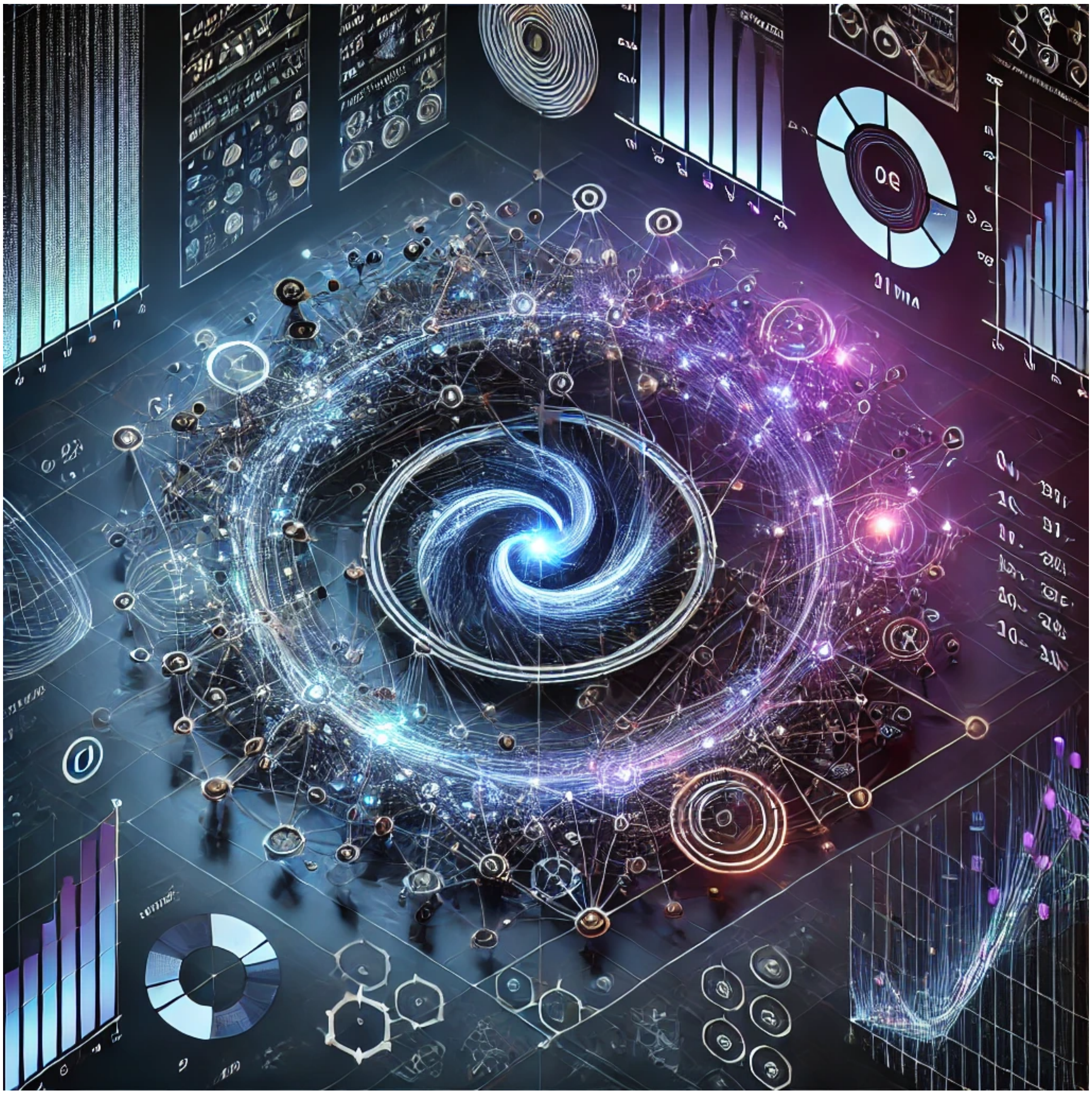
HSR.*health*

An Investigation into Recursive Regression in Health Risk Analytics



By Saiajay Chigurupati and Pau A Churchyard, HSR.*health*

Impact@HSR.*health* | 240-731-0756 | www.HSR.*health*

HSR.*health*+

An Investigation into Recursive Regression in Health Risk Analytics

Table of Contents

Table of Figures

Table of Tables

An Investigation into Recursive Regression in Health Risk Analytics

Executive Summary

To investigate the potential of HSR.*health*'s recursive regression algorithm in analyzing state and zip-code level data for mental health outcomes, I compared the results of the algorithm to established literature trends in income, sex, and employment status. The algorithm, which also involves z-score standardization, Spearman correlations, and variance inflation factor (VIF) analysis, led to mixed results. While most of the regression aligned with the general trends on sex and income vs. mental health, Utah's regression was an outlier in terms of income and Iowa's regression was an outlier in terms of sex. For employment, many of the regressions disagreed with the literature trends. These mixed results provide promise for further evaluation and expansion, like in introducing more complexity, while providing room for improvement, potentially within the VIF analysis to eventually lead to localized mental health outcomes analysis with the recursive regression algorithm.

Introduction

While awareness has grown for it, mental health continues to be a major problem across the country. In 2022, more than 1 in 5 US adults [59.3 million] were facing mental health problems, with only 50.3% [30 million] of those adults receiving mental health treatment in response (National Institute of Mental Health, 2024). To address the mental health crisis in our country, it is important to understand the social and demographic factors that drive these outcomes. While research has been done on how factors like employment, gender, income, and race affect mental health outcomes, it may prove more informative to work past the national level into the state and zip-code levels.

To investigate the potential of more local state-level/zip-code level analysis, HSR.health has developed a multi-step recursive regression analysis to determine which social predictors are the most significant in each state and calculate risk adjustment scores for each zip-code based on these regressions. In this report, I will outline the steps of the method and investigate the regressions that result from this algorithm, specifically on how they compare to literature trends on income, gender, and employment.

Methodology

Through the GeoMD Platform, ZIP Code level data was accessed on various demographics and social factors, including age, sex, income, gender, employment. The response variable to represent mental health at each zip code was the percent of each ZIP Code that faced poor mental health - “Percent Mental Health Not Good”.

The data was first standardized into z-scores to prevent certain variables from impacting regression results more simply because of the size of their units. Z-score standardization involves subtracting the overall column mean value from each data entry and dividing by the overall column standard deviation.

After Z-score standardization, a Spearman correlation was conducted on the data. While a typical Pearson correlation measures the linear correlation of a predictor variable to the response variable without further modification of the data, a Spearman correlation instead measures the linear correlation of the rankings of the predictor variable and response variable. The Spearman correlation has the advantage of not being non-parametric, which avoids having to meet certain assumptions the Pearson correlation requires.

Once Spearman correlations are calculated for every predictor variable with the response variable, the next step is to implement Variance Inflation Factor (VIF) analysis and filter based on those factors. VIF helps determine the presence of multicollinearity, which occurs when multiple predictor variables are highly correlated with each other. With multicollinearity, a major issue, especially for the next steps, is that the statistical significance of each predictor variable implicated in multicollinearity is lowered and obscured. A prime example of multicollinearity is the dummy variable trap, which occurs when a categorical variable is coded for with too many variables. In the case of our sex predictor variables, which are limited to male and female, certain multicollinearity would occur if a regression included both a percent male and percent female variable. Other variables may have less obvious multicollinearity. Thus, using a recursive VIF filtering method, we drop variables that most strongly contribute to multicollinearity. For variables implicated together in multicollinearity, the variable with the highest Spearman correlation coefficient is dropped.

After VIF analysis is completed to remove variables implicated in collinearity, the next step is the recursive regression analysis. Here, the regression is run multiple times, starting at a p-value cutoff of 0.5. At the first cutoff, an ordinary least squares regression is run with the remaining variables from the previous step. Here, only predictors with p-values under the cutoff of 0.5 are kept. The regression analysis is then run again, lowering the cutoff value by 0.1 with each round until the cutoff reaches 0.1 or no more new variables are kept.

While this report will focus on the results of the recursive regression, the method continues on to determine risk adjustment factors (RAFs) for each ZIP Code. The corresponding state regression recalculates the response variable, Percent Mental Health Not Good, for each ZIP Code. These scores, which are in terms of Z-score standardized variables, are reverted to generate a single value, the RAF, for each county. In theory, this value represents how much the variables identified as significant by the recursive regression specifically contribute to the proportion of those with poor mental health.

Discussion

After the recursive regression was run, regressions were generated for 48 out of 50 states. The exceptions were Florida and Delaware, which did not have data available. Three other states generated questionable coefficients that were orders of magnitudes larger than the other coefficients, which were primarily between 0 and 1 due to the standardized variables. These states were Minnesota, Montana, and New Hampshire:

Table 1: States with the Maximum Coefficient Value.

State	Maximum Coefficient Value
Minnesota (MN)	3447.37
Montana (MT)	13591.25
New Hampshire (NH)	26506.7

For now, these states will be excluded, although further analysis will be required to determine why these states generated outlier regressions. One potential explanation could be that the VIF step failed to exclude highly covariant predictor variables for these states, which then led to outlier coefficients in the following regression step.

For the remaining forty-five states, a distribution of the adjusted R-squares is plotted below. A majority of regressions fall within the 0.4-0.5 and 0.5-0.6 range, with a few regressions on either side of that outperforming or underperforming the majority.

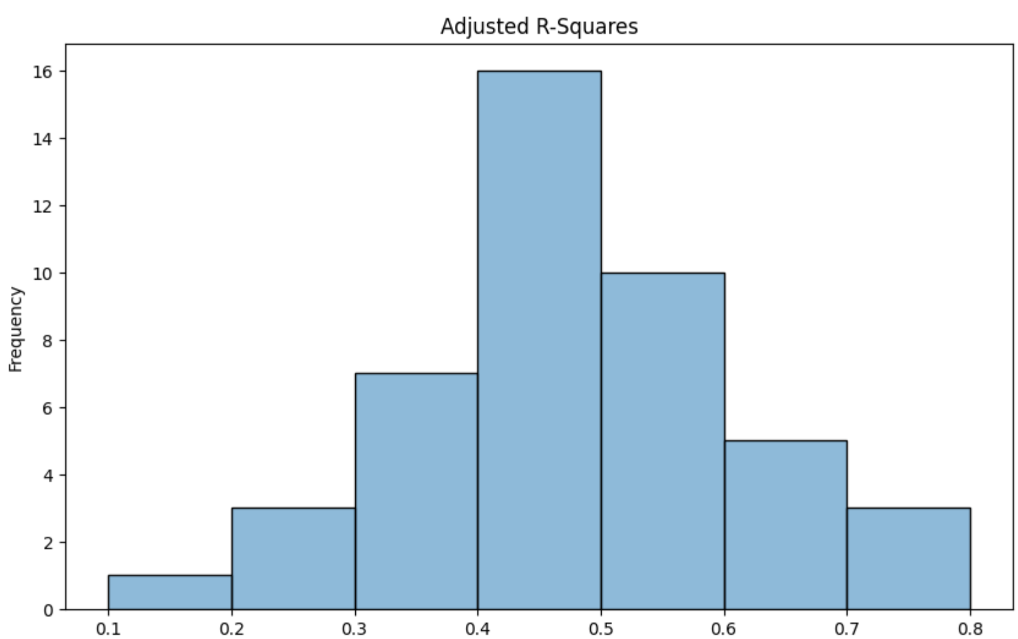


Figure 1: Distribution of adjusted R-squares values.

In the table below, the top and bottom six R-squares are displayed:

Table 2: States with the top and bottom adjusted R-square values.

State	Adjusted R-Square Value
Top Six	
Connecticut (CT)	0.78
South Carolina (SC)	0.74
New Jersey (NJ)	0.72
Arizona (AZ)	0.68
North Dakota (ND)	0.64
Tennessee (TN)	0.64
Bottom Six	
West Virginia (WV)	0.18
Iowa (IA)	0.20
Pennsylvania (PA)	0.28
Kentucky (KY)	0.30
Louisiana (LA)	0.32
California (CA)	0.37

If the adjusted R-square values are the only metric to judge the regression's ability to account for the variability in poor mental health within a ZIP code, these regressions show mixed results. While states like Connecticut, South Carolina, and New Jersey have strong adjusted R-square values, they are countered by states like West Virginia, Iowa, and Pennsylvania, which have poor adjusted R-square values. Beyond these standouts, a majority fall within R-square ranges of 0.40 to 0.60.

To further analyze these results, the frequency of each predictor variable determined as significant across the forty-five regressions was tallied. The results of that tally are detailed below in the following table:

Table 3: Number of State Regressions with each Predictor Variable.

Predictor Variable Name	# of State Regressions with Predictor
Percent Employed	31
Percent Unemployed	30
Percent Not In Labor Force	26
Percent Have No Health Insurance	25
Percent Mortgage greater than 30 Percent of Income	23
Percent Between the ages of 10 and 19	22
Percent Military	22
Percent Have Private Health Insurance	22
Percent Public Administration	22
Percent Asian	21
Percent Disabled	21
Percent Retail Trade	21
Percent Households With Food Stamps or SNAP Benefits	21
Percent Transportation and warehousing, and utilities	21
Percent Under the age of 5	21
Percent Between the ages of 20 and 29	20
Percent Between the ages of 30 and 39	20
Percent Not Fluent in English	20

Predictor Variable Name	# of State Regressions with Predictor
Percent Veteran	20
Percent Between the ages of 50 and 59	19
Percent Over the age of 80	19
Percent Arts, entertainment, and recreation, and accommodation and food services	19
Percent Educational services, and health care and social assistance	19
Percent Other services, except public administration	19
Percent Between the ages of 60 and 69	18
Percent Have Public Health Insurance	18
Percent Households With Retirement Income	18
Percent Households With Supplemental Security Income	18
Percent Between the ages of 40 and 49	17
Percent Income Between 50 and 74k	17
Percent Black	17
Percent Rent greater than 30 Percent of Income	17
Percent Single Parent Household	17
Percent Households With Social Security	17
Percent Income Between 75 and 99k	16

Predictor Variable Name	# of State Regressions with Predictor
Percent Finance and insurance, and real estate and rental and leasing	16
Percent Information	16
Percent Two or more Races	16
Percent Wholesale Trade	16
Percent Income Between 100 and 149k	15
Percent Households With Earnings	15
Percent Male	15
Percent Median Housing Value	15
Percent Income Between 150 and 199k	14
Percent Income Between 25 and 49k	14
Percent Lack of Completed Kitchen	14
Percent Manufacturing	14
Percent Income Greater than 200k	13
Percent Households With Cash Public Assistance Income	13
Percent Lack of Completed Plumbing	13
Percent No Home Heating Fuel	13
Percent Income Under 25k	13
Percent Construction	12
Percent Other Race	12

Predictor Variable Name	# of State Regressions with Predictor
Percent Professional, scientific, and management, and administrative and waste management services	11
Percent Between the ages of 70 and 79	10
Percent Housing Units built after 2020	10
Percent Median Rent Cost	10
Percent Normal Occupant Density	10
Percent Renter Occupied Housing	9
Percent Agriculture, forestry, fishing and hunting, and mining	8
Percent High Occupant Density	8
Percent Median Household Income	8
Percent Native Hawaiian or other Pacific Islander	8
Percent Female	7
Percent Have Health Insurance	7
Percent Under the age of 10	6
Percent Housing Units built before 1950	5
Percent White	4
Percent Hispanic	3
Percent Low Occupant Density	2
Percent Over the age of 65	1

The three most frequent variables all tie to employment/working, with percent employed/unemployed and percent not in the labor force. Other variable types that appear on this list include income ranges, sex, age ranges, and race. For this report, we will focus on income, gender, and employment status.

Income Brackets

In the literature on how income and mental health relate to one another, it has been generally found that higher income correlates with better mental health outcomes (Ettner, 1996; Li et al., 2022; Sareen et al., 2011; Thomson et al., 2022). Thus, for the regressions that include income brackets [ex. “Percent Income Under 25k”, “Percent Income Between 75 and 99k”, “Percent Income Greater Than 200k”, etc.], we would expect the coefficients to match this correlation. To investigate this, we first isolated the states that had income brackets in their predictors and found their coefficients:

Table 4: Percent income range coefficients by state.

State	Percent Income Ranges Coefficients (Ranges in Thousands of Dollars)						
	< 25	25-49	50-74	75-99	100-149	150-199	≥ 200
Alabama (AL)			-0.04		-0.10		-0.14
Arkansas (AR)				-0.21	-0.18		
Connecticut (CT)	0.31	-0.11	0.26		0.12	0.17	
Georgia (GA)				-0.09			-0.32
Idaho (ID)		-0.14	-0.15				
Louisiana (LA)		-0.07	0.15	0.16		0.11	-0.18
Massachusetts (MA)	0.20	0.15	0.19	0.15		0.13	
Maryland (MD)		-0.25	-0.13	-0.19	-0.25	-0.16	-0.51

State	Percent Income Ranges Coefficients (Ranges in Thousands of Dollars)						
	< 25	25-49	50-74	75-99	100-149	150-199	≥ 200
Maine (ME)			-0.10	-0.27	-0.14	-0.14	-0.21
Mississippi (MS)	0.22	0.15	0.13		0.11		
North Carolina (NC)	0.27	0.16	0.08				
North Dakota (ND)				-0.14	-0.20		-0.14
New Jersey (NJ)	0.16	0.29	0.26	0.24	0.17	0.08	
Ohio (OH)	0.35	0.30	0.19	0.19	0.25	0.08	0.10
Rhode Island (RI)							-0.46
South Carolina (SC)			-0.18	-0.06	-0.14	-0.18	
South Dakota (SD)						-0.10	
Tennessee (TN)	0.35	0.23	0.20	0.11	0.13	0.18	
Texas (TX)	0.09	0.08		0.04		-0.05	-0.12
Utah (UT)	-0.21						
Virginia (VA)	0.54	0.29	0.37	0.31	0.27	0.08	0.37
Vermont (VT)					0.19		
Wisconsin (WI)	0.28	0.16	0.15	0.14		0.19	

State	Percent Income Ranges Coefficients (Ranges in Thousands of Dollars)						
	< 25	25-49	50-74	75-99	100-149	150-199	≥ 200
West Virginia (WV)	0.16		0.08		-0.07		-0.15
Wyoming (WY)				-0.08			-0.18

Since the predictor variable is the percent of a zip code area with poor mental health, we expect coefficients to decrease in value as we move across the table. For example, Alabama's regression has the coefficients decrease from -0.04 to -0.10 to -0.14 as the income ranges increase. Thus, the regression for Alabama predicts a 0.04 unit decrease in poor mental health for a unit increase in those making between 50 and 74 thousand dollars, but a larger 0.14 unit decrease in poor mental health for a unit increase in those making more than 200 thousand dollars.

However, not every state shows this monotonous drop in coefficient values across the table. For example, Connecticut, while showing a generally decreasing trend, has its smallest coefficient at -0.11 for the 25 to 49 thousand range. Alternatively, Virginia also has its lowest coefficient at the 150 to 199 thousand range, with a jump in the coefficient for the greater than 200 thousand range.

There are multiple plausible explanations for these breaks in the trends. One explanation that could account for Connecticut is the general direction of income change from area to area. In the case of Connecticut, the general trend could be that the increase in those who make between 25 to 49 thousand could primarily come from upward movement from those making less than 25 thousand dollars. Another explanation, which could account for Virginia, is the theory of marginal returns in improvement as income improves. From the literature, studies have shown that the improvements in mental health begin to fall off as income increases more and more (Li et al., 2022). One study applied quadratic regressions to even suggest that those in the highest income brackets could have worse mental health outcomes than those in the bracket below (Li et al., 2022). A third plausible explanation is that regressions aren't showing the monotonous drop in coefficient values due to standard errors on the coefficients. For example, in Wisconsin, the 150 to 199 thousand range has a slightly higher coefficient than lower income brackets, which could easily be a result of overlapping standard error ranges.

Outside of these explanations, the prime outlier in these regressions is the Utah regression. For Utah, the regression surprisingly predicts that a one-unit increase in those who make under 25k, which can only occur with a corresponding decrease in those who make more, improves mental health by 0.21 units. This result opposes the generally agreed-upon trend in the public health literature of low income directly correlating to worse mental health outcomes.

Looking deeper at Utah's regression specifically, along with the -0.21 coefficient value for Percent Income Under 25k, there is also an interesting trend with the three health insurance variables. The no health insurance group has a smaller coefficient compared to the private and public health insurance groups; the regression implies that zip codes in Utah with more people not on insurance have better mental health. The insurance coefficients similarly go against trends connecting improved health insurance to better mental health outcomes (Hamersma & Ye, 2021; Kozloff, 2017; Lang, 2011).

Table 5: Coefficient for each predictor variable.

Predictor Variable	Coefficient
Percent Have No Health Insurance	0.13
Percent Have Private Health Insurance	0.26
Percent Have Public Health Insurance	0.23
Percent Between the ages of 10 and 19	0.13
Percent Between the ages of 20 and 29	0.32
Percent Between the ages of 30 and 39	0.16
Percent Between the ages of 40 and 49	0.19
Percent Between the ages of 50 and 59	0.13
Percent Between the ages of 60 and 69	0.26
Percent Over the age of 80	0.13
Percent Under the age of 10	0.14
Percent Housing Units built after 2020	-0.09
Percent Lack of Completed Plumbing	0.12
Percent No Home Heating Fuel	0.11
Percent Mortgage greater than 30 Percent of Income	-0.13

Predictor Variable	Coefficient
Percent Households With Food Stamps or SNAP Benefits	0.29
Percent Income Under 25k	-0.21
Percent Single Parent Household	0.11
Percent Asian	0.12
Percent Native Hawaiian or other Pacific Islander	-0.10

Literature review showed that Utah has had specific policy actions directed toward unhoused individuals, which could potentially explain the regression results above. In 2021, The Other Side Village was a community-based village for unhoused individuals to stay and receive the resources and support they require to grow (Gochnour, n.d.). Additionally, in Salt Lake County, the Rapid Rehousing Program has worked to establish security for unhoused individuals and thus, improve mental health outcomes (García & Kim, 2020).

The recursive regressions generally follow literature-established trends between income and mental health status. With Utah as the primary outlier, this could either point to a true break in the trend, as the literature on Utah's policy actions may suggest, or a methodological fault in the process.

Gender

In the literature, the general trend is that women disproportionately face more frequently poor mental health outcomes when compared to men (Astbury, 2001; Emslie et al., 2002; Picco et al., 2017). Specifically, women tend to face more "internalizing" issues like anxiety and depression, while men face more "externalizing" issues like substance use disorder (Needham & Hill, 2010). With these in mind, the expectation would be that the regressions that include sex as significant predictors will likely reflect this. To investigate this, we draw all regressions that include either "Percent Male" or "Percent Female" as a predictor:

Table 6: Coefficient value by gender and state.

State	Percent Female Coefficient	Percent Male Coefficient
Arkansas (AR)		-0.24

State	Percent Female Coefficient	Percent Male Coefficient
Arizona (AZ)	-0.13	-0.31
Iowa (IA)		0.13
Illinois (IL)		-0.11
Michigan (MI)	0.28	0.10
Missouri (MO)		-0.10
North Dakota (ND)	0.46	0.34
Ohio (OH)	0.16	
Oregon (OR)	0.12	
South Carolina (SC)		-0.13
South Dakota (SD)		-0.12
Texas (TX)		-0.09
Washington (WA)		-0.10
Wisconsin (WI)		-0.08

If the regressions were to follow the established gender trends, there would be three cases:

1) If only the percent female predictor is included, then it would be a positive coefficient (as the proportion of female individuals increases, poor mental health would be more prevalent).

2) If only the percent male predictor is included, then it would be a negative coefficient (as the proportion of male individuals decreases [in turn increasing the proportion of female individuals in the data], poor mental health would be less prevalent).
3) If both predictors are included, then the female coefficient would be more positive than the male predictor (for every one unit shift from the male proportion to the female proportion, the net predicted effect on poor mental health prevalence would be an increase).

Ohio and Oregon fall under case 1. Arkansas, Illinois, Missouri, South Carolina, South Dakota, Texas, Washington, and Wisconsin fall under case 2. Arizona, Michigan, and North Dakota fall under case 3. The sole state that does not fall under the three cases and breaks from the gender trend is Iowa, which predicts that zip codes with a higher male proportion would have more prevalent poor mental health (1 unit increase in percent male leads to a 0.12 unit increase in proportion mental health not good).

While Utah's regression had health insurance coefficients as an aligning trend with its income coefficients, there are no similarly apparent alignments with Iowa's male coefficients.

Table 7: Coefficient value for each predictor variable.

Predictor Variable	Coefficient
Percent Between the ages of 10 and 19	0.10
Percent Between the ages of 70 and 79	-0.14
Percent Median Rent Cost	0.20
Percent Mortgage greater than 30 Percent of Income	0.11
Percent Renter Occupied Housing	0.14
Percent Disabled	0.07
Percent Not Fluent in English	0.09
Percent Veteran	-0.12
Percent Asian	-0.15

Predictor Variable	Coefficient
Percent Black	0.08
Percent Male	0.13

Diving into the literature on Iowa male mental health, one study identified that in an Iowa cohort, 75.7% of the suicides were male (Persons et al., 2019). However, this ratio of male-to-female suicides is not exclusive to Iowa; the CDC reported that the suicide rate among males was four times higher than among females (National Institute of Mental Health, 2024). Another study looked at Iowa farmers, which is relevant here, as 67% of Iowa farmers were male in 2022 (Vilsack & Hamer, 2024). When specifically compared to Colorado farmers, Iowa farmers were 1.74 times more likely to face depressive symptoms (Scarth et al., 2000). Generally, male-dominated occupations, which farming falls under, face higher rates of depression compared to national baselines (Roche et al., 2016). Interestingly, if male-dominated occupations are the explanation for this break in gender trend for Iowa, one would expect occupational variables that are already present in the data to arise as significant predictors, but they are absent here. This could be partly due to the VIF filter opting to keep the covarying gender variable over the covarying occupational predictors.

Much like income, the regressions followed previously established trends save for one outlier. Iowa's regression could point to a true break in trend, potentially in relation to male-dominated occupations, or a fault in the recursive regression analysis.

Employment

In the literature, employment correlates with better mental health outcomes when compared to unemployment, with the theory being that employment provides a structure for community integration and room for self-confidence and self-reliance to grow (Drake & Wallach, 2020; Frijters et al., 2014; Modini et al., 2016). To investigate whether the recursive regressions follow this trend, we pull the regressions that have one of the aforementioned top three frequent predictors in their regression.

Table 8: Employment related coefficients by state.

State	Percent Employed Coefficient	Percent Unemployed Coefficient	Percent Not in Labor Force Coefficient
Alabama (AL)	0.33		

State	Percent Employed Coefficient	Percent Unemployed Coefficient	Percent Not in Labor Force Coefficient
Arizona (AZ)		0.15	0.15
California (CA)	0.33	0.11	
Colorado (CO)	0.75	0.28	0.23
Connecticut (CT)			-0.31
Georgia (GA)		0.18	0.15
Idaho (ID)			0.40
Illinois (IL)			0.12
Indiana (IN)	0.32	0.14	0.30
Kansas (KS)	0.30	0.21	0.15
Kentucky (KY)	0.25	0.14	0.24
Louisiana (LA)	0.17	0.09	0.32
Massachusetts (MA)		0.21	0.19
Maryland (MD)	-0.46	-0.25	
Maine (ME)	0.57	0.19	-0.24
Michigan (MI)			0.09
Missouri (MO)	-0.40	-0.20	

State	Percent Employed Coefficient	Percent Unemployed Coefficient	Percent Not in Labor Force Coefficient
Mississippi (MS)	0.27	0.15	0.12
North Carolina (NC)	0.60	0.11	0.08
Nebraska (NE)	0.15		
New Jersey (NJ)			0.11
New Mexico (NM)	0.18	0.17	
New York (NY)	-0.44	-0.10	
Ohio (OH)	0.12		
Oklahoma (OK)	0.12	-0.11	0.18
Oregon (OR)	0.38	0.25	
Pennsylvania (PA)	0.24	0.14	
South Carolina (SC)	0.21		0.12
South Dakota (SD)	0.29	0.20	
Tennessee (TN)	0.30	0.07	
Texas (TX)	0.22	0.10	0.15
Virginia (VA)	0.45	0.13	0.23
Vermont (VT)	-0.17		0.20

State	Percent Employed Coefficient	Percent Unemployed Coefficient	Percent Not in Labor Force Coefficient
Washington (WA)		0.08	
Wisconsin (WI)	0.28	0.19	0.20
West Virginia (WV)	0.18	0.09	0.23
Wyoming (WY)	0.77	0.29	

For the regressions above, if they were to follow the literature trend, the coefficient for unemployment would likely be more positive than the coefficient for employment. However, we do not see this for many states, including California (0.33 vs 0.11), Colorado (0.75 vs 0.11), Indiana (0.33 vs 0.11), and more. Unlike income and gender, there is no single outlier state regression. In this case, it is difficult to isolate one state and look into specific reasons as to why the regressions disagree with the literature trend of employment and mental health. In this situation, there is not as strong of, if any, agreement between the employment coefficients and the literature trends.

Conclusion

Overall, the results of the recursive regression were mixed. To begin, the three outlier regressions with extreme coefficients draw concern for either the original data source or methodology. One proposed explanation for the large coefficients was a large degree of covariability between those predictors, but in theory, this should have been addressed in the VIF filtering step. Furthermore, many regressions still had the dummy variable trap.

For example, for sex, Michigan and North Dakota included “Percent Male” and “Percent Female”. Since the data did not have a category for non-binary gender, these should be the only two options in the data, which means VIF filtering, in theory, should have led to at most one of these being included.

When compared to the literature trends, the recursive regressions showed mixed agreement. For income brackets, most regressions showed the trend of higher incomes leading to better mental health outcomes, aside from Utah. Similarly, for gender, most regressions showed the trend of females having worse mental health outcomes than men, aside from Iowa.

However, for employment, many regressions disagreed with the literature, predicting better mental health outcomes for unemployed individuals versus employed individuals. It is difficult at this point to argue whether the regressions have unearthed specific state trends that break from the overall established literature trends or if there is some fault in the recursive regression analysis. We found potentially plausible explanations for Utah's and Iowa's breaks in trend (focused policy for homelessness and prevalence of male farmers/poor farming conditions).

There were key limitations in the process that may provide room for improvement. Ideally, the process can be implemented with more robust data. Data on the whole was missing from Florida and Delaware; having regressions for all fifty states would allow for better comparisons. The response variable used, "Percent Mental Health Not Good", also lies as a flawed response variable. While it succeeds in being widely available at the ZIP Code across the nation, it fails to capture specific nuances of various types of mental issues, like anxiety, depression, substance use disorder, schizophrenia, etc. However, in general, finding quality ZIP Code data is difficult; previous attempts to find more specific metrics that hone in on depression and suicide at this level of granularity failed before this analysis, and collecting that data could be an incredibly large endeavor.

In terms of the regression, once ironed out, there could be room for interaction terms and higher-order polynomial terms. For example, one study on income and mental health included quadratic regression terms for income to account for a U-shaped relation between income and mental health outcomes. However, including these terms could come at a high cost of complexity and processing time which may make this analysis not feasible. Finally, for the analysis of results, I was only able to look into three areas of predictors, but there are more questions to be answered regarding how the regressions handle other social factors like age and race.

In conclusion, the mixed results of the recursive regressions show promise for a method that can ideally isolate potentially relevant social predictors from state to state. While the literature has worked to establish overall trends, the localized recursive regressions can help identify future research areas and inform policy actions from state to state – potentially increasing the granularity of the analysis. In other words, a factor that may be more relevant in one state may not be as relevant in another. Identifying such can help resource allocation and promote better mental health outcomes state to state.

References

1. Astbury, J. (2001). *Gender disparities in mental health*. Vuir.vu.edu.au. https://vuir.vu.edu.au/1656/
2. Drake, R. E., & Wallach, M. A. (2020). Employment is a critical mental health intervention. *Epidemiology and Psychiatric Sciences*, *29*(e178). https://doi.org/10.1017/s2045796020000906
3. Emslie, C., Fuhrer, R., Hunt, K., Macintyre, S., Shipley, M., & Stansfeld, S. (2002). Gender differences in mental health: evidence from three organisations. *Social Science & Medicine*, *54*(4), 621–624. https://doi.org/10.1016/s0277-9536(01)00056-9
4. Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of Health Economics*, *15*(1), 67–85. https://doi.org/10.1016/0167-6296(95)00032-1
5. Frijters, P., Johnston, D. W., & Shields, M. A. (2014). THE EFFECT OF MENTAL HEALTH ON EMPLOYMENT: EVIDENCE FROM AUSTRALIAN PANEL DATA. *Health Economics*, *23*(9), 1058–1071. https://doi.org/10.1002/hec.3083
6. García, I., & Kim, K. (2020). "I Felt Safe": The Role of the Rapid Rehousing Program in Supporting the Security of Families Experiencing Homelessness in Salt Lake County, Utah. *International Journal of Environmental Research and Public Health*, *17*(13), 4840. https://doi.org/10.3390/ijerph17134840
7. Gochnour, N. (n.d.). *Utah and Salt Lake City Policy Innovations in Homelessness, Poverty, and Health*. https://joelkotkin.com/wp-content/uploads/2023/01/The-Future-of-Cities_Utah-Innovations.pdf
8. Hamersma, S., & Ye, J. (2021). The effect of public health insurance expansions on the mental and behavioral health of girls and boys. *Social Science & Medicine*, *280*, 113998. https://doi.org/10.1016/j.socscimed.2021.113998
9. Kozloff, N. (2017, August 23). *Insurance Coverage and Health Outcomes in Young Adults With Mental Illness Following the Affordable Care Act Dependent Coverage Expansion*. Psychiatrist.com; Primary Care Companion for CNS Disorders. https://www.psychiatrist.com/jcp/insurance-coverage-and-health-outcomes-in-young-adults-with-mental-illness-following-the-aca/
10. Lang, M. (2011). THE IMPACT OF MENTAL HEALTH INSURANCE LAWS ON STATE SUICIDE RATES. *Health Economics*, *22*(1), 73–88. https://doi.org/10.1002/hec.1816
11. Li, C., Ning, G., Wang, L., & Chen, F. (2022). More income, less depression? Revisiting the nonlinear and heterogeneous relationship between income and mental health. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1016286
12. Modini, M., Joyce, S., Mykletun, A., Christensen, H., Bryant, R. A., Mitchell, P. B., & Harvey, S. B. (2016). The mental health benefits of employment: Results of a systematic meta-review. *Australasian Psychiatry*, *24*(4), 331–336. https://doi.org/10.1177/1039856215618523
13. National Institute of Mental Health. (2024, September). *Mental illness*. National Institute of Mental Health. https://www.nimh.nih.gov/health/statistics/mental-illness
14. Needham, B., & Hill, T. D. (2010). Do gender differences in mental health contribute to gender differences in physical health? *Social Science & Medicine*, *71*(8), 1472–1479. https://doi.org/10.1016/j.socscimed.2010.07.016
15. Persons, J. E., Hefti, M. M., & Nashelsky, M. B. (2019). Epidemiology of suicide in an Iowa cohort. *Public Health*, *177*, 128–134. https://doi.org/10.1016/j.puhe.2019.08.005
16. Picco, L., Subramaniam, M., Abdin, E., Vaingankar, J., & Chong, S. (2017). Gender differences in major depressive disorder: findings from the Singapore Mental Health Study. *Singapore Medical Journal*, *58*(11), 649–655. https://doi.org/10.11622/smedj.2016144
17. Roche, A. M., Pidd, K., Fischer, J. A., Lee, N., Scarfe, A., & Kostadinov, V. (2016). Men, Work, and Mental Health: A Systematic Review of Depression in Male-dominated Industries and Occupations. *Safety and Health at Work*, *7*(4), 268–283. https://doi.org/10.1016/j.shaw.2016.04.005

18. Sareen, J., Afifi, T. O., McMillan, K. A., & Asmundson, G. J. G. (2011). Relationship Between Household Income and Mental Disorders. *Archives of General Psychiatry*, *68*(4), 419. https://doi.org/10.1001/archgenpsychiatry.2011.15
19. Scarth, R. D., Stallones, L., Zwerling, C., & Burmeister, L. F. (2000). The prevalence of depressive symptoms and risk factors among Iowa and Colorado farmers. *American Journal of Industrial Medicine*, *37*(4), 382–389. https://doi.org/10.1002/(sici)1097-0274(200004)37:4%3C382::aid-ajim8%3E3.0.co;2-4
20. Thomson, R. M., Igelström, E., Purba, A. K., Shimonovich, M., Thomson, H., McCartney, G., Reeves, A., Leyland, A., Pearce, A., & Katikireddi, S. V. (2022). How do income changes impact on mental health and wellbeing for working-age adults? A systematic review and meta-analysis. *The Lancet Public Health*, *7*(6), e515–e528. https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(22)00058-5/fulltext
21. Vilsack, T., & Hamer, H. (2024). *Iowa State and County Data*. United States Department of Agriculture. https://www.nass.usda.gov/Publications/AgCensus/2022/Full_Report/Volume_1,_Chapter_1_State_Level/iowa/iav1.pdf