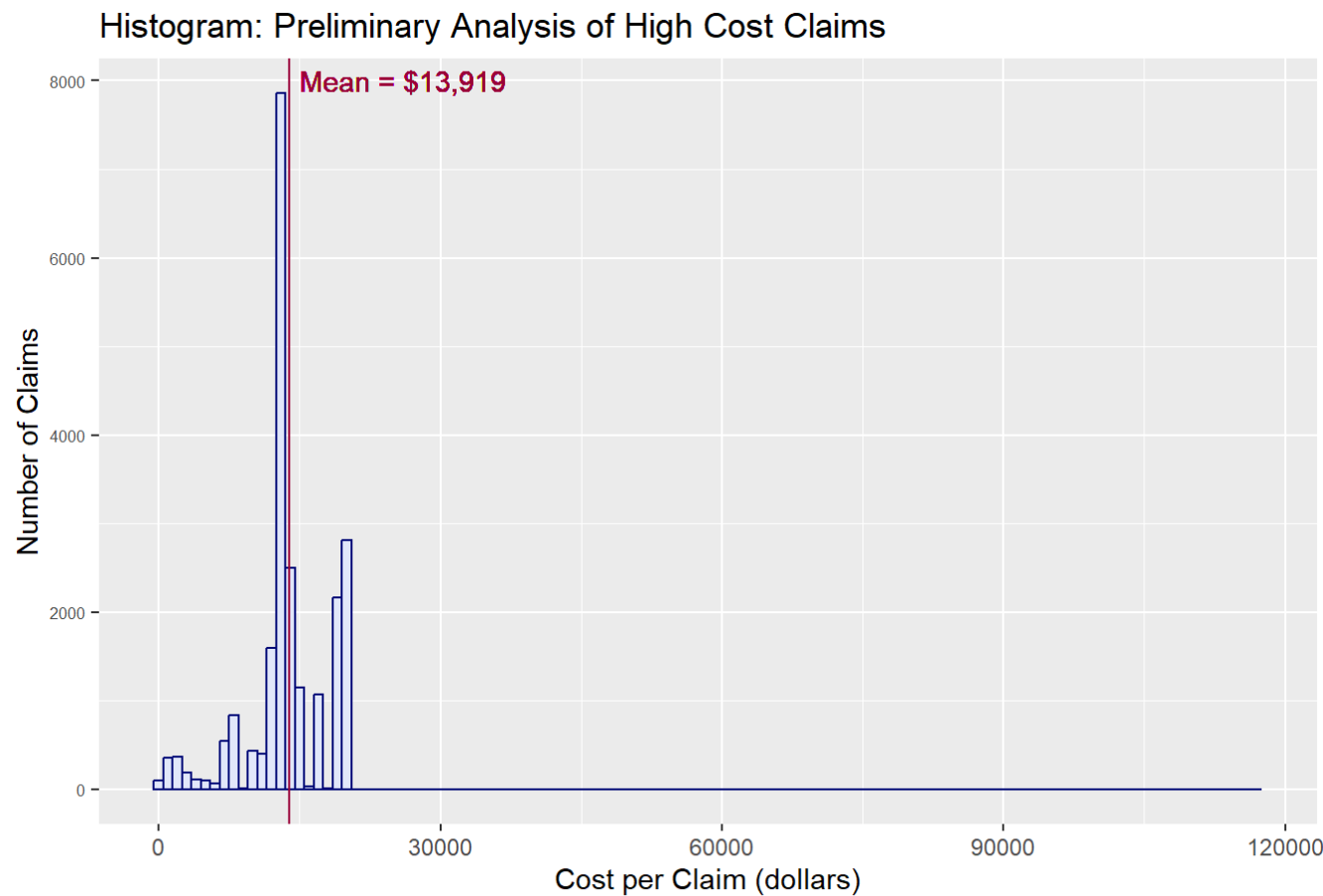
# Results

## Assessment of Cost Distribution for All Providers and MH Groups

A series of analyses were conducted to observe the overall distribution of high cost claims according to Alliance Provider and mental health diagnostic grouping.

### *Preliminary analysis*

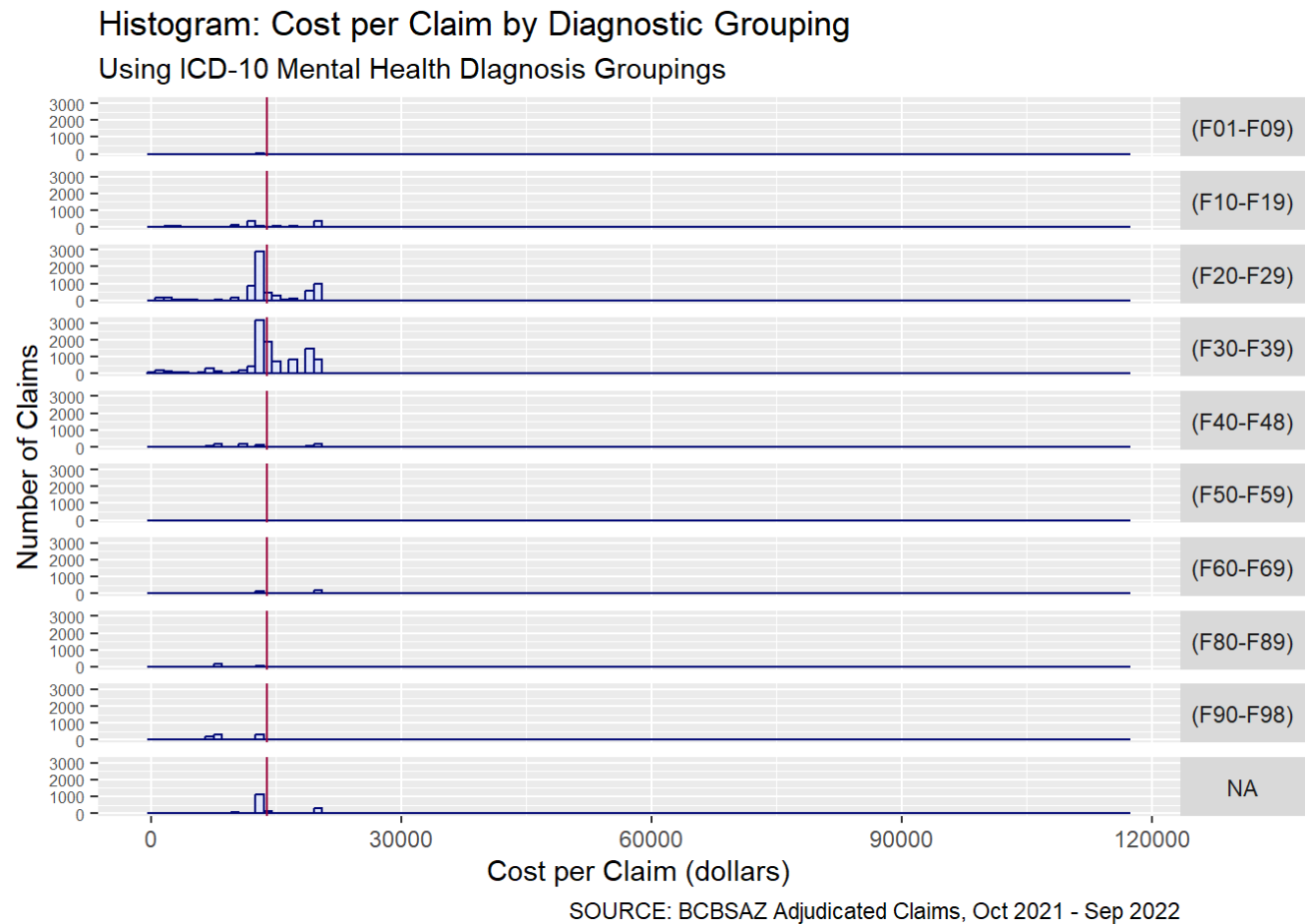
First, a basic histogram for cost of claim was plotted to observe the distribution of data. This provided a very clear visual indicator that the distribution is negatively skewed to the right by a small number of high cost outliers. Bin width for the histogram was set to \$1,000.



SOURCE: BCBSAZ Adjudicated Claims, Oct 2021 - Sep 2022

The distribution of cost was further investigated by expanding the histogram to create facets for each mental health grouping of ICD10 codes<sup>6</sup>. These results further demonstrated the data being skewed to the right by very few cases.

In addition, it was noted that the distribution appears to be concentrated in the psychotic disorders, and mood disorders groups, (F20-F29) and (F30-F39).



Descriptive statistics were calculated to further observe the distribution and assist in decision making about suppressing outlying high cost claims.

### Descriptive Statistics: Preliminary Analysis

Adjudicated Claim Cost Distribution (October 2021 - September 2022)

n	mean	sd	min	max	skew	kurtosis
22,675	\$13,919	\$4,459	\$0	\$117,040	-0.070	13.780

SOURCE: BCBSAZ Adjudicated Claims, extracted 12/31/22

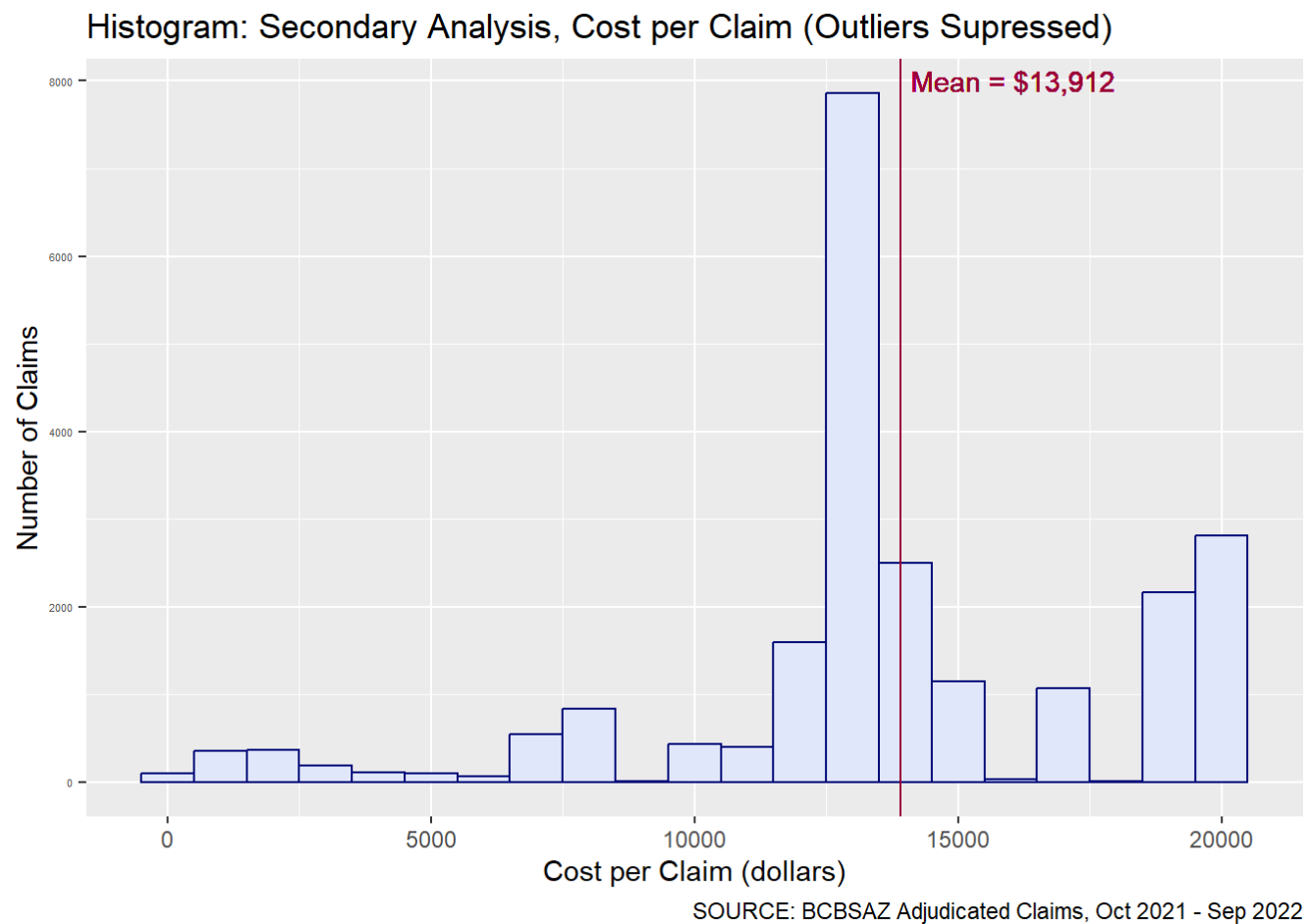
The results of these statistics were: Mean = \$13,919.45, and Standard Deviation = \$4,458.63. The statistics confirmed our visual observation that there is a negative skew in the data (Skewness = -0.07); however the shape of the distribution is also impacted by a high kurtosis (Kurtosis = 13.78), or a *heavy upper tail*, due to extreme outliers.

It was therefore determined that outliers beyond 3 standard deviations of the mean in either direction will be removed. The outlier cutoff, defined as the Mean added to three times the Standard Deviation (Cutoff Limit =  $M + (3 * SD)$ ), gives us a cutoff limit of +/- \$27,295.34.

There were a total of 3 outliers removed. The final sample size is 22672 claims.

## Secondary Analysis

The histograms were re-constructed utilizing the new data set with outliers suppressed. Bin width was again set to \$1000. The new chart for the entire sample reveals a more symmetrical distribution with a slight skew to the left (Skewness = -0.68).



Updated descriptive statistics were calculated, revealing a new Mean = \$13,911.91, and Standard Deviation = \$4,394.47. The mean was not greatly impacted by the removal of the outlying values. The kurtosis of the distribution was relieved by the removal of the outliers (Kurtosis = 1.09).

Altogether, the characteristics of the distribution indicate that this data is acceptable to use for predictive analyses.

## Descriptive Statistics: Secondary Analysis

Outliers Removed: Adjudicated Claim Cost Distribution

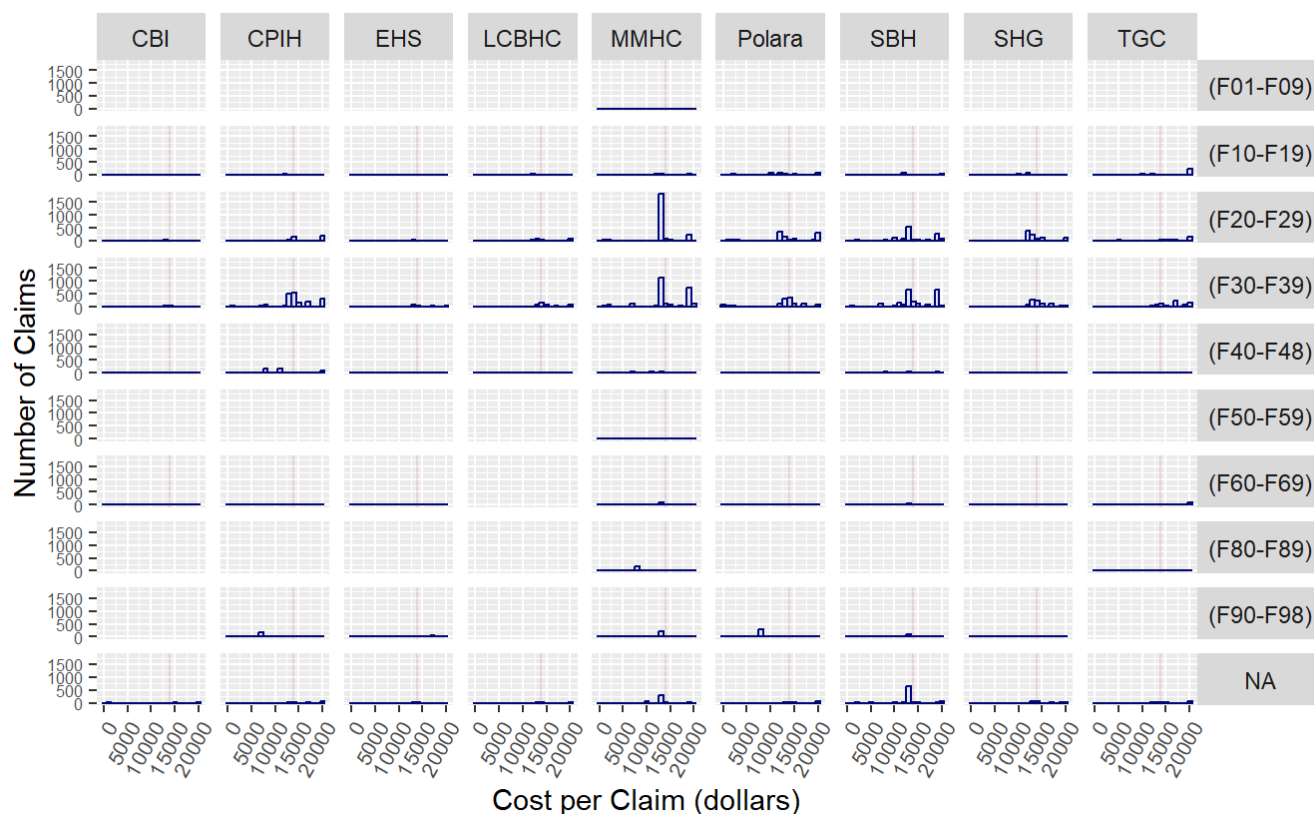
n	mean	sd	min	max	skew	kurtosis
22,672	\$13,912	\$4,394	\$0	\$20,465	-0.680	1.090

*SOURCE: BCBSAZ Adjudicated Claims, extracted 12/31/22*

An expanded histogram was constructed investigating facets of the total cost of services including mental health ICD10 code groupings, and responsible Alliance Provider. These results again appear to show a concentration around (F20-F29) and (F30-F39). While the plot below is somewhat complex, the expanded histogram provides additional information concerning the frequency of high cost claims for Alliance Providers.

### Histogram: Cost per Claim by Diagnostic Grouping (Outliers Suppressed)

Using ICD-10 Mental Health Diagnosis Groupings



Data Source

## Regression Analysis

A multiple linear regression was conducted to test if ICD10 diagnostic groupings for mental health significantly predicted the occurrence of high cost claims. Preliminary analysis of the data reveals a distribution that appears to be more heavily concentrated in the psychotic disorders and mood disorders,

(F20-F29) and (F30-F39) groupings.

The purpose of this test is to confirm if these groups are significant predictors of high cost claims.

## Preliminary regression analysis

The results of the multiple linear regression reveal a significant effect of diagnosis group on the occurrence of high cost claims overall ( $F(8,3)=16.39$ ,  $p=0.02$ ), with ( $R^2 = 0.918$ ).

This suggests that about **92% of the variation** is predicted by diagnostic grouping. (F20-F29) and (F30-F39), psychotic disorders, and mood disorders, were the highest predictors, with significance levels of  $p=0.043$  and  $p=0.013$ , respectively.

### Linear Regression: High Cost Claims Predicted by MH Grouping

*Initial regression includes all groupings*

Characteristic	Beta	95% CI <sup>1</sup>	p-value
(Intercept)	1,192,409	-11,242,775, 13,627,593	0.8
(F10-F19)	1.9	-0.60, 4.3	0.095
(F20-F29)	0.98	0.05, 1.9	0.044
(F30-F39)	1.0	0.41, 1.7	0.013
(F40-F48)	-0.09	-3.2, 3.0	>0.9
(F60-F69)	0.09	-7.0, 7.2	>0.9
(F90-F98)	2.1	-1.1, 5.2	0.13
(F80-F89)	0.92	-9.6, 11	0.8
(F01-F09)	1.4	-30, 32	0.9

<sup>1</sup> CI = Confidence Interval

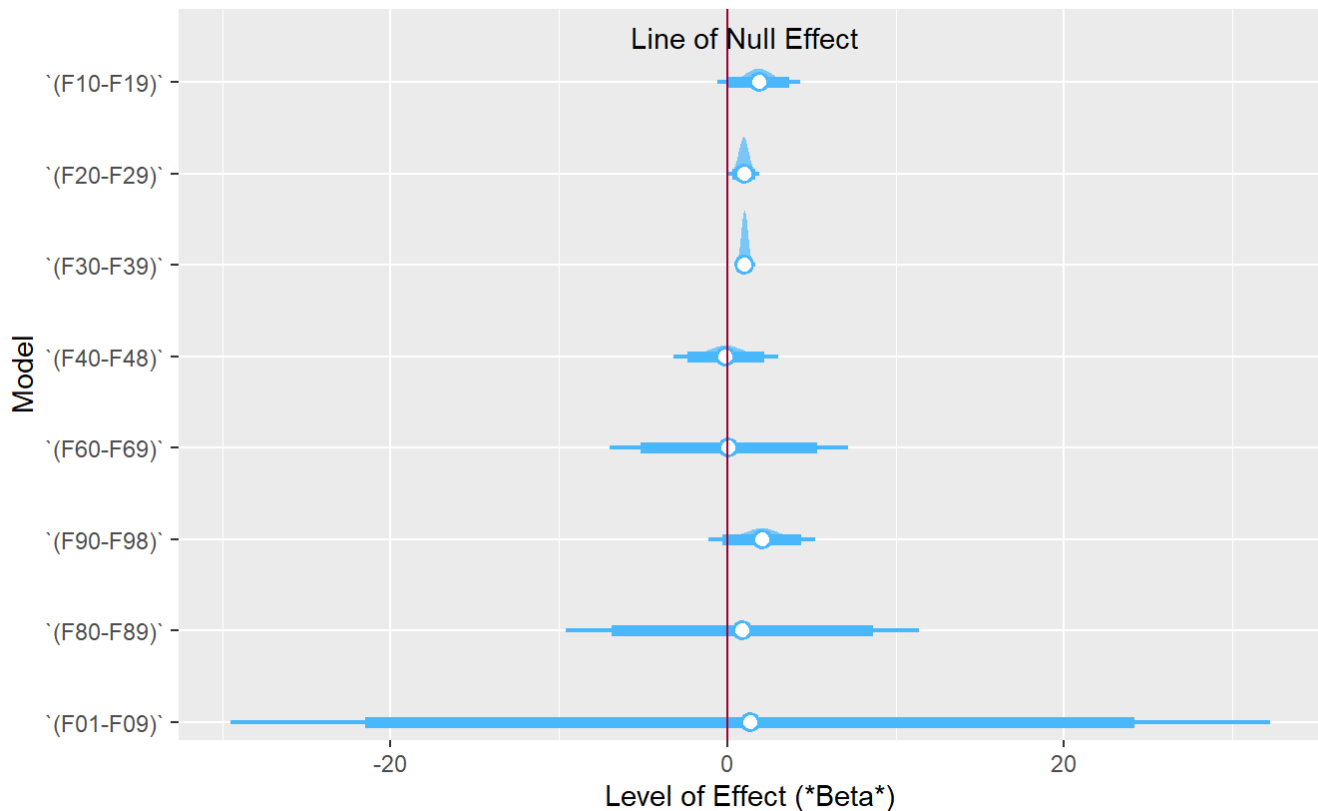
*SOURCE: BCBSAZ Adjudicated Claims, extracted 12/31/22*

The forest plot below provides a visual way to observe the significance of predictive variables. For each grouping the horizontal line represents the confidence interval. This demonstrates a range of values that we can be 95% certain contain the predicted true value. The vertical line represents the value at which there is no significant effect.

If the confidence interval line crosses the line of null effect, then the result is not significant. (F20-29) and (F30-F39) are seen here as significant predictors of high cost claims.

## Effect level of Mental Health Groupings on High Cost Claims

\*Using ICD-10 Mental Health Diagnosis Groupings\*



SOURCE: BCBSAZ Adjudicated Claims, extracted 12/31/22

## Predictive grouping regression analysis

The multiple regression equation was updated to focus on the significant predictive groupings (F20-F29) and (F30-F39). Overall, the updated regression revealed a stronger statistically significant relationship ( $F(2,9)=49.55$ ,  $p<0.001$ ), with ( $R^2 = 0.898$ ) when compared to the original regression.

This highly significant result suggests that about **90% of the variation** in high cost claim occurrence is predicted by **these two groupings alone!**.

The value " $\beta$ " represent the slope of the regression line through the data. The level of significance was set to " $p<.05$ ", or a 95% Confidence Interval.

It was observed that (F20-F29) was a significant predictor of high cost claims ( $\beta=1.2$ ,  $p<.001$ ).

It was observed that (F30-F39) was a significant predictor of high cost claims ( $\beta=1.0$ ,  $p<.001$ ).



## Linear Regression: High Cost Claims Predicted by MH Grouping

*Non-predictive groupings suppressed*

Characteristic	Beta	95% CI <sup>1</sup>	p-value
(Intercept)	3,867,130	-1,252,778, 8,987,039	0.12
(F20-F29)	1.2	0.82, 1.6	<0.001
(F30-F39)	1.0	0.72, 1.3	<0.001

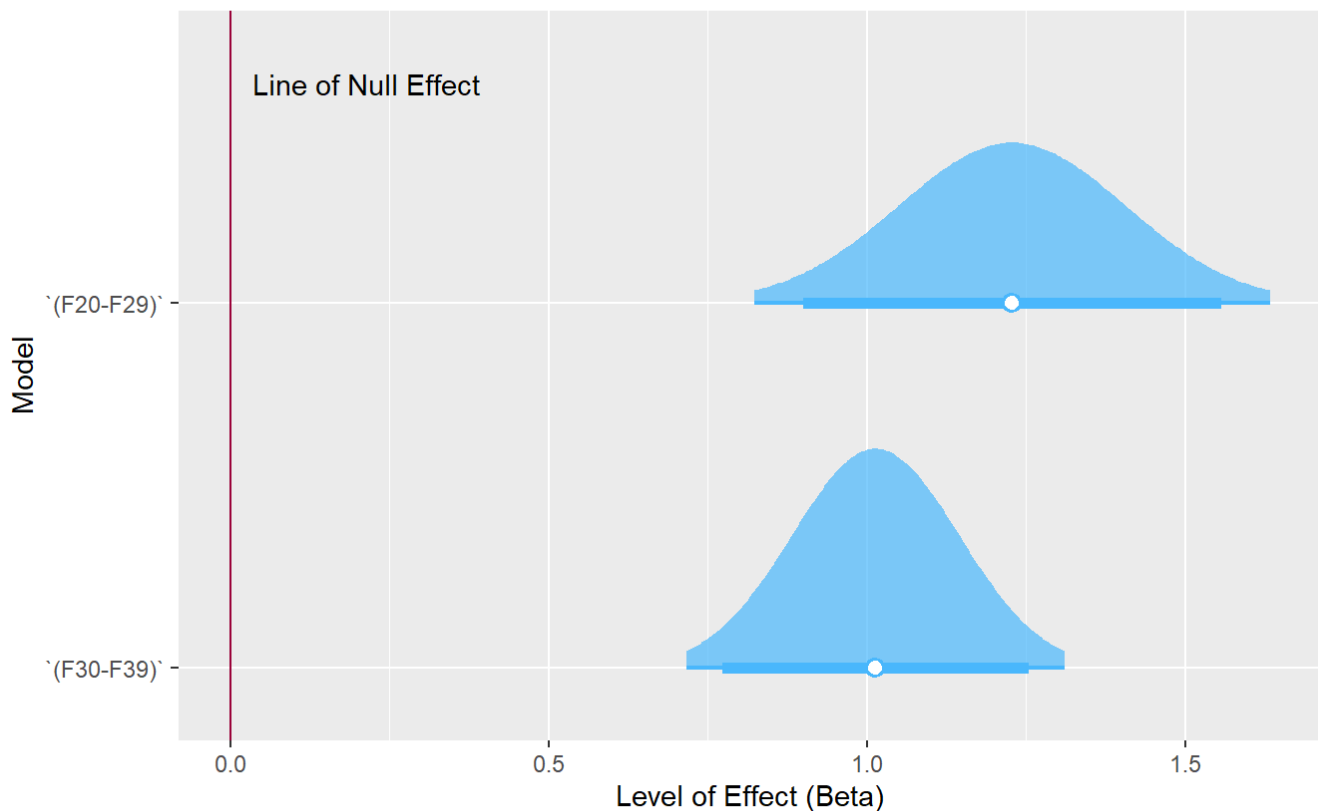
<sup>1</sup> CI = Confidence Interval

*SOURCE: BCBSAZ Adjudicated Claims, extracted 12/31/22*

By suppressing the non-predictive groupings, the significance of the predictive value of (F20-F29) and (F30-F39) is more visible. The forest plot below demonstrates the distribution of values contained within a 95% confidence interval for the 2 respective groupings. The distributions are closely contained within the confidence interval, indicating that each distribution is statistically significant.

### Effect level of Predictive MH Groupings on High Cost Claims)

Using ICD-10 Mental Health Diagnosis Groupings



SOURCE: BCBSAZ Adjudicated Claims, extracted 12/31/22

# Discussion

## Regression Analysis

The results of the analysis **confirmed that the mental health diagnostic group of the patient is a strong predictor** for the occurrence of high cost claims for inpatient residential services. In particular, Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders, and Mood [affective] disorders, (F20-F29) and (F30-F39) respectively, were the highest predictors of all groupings, accounting for **90% of the variance** across over 22,000 adjudicated claims for a period of 1 year.

## Conclusion

### Why is this important?

Significant predictors of behaviors, services, and claim costs can be hidden among the immense volume of data and clinical work that is required to provide quality behavioral health services. It is common knowledge in the field that some disorders tend to occur more frequently than others in inpatient care. This research allows us to put a name, and a significance level to the **top predictors**.

These results also helps us understand the power of psychotic disorders, and mood disorders to predict higher cost. These 2 diagnostic groups, out of 9, are responsible for **at least 90%** of the variance in our sample.

These results provide evidence based decision making power to address the clinical and administrative needs of members living with psychotic and mood disorders. It also provides the ability to estimate performance targets for reduction in inpatient service costs.

## Suggestions for Action

Further investigation into the inpatient experience of members with predictive diagnosis is indicated. Areas of emphasis may include:

- The impact of diagnostic groupings on length of stay
- The impact of diagnostic groupings on re-admission
- Frequency and duration of inpatient stays by facility

Additional training could be provided to enable the work force to provide care to individuals with disorders that are highly predictive of higher cost claims. Collaboration with health medical management resources could also help to ensure clear and effective communication between the health plan and Alliance Providers.

- 
1. Kivela J.R., McMillian, J. & Tewa, V. (2023) Alliance ACO Progress Report, January 2023. *The Northern Arizona Regional Behavioral Health Authority*.↵
  2. Encryption in Microsoft 365 - Microsoft Purview (Compliance), 2022↵
  3. <https://www.ncqa.org/hedis/> (<https://www.ncqa.org/hedis/>)↵

4. <https://www.ncqa.org/hedis/> (<https://www.ncqa.org/hedis/>)↵
5. ICD-10-CM official guidelines for coding and reporting. (n.d.). Retrieved January 18, 2023, from <https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf> (<https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf>)↵
6. ICD-10-CM official guidelines for coding and reporting. (n.d.). Retrieved January 18, 2023, from <https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf> (<https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf>)↵