The Relationship Between Healthcare Resources and Academic Performance in the State of Connecticut

Emily Nolan

Department of Statistics

University of Connecticut

June 7, 2023

Abstract

This paper analyzes the relationship between childhood health insurance status and school performance, specifically in the State of Connecticut. Socieconomic status impacts childhood development in many ways. Studies have found a connection between childhood stress and poor school performance. This analysis shows how much, if at all, health insurance status, as a metric of socioecnomic status, impacts academic performance. When children are focused on staying healthy or overcoming illness, it becomes harder to prioritize school work. Stress at school can lead to multiple larger issues later in life. If improved healthcare can avoid some of this childhood stress, in turn creating healthier and more successful adults, then the State of Connecticut may consider more focus on healthcare resources. In this paper, a multiple linear regression model is created to analyze the significance of different variables in predicting school performance.

Keywords: childhood health, education, health insurance, public school.

1 Introduction

This paper plans to look at the correlation between healthcare and school performance (grades K-12 with an emphasis on high schools) in Connecticut Public Schools. "Although [socioeconomic status] has been at the core of a very active field of research, there seems to be an ongoing dispute about its conceptual meaning and empirical measurement in studies conducted with children and adolescents," Sirin (2005). The purpose of this paper is to determine if healthcare status, as an indicator of socioeconomic status, is a good predictor of school performance.

At home conditions, such as poor nutrition or health issues can impact a child's ability to perform at school. The stress of poor grades can create more illness and other ailments that can further harm school performance. Illness, both physical and mental, can lead to chronic absenteeism which is one of the better predictors of poor school performance. Brown et al. (2010) found that adverse experience in childhood, such as stress in school, can cause illness in adulthood such as heart disease, cancer, stroke, chronic lung disease, or diabetes. In order to create a good foundation for both physical and mental health of our children that can carry into adulthood, it is important to look at a well rounded picture of what causes good or bad school performance. If it is shown that healthcare may impact school performance, improving access to good healthcare may be a good step in order to avoid some of the adverse experiences that a child might face when getting bad grades. Liu et al. (2013) found that adverse childhood experiences leads to less success in employment and even unemployment later in life. Illness and unemployment can contribute to the cycle of low socioeconomic status, in order to break this cycle the cause of the childhood stress must be identified in the first place. One cause of these adverse childhood experiences is excess stress while in school. This stress itself can be caused by many factors, this paper is trying to identify if health insurance status is one of those factors. Access to good health care, which is a direct impact of having good health insurance, will lead to healthier and better cared for children. Without the stress of illness, children can focus on their school work, among other important aspects of growing up.

The rest of the paper is organized as follows. The data will be presented in Section 2. Section 3 describes the methods. The results are reported in Section 4. A discussion concludes in Section 5.

2 Data Description

The data used in this study is from either 2019 or the 2018-2019 school year. Later datasets of the variables that are available were from 2020 and 2021, but the Covid-19 Pandemic strongly impacted public school data, therefore, to keep consistent timing and remove barriers created by Covid, all sets are from 2019.

All datasets used in the analysis contain 169 rows for the 169 towns and municipalities of Connecticut. Many different datasets were accessed to create the table used in this report. "Next Generation Accountability" is from a measure used by the State of Connecticut to "tell the story of how well a school is preparing its students for success in college, careers, and life" (public edsight.ct.gov, 2023). It is calculated from 12 indicators including, achievement, growth, chronic absenteeism, college and career readiness, high school graduation rate, post-secondary entrance, physical fitness, and the arts. The measure is a number from 1 to 100, the State of Connecticut sets the target to be 85 for districts.

The variable "Percent of Free Lunch Students" shows the number of students who qualify for Free/Reduced Lunch Prices divided by the number of students in the district to create a percent. "Percent of Student English Learners" is the number of students who "lack sufficient mastery of English" (public edsight.ct.gov, 2023) divided by the student population. "Percent of SpEd Students" is the number of students in a districts special education program divided by the student population. "Percent of Chronic Absenteeism" is the percent of the student population that "miss[ed] ten percent or greater of the total number of days enrolled in the school year for any reason (public edsight.ct.gov, 2023). "Four Year Graduation Rate" is the percent of the student population that graduated high school after four years.

The two insurance variables, “Percent Under 18 With Public Health Insurance” and “Percent Under 18 With Private Health Insurance” both come from a dataset from CTData.org. Both only account for those under the age of 18 residing in the State of Connecticut covered under public or private health insurance, respectively. The variable used in the analysis, “Percent Under 18 Without Health Insurance” is the sum of “Percent Under 18 with Public Health Insurance” and “Percent Under 18 with Private Health Insurance” subtracted from 100%.

The variables “Population”, “Median Household Income”, and “Equalized Net Grand List” (a measure of wealth) were pulled from the Municipal Grand List dataset from CTData.org.

Some datasets used had missing information. According to Ct.gov EdSight, the source of the following datasets discussed, missing data is either suppressed to ensure confidentiality, or the town has fewer than 5 students in the category. The next two assumptions were made only for towns in which the population is less than 8,000. For the dataset used to get the variable “Percent of Student English Learners”, there were 24 (of 169) towns missing data. For those missing points, the assumption 0% was used because if the data is missing due to confidentiality then there is no way of knowing the true number and if the data is missing because it is fewer than 5 students then it is rounded down to 0. The variable “Four Year Graduation Rate” had similar issues. If data is missing for a district it was rounded up to 100% because it is most likely that less than 5 students did not graduate.

3 Methods

After creating the table used by pulling in variables and cleansing the data, a correlation matrix was created. All variables in the data table were included in the correlation matrix. Due to significant multicollinearity among the predictors, only a subset of the predictors

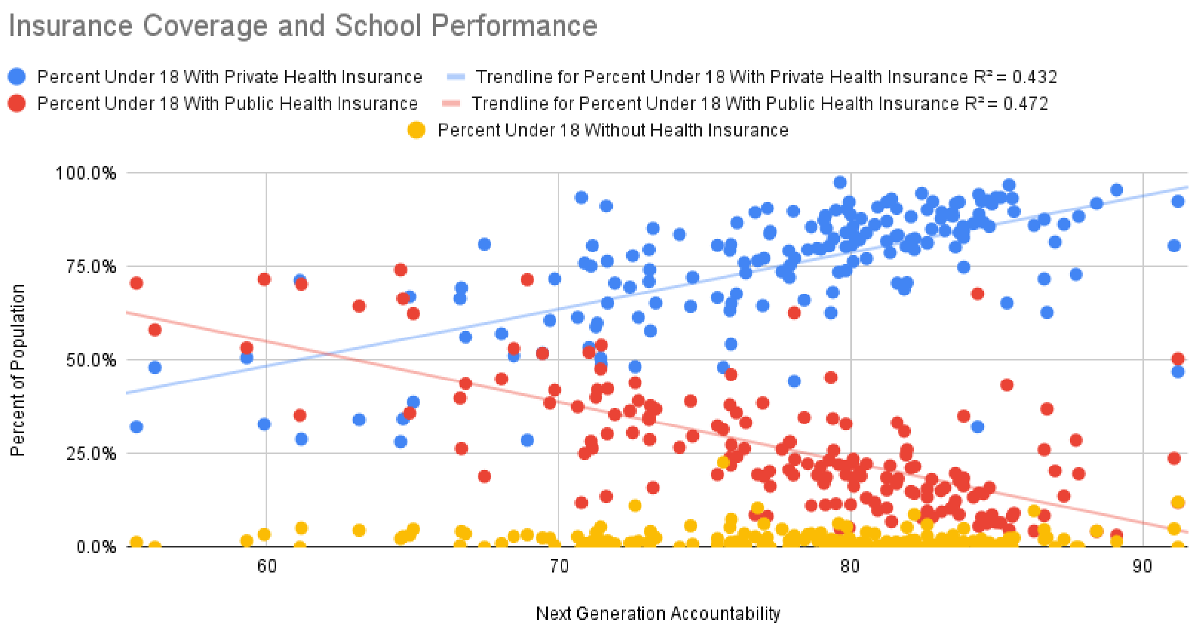


Figure 1: Scatter plot with Next Generation Accountability on the x axis and percent of population on the y axis. Graphed is the percent of a town's population under the age of 18 with public, private, and no health insurance.

that are not highly correlated will be used in the multiple linear regression model.

After choosing the subset of predictors, each variable is graphed with Next Generation Accountability separately to address the linearity assumption.

The assumptions required for a multiple regression model are listed below.

1. The random error is normally distributed with constant variance. This is addressed by showing residual histograms and residual plots.

2. Homoscedasticity is suggested.

3. There is no multicollinearity. This is addressed by the correlation matrix.

4. The predictors have a linear relationship with the response variable.

5. Observations are independent.

After confirming all the assumptions have been addressed, two multiple regression models were created and analyzed. One included Percent Under 18 with Public Health Insurance,

Correlations

	Equalized Net Grand List Per Ca	Median Household Income	Percent of Free Lunch Students	Percent of Student English Lear	Percent of SpEd Students
Median Household Income	0.637				
Percent of Free Lunch Students	-0.451	-0.710			
Percent of Student English Lear	-0.231	-0.421	0.720		
Percent of SpEd Students	-0.205	-0.363	0.531	0.231	
Percent of Chronic Absenteeism	-0.223	-0.520	0.738	0.549	0.424
Four Year Graduation Rate	0.399	0.581	-0.846	-0.700	-0.417
Percent of SNAP Recipients	-0.468	-0.636	0.872	0.724	0.413
Percent Under 18 With Public He	-0.374	-0.668	0.823	0.565	0.446
Percent Under 18 With Private H	0.322	0.630	-0.819	-0.619	-0.437

	Percent of Chronic Absenteeism	Four Year Graduation Rate	Percent of SNAP Recipients	Percent Under 18 With Public He
Median Household Income				
Percent of Free Lunch Students				
Percent of Student English Lear				
Percent of SpEd Students				
Percent of Chronic Absenteeism				
Four Year Graduation Rate	-0.732			
Percent of SNAP Recipients	0.712	-0.770		
Percent Under 18 With Public He	0.699	-0.715	0.798	
Percent Under 18 With Private H	-0.710	0.710	-0.776	-0.967

Figure 2: Correlation Matrix of all variables.

and the other Percent Under 18 with Private Health Insurance. These variables had to be separated into two different models because they are too highly correlated with each other. From the models created, the coefficients tables, model summaries, and analysis of variance tables are examined to draw conclusions. The MSE for each model is calculated and briefly discussed.

4 Results

The Variables Four Year Graduation, Percent of Chronic Absenteeism, and Equalized Net Grand List Per Capita were chosen because they seem to have the least correlation with each other. Variables with high correlation, as seen in Figure 2, were not included in this

Correlations

	Equalized Net Grand List Per Ca	Percent of Chronic Absenteeism	Four Year Graduation Rate	Percent Under 18 With Public He
Percent of Chronic Absenteeism	-0.223			
Four Year Graduation Rate	0.399	-0.732		
Percent Under 18 With Public He	-0.374	0.699	-0.715	
Percent Under 18 With Private H	0.322	-0.710	0.710	-0.967

Figure 3: Correlation Matrix of the variables used in the analysis.

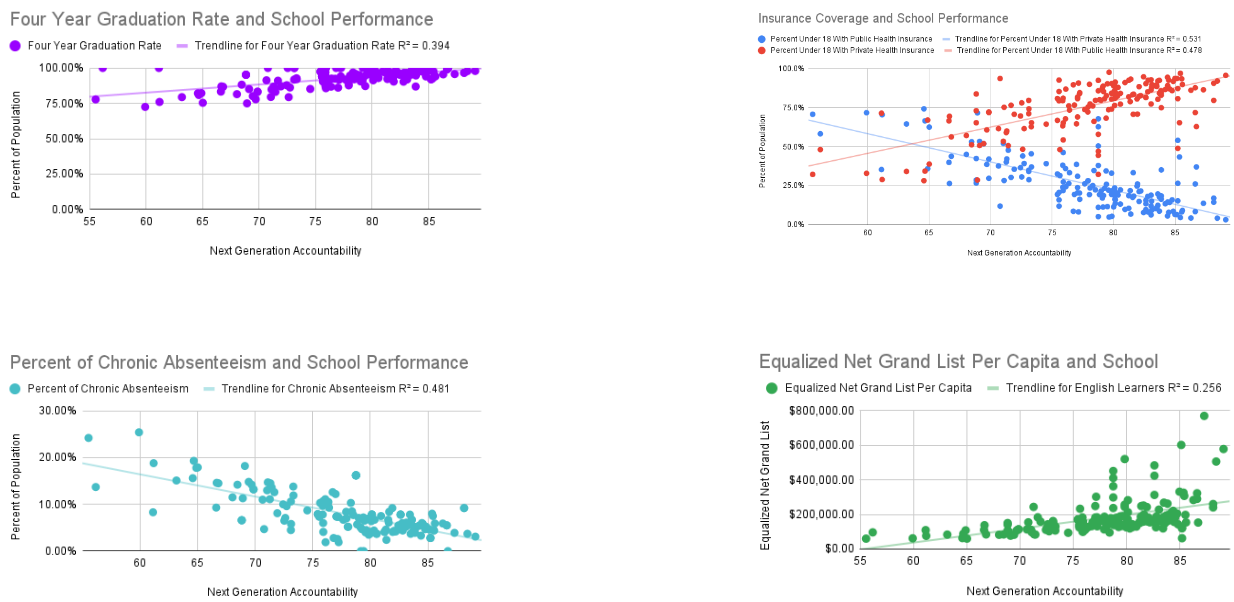


Figure 4: Linear relationships between each variable and Next Generation Accountability.

analysis.

As seen in Figure 3, Percent Under 18 with Public Health Insurance and Percent Under 18 with Private Health Insurance have an extremely high correlation, -0.967, which is why this analysis was broken into two different models. The correlation between Four Year Graduation Rate and Percent of Chronic Absenteeism is fairly strong at -0.732, this may cause some multicollinearity issues but the model was continued anyway.

Each variable has an approximate linear relationship with Next Generation Accountability. Equalized Net Grand List has the weakest linear correlation with Next Generation

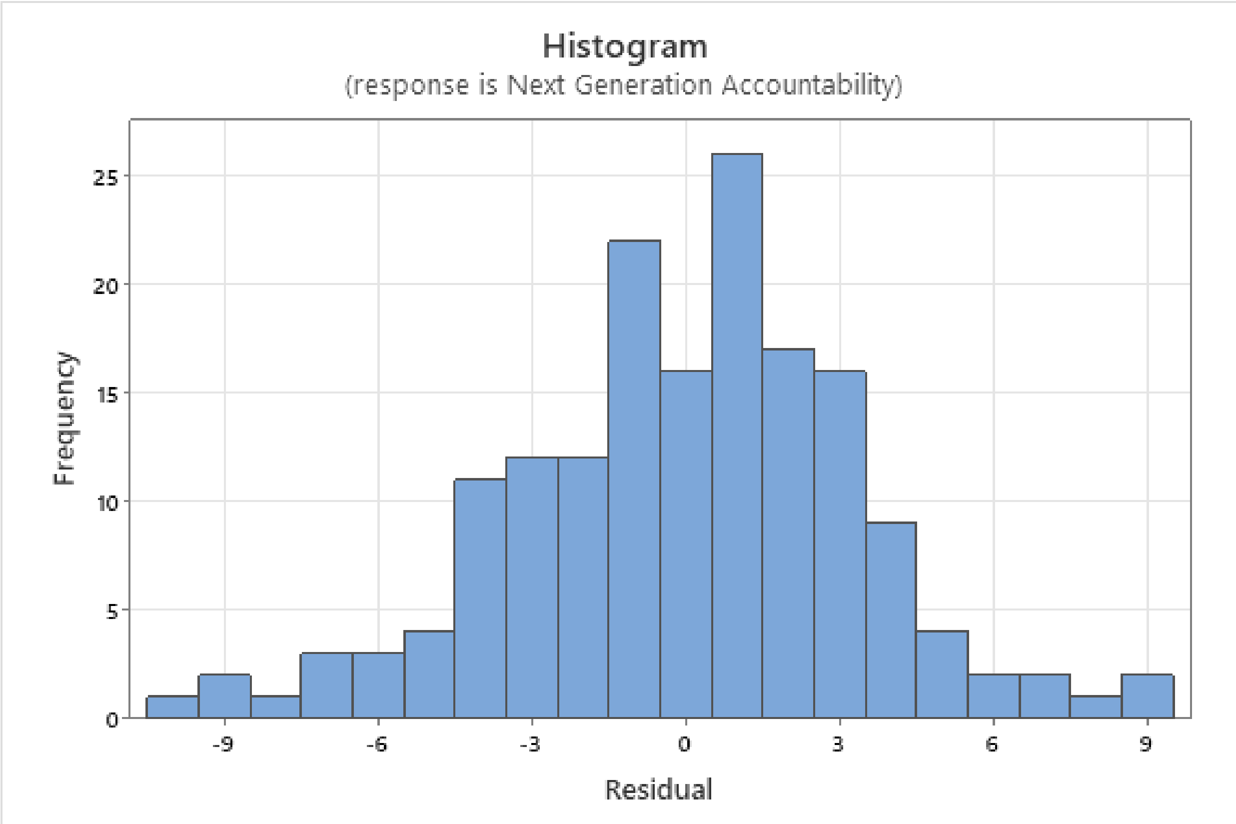


Figure 5: Residual Histogram for Next Generation Accountability from the Public Health Insurance model.

Accountability.

The histogram in Figure 5 shows that the distribution of residuals in the Public health Insurance model is approximately normal.

The Histogram in Figure 6 shows that the distribution of residuals in the Private Health Insurance model is also approximately normal, but slightly skewed in comparison to the Public Health Insurance histogram.

The four graphs, Residuals vs Fits and Plot of Residuals for each model, Public and Private Health Insurance (Figures 7 and 8) suggest homoscedasticity for both models.

Looking at the table for Public Health Insurance coefficients, Figure 9, the Variance Inflation Factor (VIF) values of the variables are fairly low (under 5), suggesting little to no multicollinearity. The coefficients for each variable match up with the scatterplots in Figure 4, with Percent of Chronic Absenteeism and Percent Under 18 with Public Health

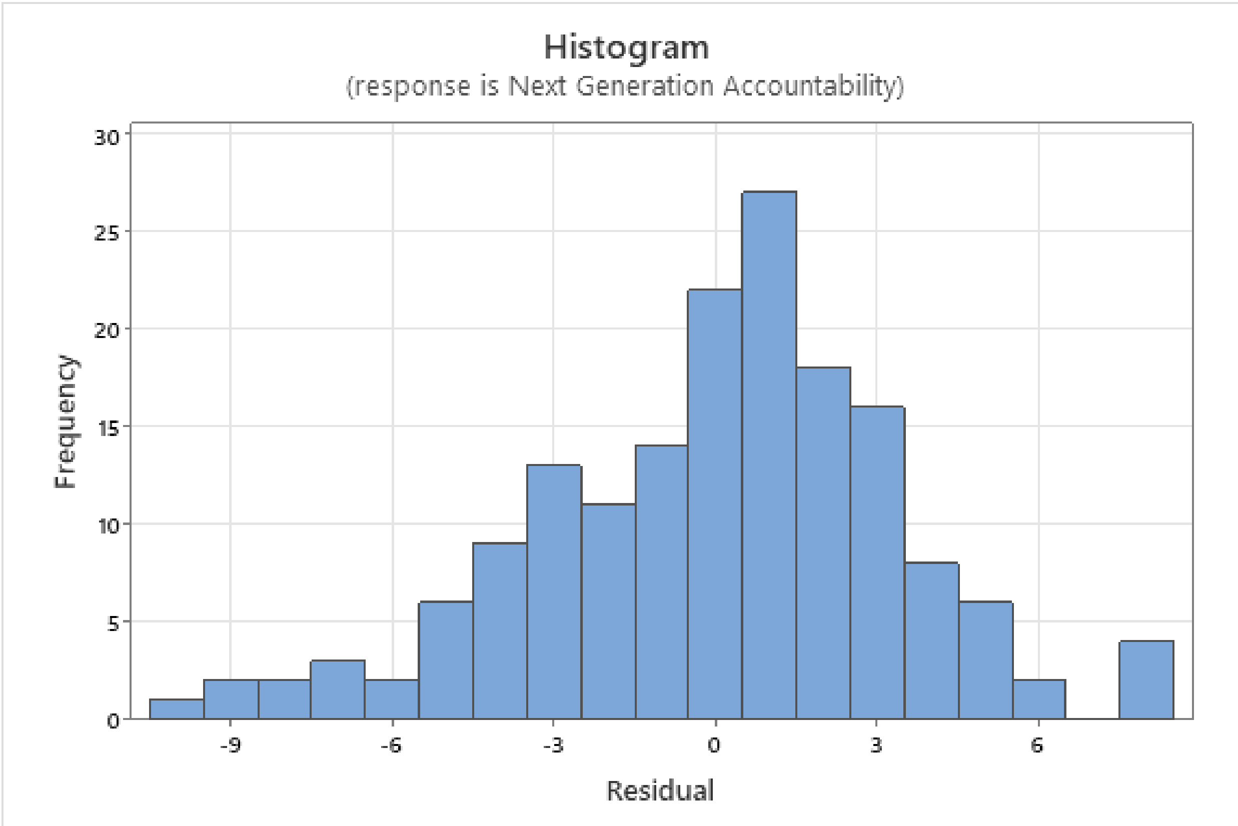


Figure 6: Residual Histogram for Next Generation Accountability from the Private Health Insurance model.

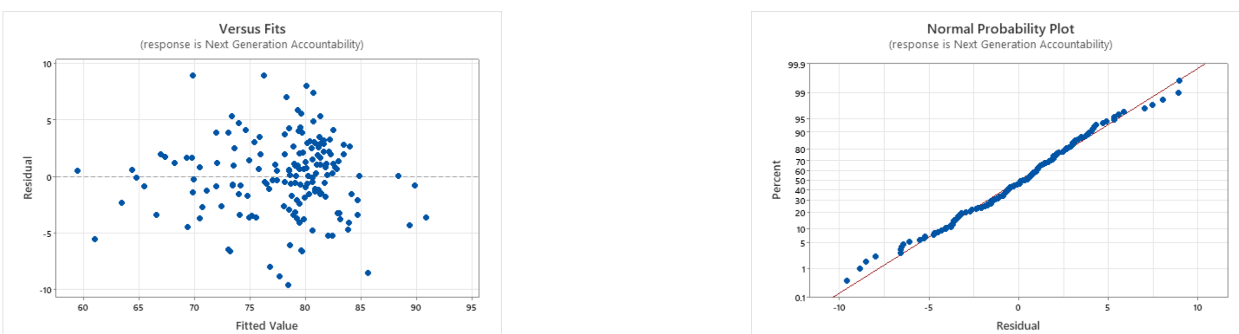


Figure 7: Residual plots of the Public Health Insurance model.

Insurance having negative relationships with the dependent variable, and with Equalized Net Grand List Per Capita and Four Year Graduation Rate having positive relationships with the dependent variable. Looking at the table for Private health Insurance coefficients, Figure 10, the VIF values of the variables are also below 5, suggesting little multicollinearity. The

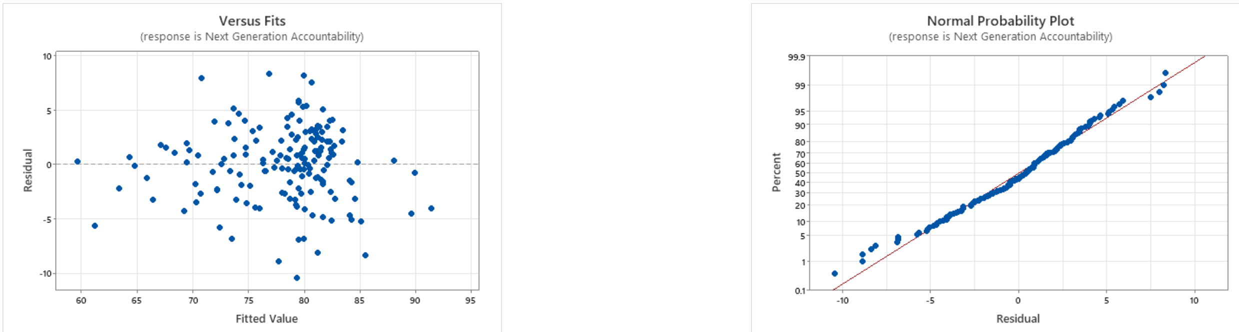


Figure 8: Residual plots of the Private Health Insurance model.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	64.83	7.38	8.79	0.000	
Percent of Chronic Absenteeism	-46.00	9.40	-4.89	0.000	2.56
Equalized Net Grand List Per Ca	0.000015	0.000003	5.46	0.000	1.25
Four Year Graduation Rate	18.30	7.41	2.47	0.015	2.82
Percent Under 18 With Public He	-11.06	2.55	-4.34	0.000	2.43

Figure 9: Coefficient Table from the Public Health Insurance multiple regression model.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	53.10	6.92	7.67	0.000	
Percent of Chronic Absenteeism	-47.85	9.63	-4.97	0.000	2.59
Equalized Net Grand List Per Ca	0.000017	0.000003	5.85	0.000	1.22
Four Year Graduation Rate	20.11	7.54	2.67	0.008	2.82
Percent Under 18 With Private H	9.40	2.62	3.59	0.000	2.40

Figure 10: Coefficient Table from the Private Health Insurance multiple regression model.

coefficients in this model also match the relationships in Figure 4. These relationships are the same as in the Public Health Insurance Model except for the fact that Private Health Insurance has a positive relationship with the dependent variable.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.40889	71.04%	70.32%	68.57%

Figure 11: Model Summary of the Public Health Insurance multiple linear regression model.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.46728	70.04%	69.29%	67.71%

Figure 12: Model Summary of the Private Health Insurance multiple linear regression model.

The adjusted r-squared value from Figure 11 shows a moderately strong fit of the Public Health Insurance model at 0.7032. The adjusted r-squared value from Figure 12, shows that the Private Health Insurance model is a bit weaker with a value of 0.6929. The Public Health Insurance model accounts for more variability in Next Generation Accountability.

When looking at the individual variables, a significance level of $\alpha = 0.05$ will be used. Both regressions have a p-value < 0.01 which is less than $\alpha = 0.05$, a sign that both Public and Private Health Insurance may predict school performance. In both models, Four Year Graduation Rate is the worst predictor of school performance. Both Private Health Insurance and Public Health Insurance as individual variables have p-values < 0.001 which is less than α which suggests these variables may be significant.

The MSE (Error Mean Sum of Squares) value of the Public Health Insurance Model is 11.62, whereas the MSE value for the Private Health Insurance Model is 12.02. This may be an indicator that Public Health Insurance is a better predictor for school performance because it has a smaller MSE. The MSE values are calculated by dividing the Error Sum of Squares value by the Error Degrees of Freedom, both found in the Analysis of Variance tables.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	4588.49	1147.12	98.72	0.000
Percent of Chronic Absenteeism	1	278.37	278.37	23.96	0.000
Equalized Net Grand List Per Ca	1	347.03	347.03	29.86	0.000
Four Year Graduation Rate	1	70.84	70.84	6.10	0.015
Percent Under 18 With Public He	1	219.23	219.23	18.87	0.000
Error	161	1870.91	11.62		
Total	165	6459.40			

Figure 13: Analysis of Variance table from the Public Health Insurance regression model.

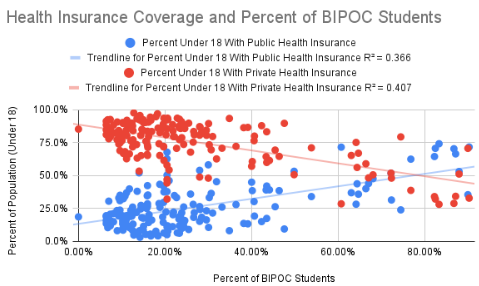
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	4523.85	1130.96	94.07	0.000
Percent of Chronic Absenteeism	1	296.89	296.89	24.70	0.000
Equalized Net Grand List Per Ca	1	411.21	411.21	34.21	0.000
Four Year Graduation Rate	1	85.63	85.63	7.12	0.008
Percent Under 18 With Private H	1	154.59	154.59	12.86	0.000
Error	161	1935.55	12.02		
Total	165	6459.40			

Figure 14: Analysis of Variance table from the Private Health Insurance regression model.

5 Discussion

There is an inability to accurately assess good quality healthcare vs poor quality healthcare. This issue doesn't seem to have an appropriate answer. Aseltine Jr et al. (2016) found a positive correlation between their health quality index and health insurance coverage. In contrast, Wray et al. (2021) found "individuals with private insurance were more likely to report poor access to care, higher costs of care, and less satisfaction with care compared with individuals covered by publicly sponsored insurance programs." (Wray et al., 2021) Although this shows that health insurance status is not a good measure of healthcare, it is still a good measure of socioeconomic status. Figure 16 shows that as median household



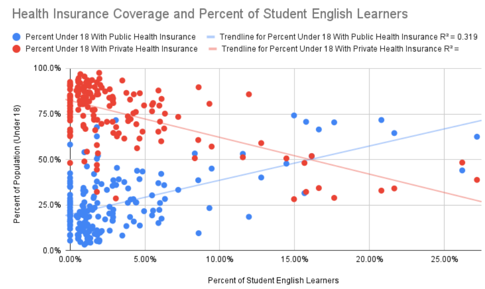
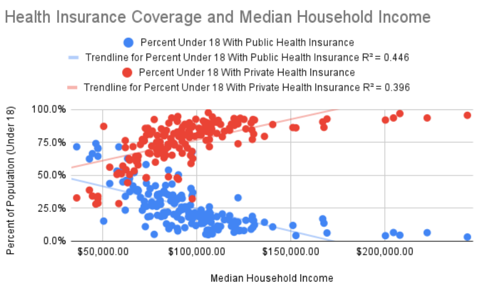


Figure 15: Scatterplots of Health Insurance Coverage vs Percent of BIPOC Students and Percent of Student English Learners.

income increases, the percent of children with private healthcare increases. Access to good healthcare is important for the development of a child and has a direct impact on school performance, but since the variable used in this analysis seems to not accurately represent health care quality, there is a larger focus on how health insurance status shows socioeconomic status instead.

Regardless of insurance coverage, many factors go into the quality of a family's healthcare. "[A] recent survey of Connecticut physicians that identified a number of shortcomings in the care provided to culturally and linguistically diverse patients. Only 38% of Connecticut physicians had received formal training in treating culturally diverse patients, and only 34% had completed CME on the subject." Aseltine Jr et al. (2016) As seen in Figure 15, there is a slight correlation between BIPOC Students, Student English Learners, and health insurance coverage. Towns that have a higher percentage of BIPOC or Student English Learners tend to have a higher percentage of people with public health insurance. Although it is hard to say if public health insurance is worse than private health insurance, there is a bit of a correlation between groups of people who do have worse quality of healthcare and have public insurance.

Something to consider when looking at the strength of this model is the multicollinearity assumption. Although the assumption was considered acceptable for the purposes of this paper, this assumption could also be considered unmet due to the high correlation between



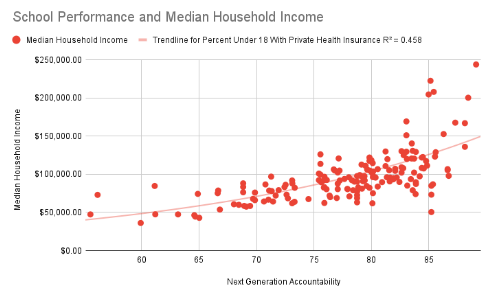


Figure 16: Scatterplots of Private Health Insurance Coverage, School Performance and Median Household Income.

variables, see Figure 3. This may impact the outcome and interpretation of this model. As well, the method in which variables were chosen, by picking least correlated from a correlation table, may impact the accuracy of this analysis.

Although public and private insurance may not be terribly different in terms of healthcare quality, this can still provide insight. Figure 16 shows that when the percent of children with private health insurance increases, household income also increases. From this, it can be concluded that wealthier families are more likely to have private health insurance. Wealthier towns have more resources in general to increase school performance. This includes access to private tutors, after school activities, better nutrition, and many more factors. Despite the fact that private health insurance may not be an indicator of school performance because of healthcare quality, it may be an indicator of school performance for other reasons.

The data used in this analysis is reliable. Since it comes from government websites we can assume that it is accurate. The data is also the full measurement of the population, not random samples which makes it more accurate. The limitation of the data is that there were a fair amount of missing data. As discussed in section 2, for the purposes of this analysis many towns have their Student English Learner percents rounded down and some had their Four Year Graduation Rate rounded up. This could possibly impact the accuracy of the model.

I believe this subject area is worth pursing further. There is some evidence that health insurance, and therefore health care access, is correlated with school performance. If better healthcare can create a better learning environment for students in Connecticut public schools, it may be worth looking more into the relationship of these two variables.

References

R. H. Aseltine Jr, A. Sabina, G. Barclay and G. Graham (2016). "Variation in patient–provider communication by patient's race and ethnicity, provider type, and continuity in and site of care: An analysis of data from the connecticut health care survey." *SAGE open medicine* **4**.

D. W. Brown, R. F. Anda, V. J. Felitti, V. J. Edwards, A. M. Malarcher, J. B. Croft and W. H. Giles (2010). "Adverse childhood experiences are associated with the risk of lung cancer: a prospective cohort study." *BMC public health* **10**, 1–12.

Y. Liu, J. B. Croft, D. P. Chapman, G. S. Perry, K. J. Greenlund, G. Zhao and V. J. Edwards (2013). "Relationship between adverse childhood experiences and unemployment among adults from five us states." *Social psychiatry and psychiatric epidemiology* **48**, 357–369.

public edsight.ct.gov (2023). "Next generation accountability system." `https://portal.ct.gov/SDE/Performance/Performance-and-Accountability/Next-Generation-Accountability-System`.

S. R. Sirin (2005). "Socioeconomic status and academic achievement: A meta-analytic review of research." *Review of educational research* **75**, 417–453.

C. M. Wray, M. Khare and S. Keyhani (2021). "Access to care, cost of care, and satisfaction with care among adults with private and public health insurance in the us." *JAMA Network Open* **4**.