Bridging the Digital Divide: The Impact of High-Speed Internet Access on Health

Team 16958

Modeling the Future Challenge 2024

March 3rd, 2024

Contents

Executive Summary

Although internet access has expanded in recent years, this increase has not accrued equally across the nation. Instead, there is a “digital divide”, where rural areas and pockets of urban centers lack adequate access to the internet. 42 million Americans currently lack access to high-speed broadband, which shuts them out of economic opportunities and hampers access to digital health resources [22]. Leaving this inequality unaddressed threatens to leave rural America behind, widening existing geographical and socioeconomic divides by exacerbating poor health outcomes in disadvantaged communities.

To analyze the risks posed by this issue and propose solutions, this work aims to determine the effects high speed internet access has on health outcomes. Specifically, we hope to determine what aspects of internet access, such as price and speed, have the greatest potential to help people across the country improve their health.

Utilizing data about broadband speeds, coverage, and price from the Federal Communication Commission’s National Broadband Map [3], we consolidated our own data set that describes the speed and affordability of different types of internet at the county level in 2023. This data included the percentage of households in each county that was covered by internet services of certain speeds, as well as the percentage of households in each county that was covered by an internet plan that is considered “low-cost”. This combination of factors aimed to capture both the quality and affordability of internet coverage across the nation.

Our data about broadband speed and price was then used alongside county level data about health outcomes, such as life expectancy and child mortality, as well as the prevalence of preventable causes of death such as diabetes and drug overdoses. We used an Ordinary Least Squares Regression model to perform linear regression on our broadband and health data. Our model revealed that the percentage of households with internet access as well as the availability of low-cost plans were the most statistically significant factors in predicting health outcomes.

Based on those results, we recommend major policy strategies to decrease the cost of broadband and increase access. First, the federal government should aim to continue existing financial assistance programs like the Affordable Connectivity Program, which provides financial aid to qualifying families to help them afford internet access. Second, the internet provider market should be pushed to move away from the current monopoly, which causes higher prices and lower quality services. By supporting the entrance of smaller providers into the market and regulating the current dominant providers, the government can encourage competition that will drive down prices and increase access across the nation. These steps will prove crucial in addressing chronic health inequities and building up robust digital healthcare systems in rural and other underserved areas. We believe that through these recommendations, the United States can begin to close the digital divide in healthcare and ensure every American can access vital internet services and improve their lives.

Introduction & Background

The internet has become the foundation of modern life, impacting almost every sector and allowing access to vital services. Thus, providing every American with internet access has become a policy priority in recent years, leading to significant efforts to increase access to high-speed broadband across the United States. Recent Federal Communications Commission (FCC) estimates report that the number of Americans living in areas without access to the Commission's current standard of adequate internet (25/3 Mbps) declined by 20% from the end of 2018 to the end of 2019. This progress occurred largely in rural areas, helping close a geographical divide in internet access. However, 14.5 million Americans still lack adequate internet access [1] and are unable to access vital services. This phenomenon is called the "digital divide" and it continues to impact peoples' lives today. It is critical to assess what aspects of current efforts are successful in improving outcomes for people to ensure that future policy continues to expand the benefits of the internet.

Increasing broadband access aims to reduce the various risks associated with inadequate access to internet. Specifically, the COVID-19 pandemic highlighted the crucial role broadband access plays in American healthcare by increasing the availability of health information, allowing remote scheduling of health services, providing access to telehealth, and more.

Previous studies have established a relationship between broadband expansion and health outcomes. In August of 2023, the National Bureau of Economic Research found that broadband expansion between 1999 and 2008 improved health outcomes by 5% by allowing patients to choose higher-quality providers [2]. However, while the benefits of internet access on health are well studied, how different aspects of internet access, such as price and speed of connection, affect health remains an evolving area of research. These insights will prove particularly valuable when tailoring policy solutions to address the digital divide – by determining the aspects of internet connection that are the most significant in improving outcomes, policymakers can work to specifically improve those areas.

This work aims to assess how different characteristics of broadband access impact health outcomes and will study a variety of health determinants to analyze what aspects of health can be improved through expanding broadband. By conducting our study on a county level, we hope to identify geographical trends and provide tailored policy recommendations for counties struggling with poor health outcomes and low levels of internet access.

Data Methodology

I. Data Collection

In order to characterize different aspects of broadband access across the U.S., we used several sets of publicly available data from the FCC's National Broadband Map that reports on broadband availability and internet providers at the county level in 2023. In order to analyze health data, we used data from the 2023 County Health Rankings by the University of Wisconsin's Population Health Institute, which aims to build awareness of the multiple factors that influence health and support leaders in growing community power to improve health equity. Both of these datasets are regularly updated to check for errors and inconsistencies, making them reliable enough to analyze. The FCC also encourages people to submit their own reports of their internet availability if the existing data is incorrect or incomplete, leading to more credible reports of internet access.

Broadband Summary by Geography Type 2023 [3]
- Type of Data: Percentage of units for broadband serviceable locations contained within geographical regions for which providers report specific types of broadband services, which vary based on whether they are residential or for businesses, the type of broadband, and broadband speed – e.g. "residential copper technology at speeds of 25/3 Mbps"
- Source: Federal Communications Commission's 2023 National Broadband Map
- Purpose: This dataset was used to characterize each county's access to broadband technology of different types and speeds. These factors were primarily used to isolate the effect of broadband price on our dependent variables.

Provider Summary by Geography Type 2023 [4]
- Type of Data: Calculated percentage of units for broadband serviceable locations contained within geographical regions for which different providers report offering broadband service.
- Source: Federal Communications Commission's 2023 National Broadband Map
- Purpose: This dataset was used to calculate the availability of low-cost broadband in each county. Because data about the price of plans was tied to each provider, data about how much of each county the provider covered was necessary to determine the prices available to each county.

Urban Rate Survey 2023 [5]
- Type of Data: Descriptors of plans of different speeds offered by each broadband provider. These descriptors include speed, monthly charge, and total charge.
- Source: Federal Communications Commission's 2023 Urban Rate Survey
- Purpose: This dataset was used to determine the prices of different plans that are available in each county.

Provider List [6]
- Type of Data: Lists internet providers across the country along with a unique ID number
- Source: Federal Communications Commission's 2023 Urban Rate Survey
- Purpose: This dataset was used to link two other data sets – the Urban Rate Survey, which contained provider names, and the Provider Summary by Geography Type, which only contained provider IDs. By merging the two datasets based on matching ID and name, we were able to consolidate data both about price and coverage.

County Health Data 2023 [7]
- Type of Data: Lists significant health factors at the county level across the country, including factors such as life expectancy, child mortality, percentage of adults uninsured, and more.
- Source: University of Washington's Population Health Institute

- Purpose: This dataset was used to characterize our dependent variables, which were health outcomes. By containing a variety of metrics for health, this dataset allowed us to get a holistic view of how the internet may affect different aspects of healthcare.

II. Data Processing

As shown in Appendix 1, the first step in our data cleaning process was focusing specifically on what we wished to examine. Data about internet access for businesses, for example, was disregarded, as our study focuses on how broadband access in households would affect their health outcomes. Our dataset regarding health was also significantly trimmed down to focus primarily on health outcomes, such as life expectancy, rather than factors that affect health, such as employment. Much of our data from the FCC also included data at various levels of geographical specificity, such as Congressional Districts, so this data had to be narrowed down to focus on the county level. After filtering for specifically county level data through geographical IDs, we worked to consolidate our data. Although all of our data about broadband came from the Federal Communications Commission, there were inconsistencies in how this data was reported. For instance, descriptors of the type of broadband technology differed across datasets, with fiber internet being categorized as “Fiber” in one dataset and “FTTH” (Fiber To The Home), in others. We performed preliminary data exploration to locate these inconsistencies and replace values to make the data more consistent.

Our data also needed to be simplified to display an understandable summary of broadband speed and price. In order to do this in terms of speed, we used the commonly accepted threshold for “adequate” internet service. Since 2015, the FCC has set the minimum speed of broadband to 25/3 Mbps which translate to a speed of 25 Mbps when downloading, which refers to the bits per second that travel to a user's device, and 3 Mbps when uploading, which refers to the bits per second that travel from a user's device [8]. However, after the COVID-19 pandemic revealed that the internet needs of the United States now may surpass that benchmark, the FCC has launched an inquiry that considers raising the threshold to 100/20 Mbps [9]. In order to examine the potential effects of this change, we chose to categorize broadband speed into three levels – at least 25/3 Mbps (the current standard), at least 100/20 Mbps (the proposed new standard), and at least 250/50 Mbps, which is considered to high-speed.

Furthermore, transforming our data to describe the prevalence of low-price internet plans was a significant part of our data processing. To calculate this coverage, we isolated providers in the Urban Rate Survey dataset who offered coverage at a monthly price of less than $50, a standard for low-cost internet that has also been used in recent studies done by BroadbandNow, an independent research institute that champions internet for all [10]. Then, using the Provider Summary by Geography dataset from the FCC, we were able to determine what percentage of each county these low-cost providers covered. By repeating this for each low-cost provider, we could total the percentages each provider covered to determine what percent of the county had access to a low-price plan. These percentages were computed for each type of internet coverage, such as fiber, cable, and wireless, just as our data regarding speed was.

III. Limitations

We recognize that the FCC has been previously criticized for significant limitations in their broadband data. The FCC Form 447 is the official source of broadband availability for the federal government and informs policy decisions despite several widely recognized flaws in the methodology. For one, this data is self-reported by internet service providers, who have the monetary incentive to overstate their coverage and speed statistics [11]. Although a challenge process has been created in the FCC's new 2022 broadband maps, it is unlikely that individuals can effectively police data that large internet companies may misreport. It has also been reported that the FCC's broadband maps overestimate coverage, as the entire census block can be considered served if a provider claims to be able to serve a single residence [12]. Despite these limitations however, we believe the FCC's National Broadband Map is still a valuable dataset to characterize internet access, as there are no suitable alternatives at a nation-wide scale.

The way our variables were calculated also serve as limitations on our data. When calculating low-cost internet coverage, we made two assumptions that limit our data's capability to accurately describe availability. First, while our $50 threshold for "low-cost" is supported by existing literature, it is unlikely that it truly captures the affordability of the internet across the country. What price is considered affordable likely varies across the country, as household income can differ greatly between counties. Different types of internet service such as fiber and cable, moreover, may have different standards for what is considered low-cost. Second, we had to make assumptions about provider coverage, as one provider rarely covered an entire county. If Provider A covered 10% of a county and Provider B covered 15% of a county, we assumed that the county had 25% total coverage. This is not entirely accurate, as it is possible for parts of a county to be covered by multiple providers. This addition process led to some counties having low-cost coverage of above 100% – in those cases, those numbers were replaced by 100%. Although this process does introduce limitations to our data, we believe that calculating low-cost coverage was essential to determining how the price of broadband affects health outcomes.

Mathematics Methodology

I. Variables

Our study examines a range of broadband-related independent variables and their relation to various health outcomes. The predictor variables consist of different types of internet connections, including copper, cable, fiber, and wireless, each with varying speeds and coverage levels. These variables aim to represent the extent and nature of internet accessibility across different categories. We also include a variable representing the percentage of households with broadband access to gauge the overall penetration and availability of internet within a given county.

The outcome variables we chose are critical health and societal indicators. These variables were selected to provide a comprehensive overview of the population's health status. By analyzing these dependent variables in relation to our predictors, we aim to uncover patterns and relationships between broadband access and health outcomes.

Predictor Variables

Variable	Description
Coverage by Speed and Type	
copper_25_3	% of copper coverage at 25/3 Mbps
copper_100_20	% of copper coverage at 100/20 Mbps
copper_250_25	% of copper coverage at 250/25 Mbps
cable_25_3	% of cable coverage at 25/3 Mbps
cable_100_20	% of cable coverage at 100/20 Mbps
cable_250_25	% of cable coverage at 250/25 Mbps
fiber_25_3	% of fiber coverage at 25/3 Mbps
fiber_100_20	% of fiber coverage at 100/20 Mbps
fiber_250_25	% of fiber coverage at 250/25 Mbps
wireless_25_3	% of wireless coverage at 25/3 Mbps
wireless_100_20	% of wireless coverage at 100/20 Mbps
wireless_250_25	% of wireless coverage at 250/25 Mbps
Low-Price-Plan Coverage by Type	
copper_low_price_coverage	% of copper low-price-plan coverage (<$50)
cable_low_price_coverage	% of cable low-price-plan coverage (<$50)
fiber_low_price_coverage	% of fiber low-price-plan coverage (<$50)
wireless_low_price_coverage	% of wireless low-price-plan coverage (<$50)
% Access Overall	
'% Households with Broadband Access	% of households with broadband access

Health Outcome Variables

Variable	Description
Life Expectancy	Average number of years a person can expect to live

Child Mortality Rate	# of deaths among residents < 18 per 100,000 population
% Adults with Diabetes	% of adults aged 20 and above with diagnosed diabetes
% Food Insecure	% of population who lack adequate access to food
Drug Overdose Mortality Rate	# of drug poisoning deaths per 100,000 population
% Uninsured Adults	% of adults under age 65 without health insurance
% Uninsured Children	% of children under age 19 without health insurance

II. Ordinary Least Squares Regression

Our study utilizes an Ordinary Least Squares (OLS) regression model to examine the impact of various factors on key health outcomes (Appendix 2). The OLS model is a fundamental method commonly used in statistics for a linear regression model. It minimizes the sum of the squared differences between observed and predicted values, ultimately providing a line-of-best-fit through the data points.

The general formula for the OLS regression model is given by:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

The variables represent the following:

- Y represents the dependent variable (e.g., life expectancy, child mortality rate)
- $X_1, X_2, \dots, X_n$ are the independent variables (e.g., % households with broadband access, wireless coverage)
- β_0 is the intercept term
- $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients that measure the impact of each independent variable on Y
- ε is the error term, representing the difference between the observed and predicted values of Y

Using the OLS model allows us to ensure that the estimations of our predictor and outcome variables are unbiased and efficient under the classical linear regression assumptions. The OLS model's strength lies in its simplicity and interpretability. This formulation allows us to quantify the impact of a one-unit change in each predictor variable on the dependent variable, providing straightforward interpretations of the data. For instance, we can interpret β_1 as the average change in Y resulting from a one-unit change in X_1, holding all other variables constant. This provides us with a clear and direct interpretation of the relationships between internet accessibility and health outcomes.

Additionally, the OLS model's assumptions, including linearity or independence of errors, are generally well-met by large datasets like ours, further justifying its application. While it is acknowledged that the complexity of real-world data can sometimes challenge the assumptions underlying OLS, the model remains a robust starting point for examining linear associations critical to our study's objectives.

III. Results

Our results can be measured on a variety of metrics. The first is the R-squared (R^2) value. The R^2 value helps to evaluate the proportion of variance in the dependent variable that can be predicted from the independent variable. In essence, a high R^2 value indicates a model that can better explain the variability of the outcomes based on the predictors included in the model. However, it's important to note that while the R^2 value is a useful indicator of model fit, it is not the sole criterion for evaluating the quality of a model. In this case, many of our predictions have a relatively low R^2 value. This can be explained by the fact that internet access is not the sole contributor to health outcomes, and thus, the variance of our dependent variables cannot be fully explained by our predictors. Despite this, our model is still valuable in answering our research question – how internet access affects health – regardless of other factors that also do so.

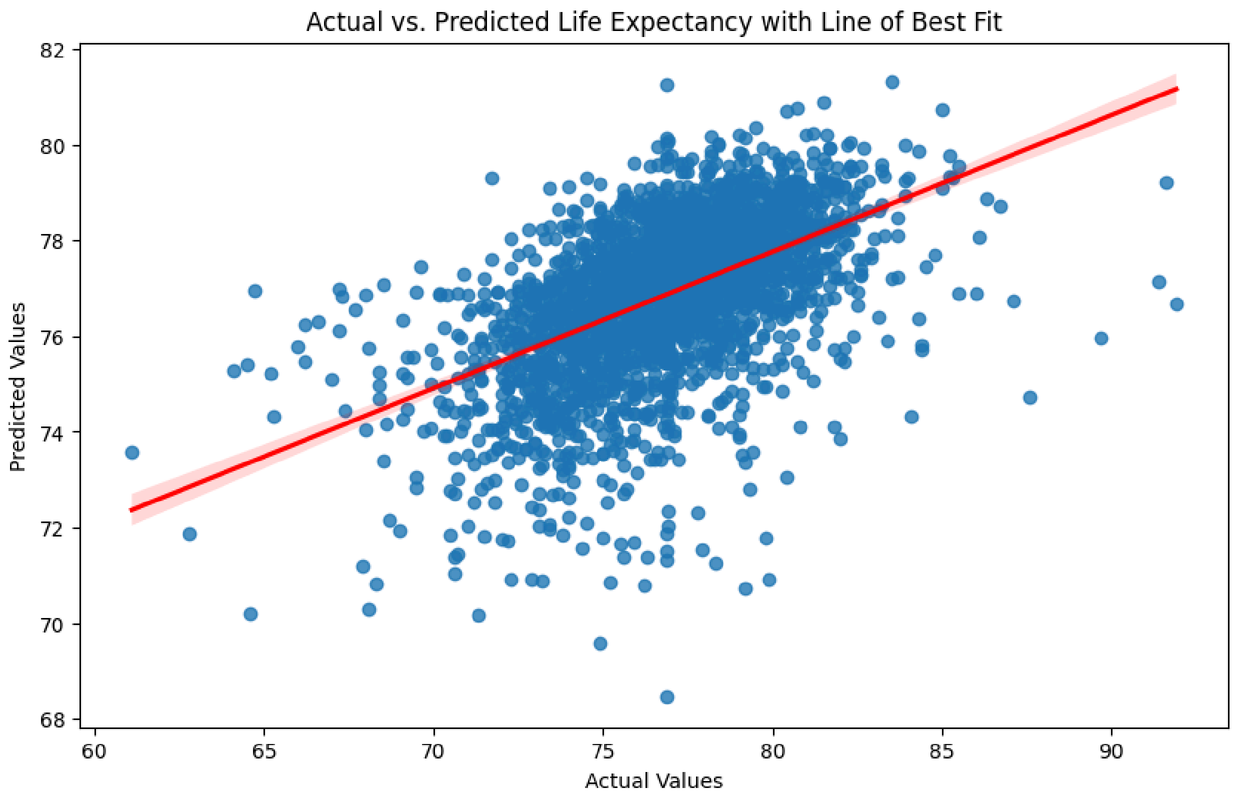
Additionally, each of our results were scrutinized for statistical significance through a combination of coefficient values and their corresponding p-values. The p-value helps determine the likelihood that an observed effect could occur by chance. A low p-value (<0.05) suggests that an effect is statistically significant, providing confidence in the reliability of the coefficient's estimation.

The magnitude of the coefficients, while important, was interpreted with caution due to differing units across variables. Instead, more emphasis was placed on the sign (positive or negative) of the coefficients, which indicates the direction of the relationship between each independent variable and the outcome variable. A positive sign suggests a direct relationship, whereas a negative sign indicates an inverse relationship.

In summary, the evaluation of our model was guided by the desire to achieve the best fit, as indicated by the R^2 value, while ensuring the statistical significance of our findings. This approach allows us to draw meaningful conclusions about the impact of broadband access on health outcomes, bearing in mind the limitations and the influence of external factors not accounted for in our model. Our results are shown on the following page, with significant results highlighted in blue.

<u>Life Expectancy</u>

$R^2 = 0.285$



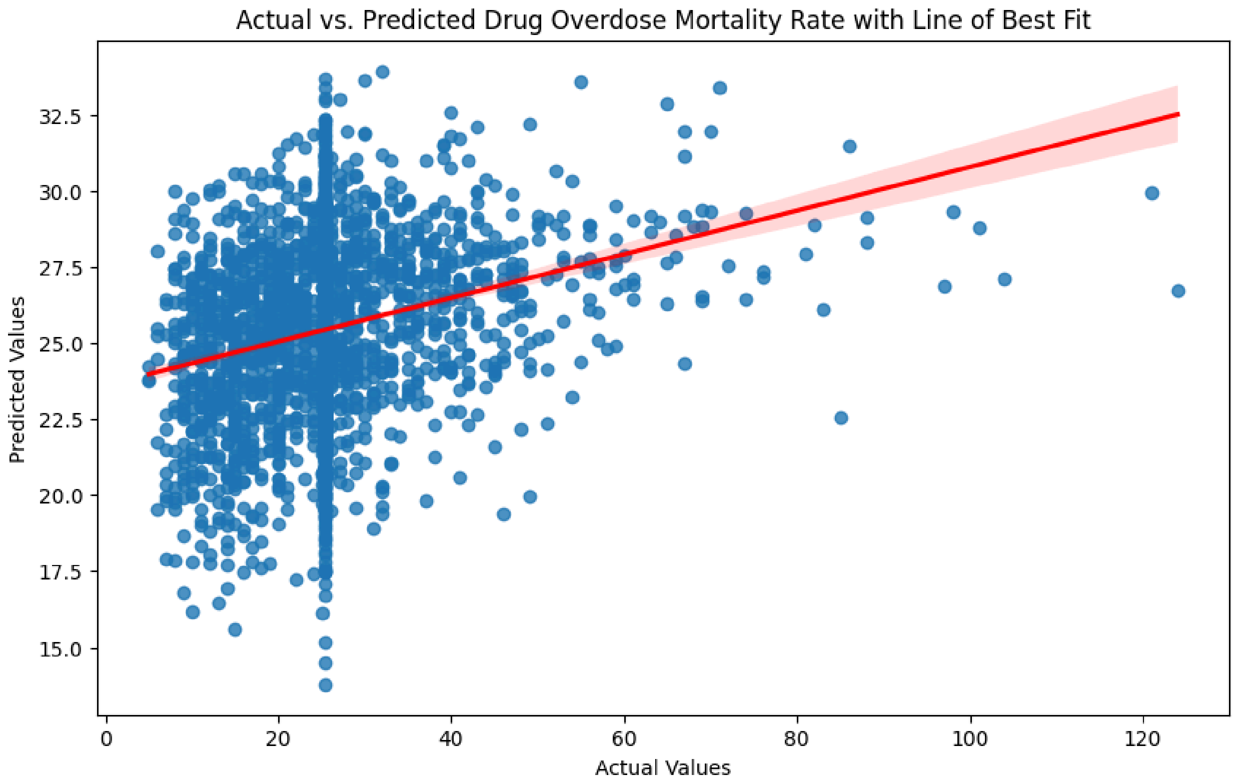
Statistically Significant Predictors (p-value < 0.05)

Variable	**Coefficient**	**P-value**
cable_25_3	3.7520	0.001
cable_250_25	-1.4039	0.011
fiber_250_25	2.8185	0.005
wireless_25_3	-1.5744	0.00
wireless_250_25	1.5620	0.001
wireless_100_20	-0.7148	0.036
wireless_low_price_coverage	0.2005	0.00
copper_low_speed_coverage	-0.1538	0.00
Percent households	-0.2797	0.00

The relevant and statistically significant variables are Cable Coverage at 25/3 Mbps, Fiber Coverage at 250/25 Mbps, Wireless Coverage at 250/25 Mbps, and Wireless Low Price Coverage. Note that we are looking for positive correlation coefficients for the given data as we expect an improvement in broadband to result in an increase in life expectancy.

Drug Overdose Mortality Rate

$R^2 = 0.072$



Statistically Significant Predictors (p-value < 0.05)

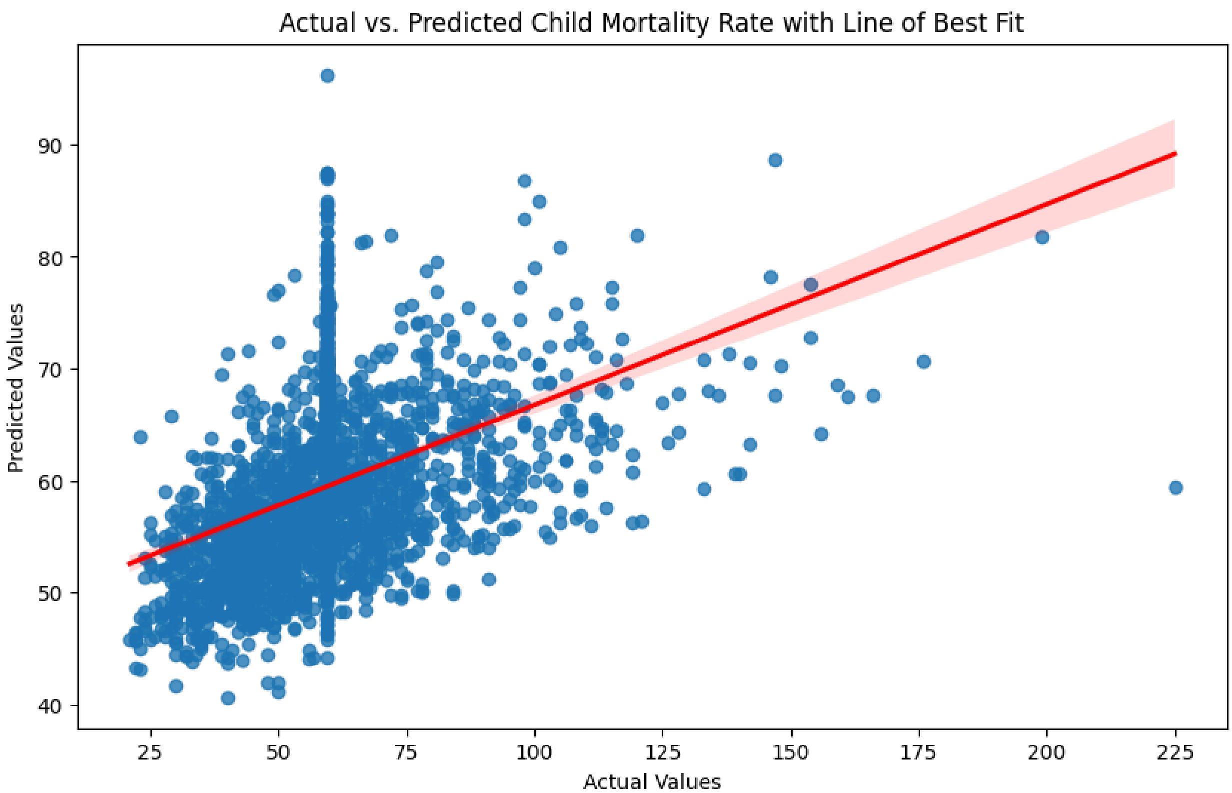
Variable	**Coefficient**	**P-value**
copper_25_3	-6.676	0.00
copper_100_20	13.0849	0.001
copper_250_25	-19.4548	0.013
cable_250_25	-4.275	0.024
wireless_25_3	-2.4931	0.021
fiber_low_speed_coverage	0.3200	0.00
cable_low_price_coverage	-0.2717	0.00

wireless_low_price_coverage	-0.4735	0.00
Percent households	-0.1527	0.00

The relevant and statistically significant variables are Copper Coverage at 25/3 Mbps, Copper Coverage at 250/25 Mbps, Cable Coverage at 250/25 Mbps, Cable Low Price Plan Coverage, Wireless Low Price Plan Coverage, and Percentage of Households with Broadband Access. Note that we are looking for negative correlation coefficients for the given data as we expect an improvement in broadband to result in a decrease in drug overdose mortality rate.

Child Mortality Rate

$R^2 = 0.179$



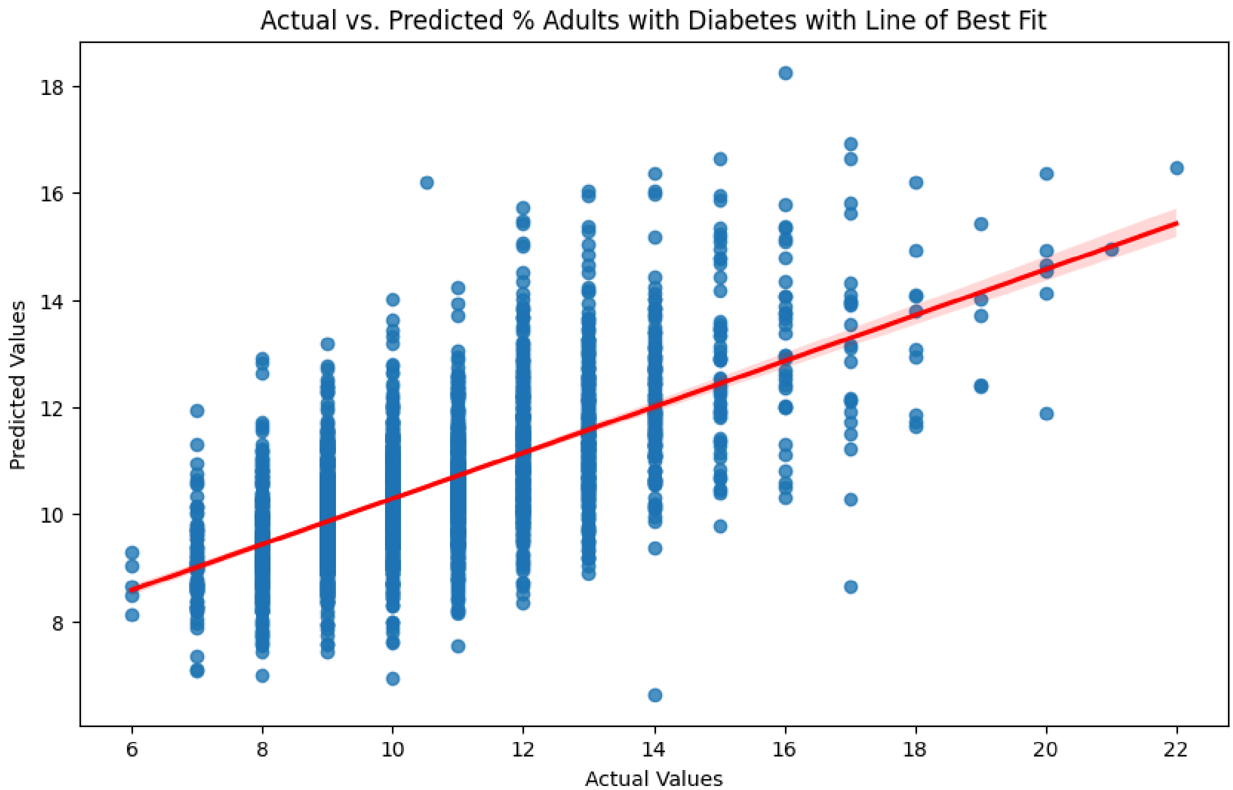
Statistically Significant Predictors (p-value < 0.05)

Variable	Coefficient	P-value
wireless_low_price_coverage	-0.3942	0.001
Percent households	-0.7513	0.00

This particular health outcome does not have many independent values that are statistically significant to its prediction. The only two are the Low-Speed Plan Coverage for wireless data and also the percentage of households with access to broadband. Both of these values demonstrate a negative correlation, which is expected as we expect an improvement in broadband to result in a decrease in child mortality rate.

<u>% Adults with diabetes</u>

$R^2 = 0.428$

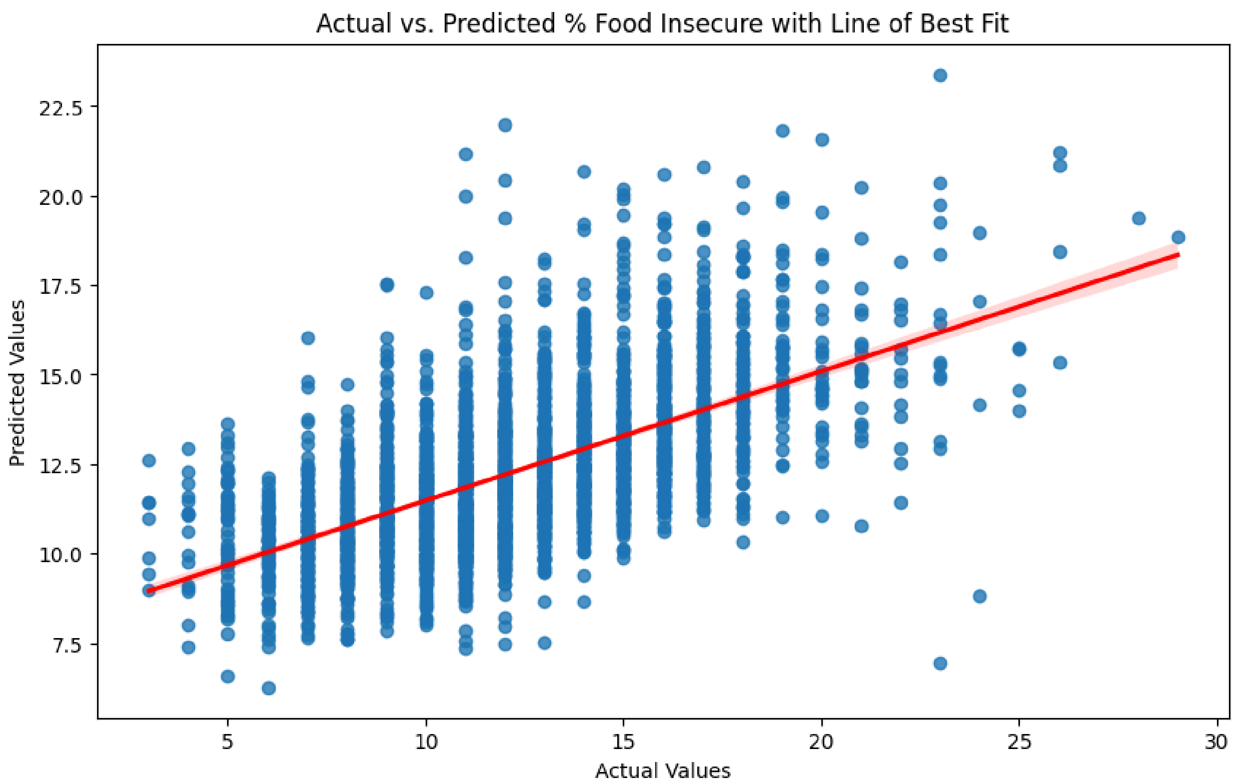


Statistically Significant Predictors (p-value < 0.05)

Variable	Coefficient	P-value
copper_100_20	1.9199	0.003
cable_25_3	2.353	0.00
wireless_25_3	-0.9210	0.00
fiber_low_price_coverage	0.0273	0.02
cable_low_price_coverage	0.0373	0.00
wireless_low_price_coverage	0.0947	0.00
% households	-0.1983	0.00

The relevant and statistically significant variables are Wireless Coverage at 25/3 Mbps and Percentage of Households with Broadband Access. Note that we are looking for negative correlation coefficients for the given data as we expect an improvement in broadband to result in a decrease in the percentage of adults with diabetes.

<u>% Food Insecure</u>
$R^2 = 0.361$



Statistically Significant Predictors (p-value < 0.05)

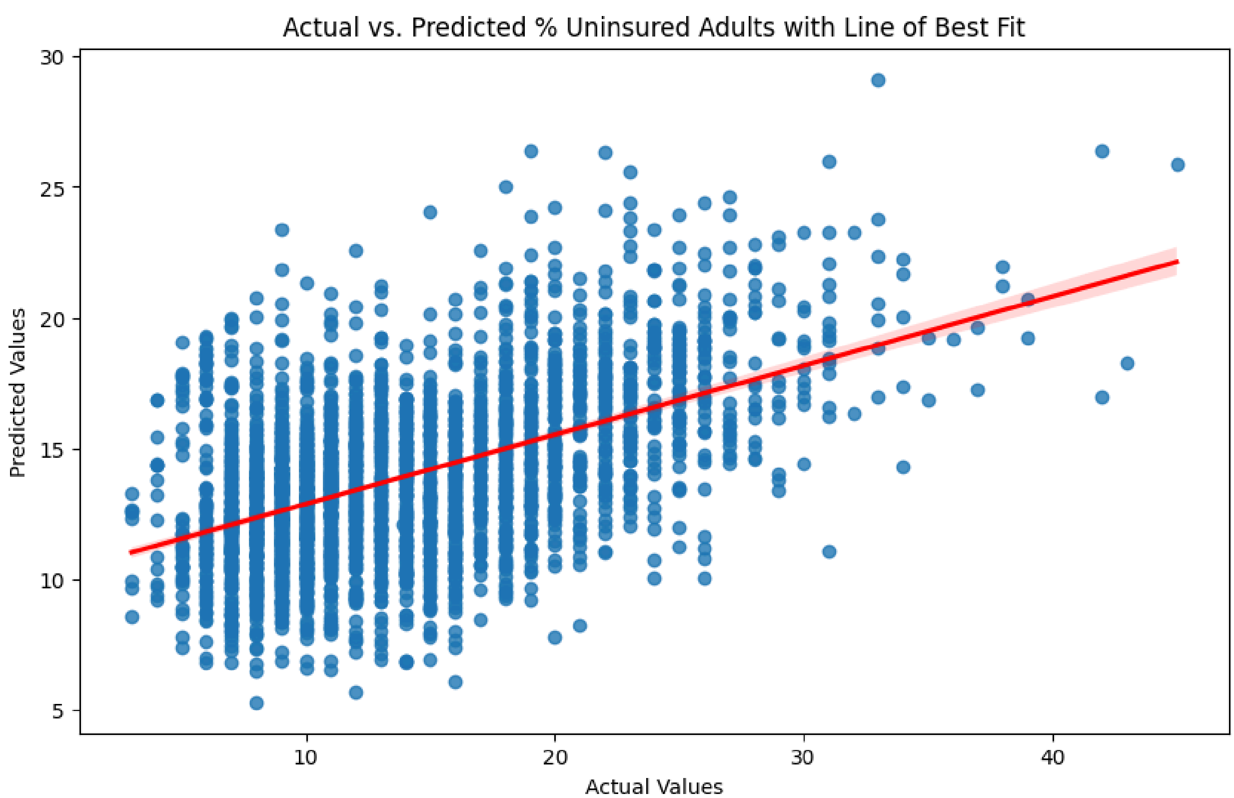
Variable	**Coefficient**	**P-value**
cable_25_3	3.7520	0.001
cable_250_25	-1.4039	0.011
fiber_250_25	2.8185	0.005
wireless_25_3	-1.5744	0.000
wireless_250_25	1.5620	0.001
wireless_100_20	-0.7148	0.036

wireless_low_price_coverage	0.2005	0.000
copper_low_price_coverage	-0.1538	0.000
% Households with Broadband Access	-0.2797	0.000

The relevant and statistically significant variables are Cable Coverage at 250/25 Mbps, Wireless Coverage at 25/3 Mbps, Copper Low Price Plan Coverage, and Percentage of Households with Broadband Access. Note that we are looking for negative correlation coefficients for the given data as we expect an improvement in broadband to result in a decrease in percentage of people experiencing food insecurity.

<u>% Uninsured Adults</u>

$R^2 = 0.265$



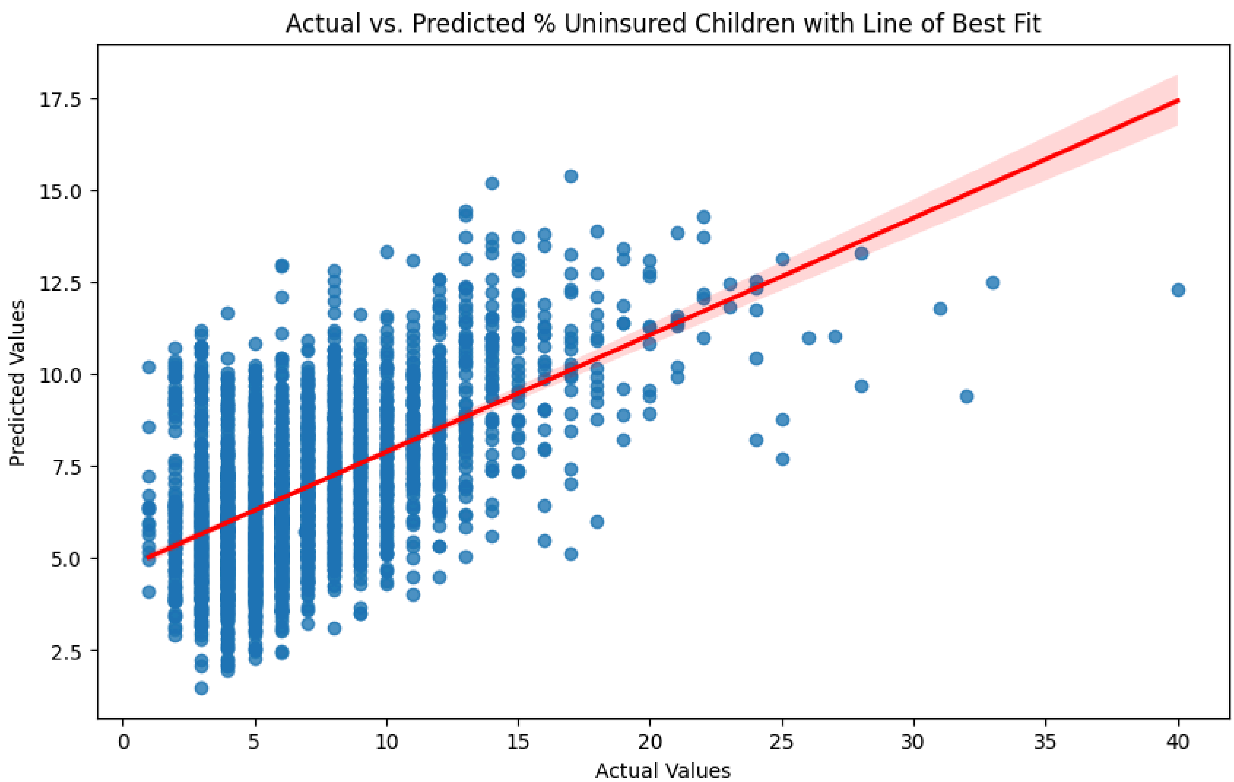
Statistically Significant Predictors (p-value < 0.05)

Variable	Coefficient	P-value
copper_25_3	2.1931	0.019
wireless_25_3	2.8027	0.000
copper_low_price_coverages	-0.4953	0.000

cable_low_price_coverage	0.0935	0.001
wireless_low_price_coverage	0.5565	0.000
copper_low_price_coverage	0.2756	0.000
% Households with Broadband Access	-0.3044	0.000

The relevant and statistically significant variables are Copper Low Price Plan Coverage and Percentage of Households with Broadband Access. Note that we are looking for negative correlation coefficients for the given data as we expect an improvement in broadband to result in a decrease in the percentage of uninsured adults.

<u>% Uninsured Children</u>
$R^2 = 0.318$



Statistically Significant Predictors (p-value < 0.05)

Variable	Coefficient	P-value
copper_100_20	4.3177	0.001
cable_100_20	-2.7886	0.031

wireless_25_3	1.7048	0.000
wireless_250_25	-1.1771	0.023
fiber_low_price_coverage	-0.2622	0.000
cable_low_price_coverage	0.0592	0.001
wireless_low_price_coverage	0.4352	0.000
copper_low_price_coverage	0.1327	0.001
% Households with Broadband Access	-0.0424	0.000

The relevant and statistically significant variables are Cable Coverage at 100/20 Mbps, Wireless Coverage at 250/25 Mbps, Fiber Low Price Plan Coverage and Percentage of Households with Broadband Access. Note that we are looking for negative correlation coefficients for the given data as we expect an improvement in broadband to result in a decrease in the percentage of uninsured children.

IV. Conclusions

For factors with p-values higher than 0.05, or those with high standard errors, we exercise caution in interpretation. High standard errors indicate a wide confidence interval around the coefficient estimate, reducing the reliability of the effect size indicated by the coefficient.

We will focus on variables with p-values of 0.05 or lower and clear directionality (as indicated by the sign of their coefficients), allowing us to identify the most reliable and significant relationships within our dataset.

Recurrent Significant Factors Across Outcomes

1) % Households with Broadband Access *(Frequency: 6)*
2) Wireless Low Price Coverage *(Frequency: 3)*
3) Copper Low Price Coverage *(Frequency: 2)*
4) Cable Coverage 25/3 Mbps *(Frequency: 2)*
5) Wireless Coverage 25/3 Mbps *(Frequency: 2)*

From the factors that are both statistically significant and make sense in the context of our analysis, we can make several determinations:

1) Overall broadband access seems to have the most impact. The frequent appearance of "% Households with Broadband Access" across multiple outcomes underscores the foundational importance of broadband connectivity, confirming the assumptions we made at the start of the study.

2) Low Price Plan Coverages also emerge as significant, suggesting that affordability is a crucial aspect of broadband access that influences health outcomes. Both Wireless and Copper Low Price Plan Coverage occur more than once as a significant factor across our outcomes, implying that notable improvements in public health metrics can be achieved when broadband services are financially accessible to a broader segment of the population.

3) The frequency of Cable and Wireless Coverage for 25/3 Mbps suggests that a baseline level of service is particularly beneficial. This specific speed tier, often considered the minimum for a service to be classified as “broadband” by many standards, appears to be a critical threshold for achieving positive health outcomes. There has been a lot of community discussion on whether to raise the baseline for broadband, but these results suggest that even at this foundational speed, significant health benefits can be realized.

Ultimately, our analysis highlights the significant role of broadband access, particularly affordable plans and baseline speed thresholds, in influencing a range of health outcomes. As we move towards increasingly digital societies, ensuring equitable access to effective and affordable broadband services emerges as a critical public health intervention.

Risk Analysis

The impact of limited internet access on health outcomes presents a significant risk to the livelihoods of Americans, both in rural areas and specific pockets of urban centers that are cut off from reliable high-speed internet. While it is challenging to precisely quantify the risks associated directly with a lack of internet access, the potential long-term ramifications of the digital divide for these communities warrants careful consideration and exploration of possible solutions.

The growing reliance on digital tools in healthcare delivery and health information dissemination raises significant concerns about the future of internet-driven health disparities. The National Telecommunications and Information Administration found that digital healthcare such as telemedicine is on the rise. Even before the COVID-19 pandemic forced much of healthcare to move online, the proportion of households that accessed health records online grew between 2017 and 2019. More than half of American households used the internet for health-related activities in 2019 [13]. Today, new developments in artificial intelligence and health informatics foresee an even more significant transition toward digital technologies in healthcare in the future, amplifying the risk of leaving communities that lack internet access behind.

Our results show that expanding internet access can help increase health metrics such as life expectancy and child mortality, and decrease preventable deaths from causes such as diabetes and drug overdoses. Lack of internet specifically poses a variety of risks towards an individual’s health outcomes, including:

- Decreased utilization of digital health services such as telehealth, which is crucial to reduce geographic inequalities in healthcare by allowing access to health for those physically far from a hospital. This lack of access to preventative care has significant risks, as less severe diseases become life-threatening when left untreated. For example, an undiagnosed and unaddressed case of high blood pressure increases the risk of heart failure, an issue that costs the American economy $216 million dollars a year. Digital health services can allow for these cases to be diagnosed early, allowing for preventative screenings, lifestyle changes, and medication to mitigate this risk [14]. Similarly, chronic health conditions like diabetes can be manageable and avoidable with proper care, but the digital divide threatens to cut off disadvantaged communities from accessing telehealth services that could prevent them.

- Reduced access to reliable health information, whether it be information about how to manage a chronic condition or the risks associated with activities such as smoking or drug use. Health information about these issues are critical – in 2020, it was estimated that the opioid epidemic, for example, cost the United States economy $1.5 trillion dollars [15], a number that is projected to only increase and does not even take into account the social and community-based costs of rampant addiction.

It is also crucial to acknowledge the potential for these risks to produce a cyclical effect. Limited access to the internet, and thus health services, contributes to the economic stagnation and poorer living conditions that are currently driving population outmigration of healthcare workers in rural areas. This phenomenon, known as “brain drain”, has been observed in the U.S. over the past 50 years as the nation has experienced major shifts in geographic mobility patterns among its highly-educated citizens, such as doctors and nurses [16]. This has left portions of the country unable to develop primary healthcare systems, creating a cycle of poor health that accelerates the failure of rural healthcare.

These current and future risks prove that examining the effects internet access has on healthcare in order to address inequities is crucial. Only through this can the United States become a nation where everyone has access to critical online resources and is able to improve their health outcomes.

Recommendations

Based on our analysis, we have identified two major recommendations for policymakers that aim to address the digital divide and mitigate its impacts on health outcomes. Our results found that access and affordability are the most influential factors on health, while the internet at speeds higher than the current baseline of 25/3 Mbps seems to have a limited effect. In accordance with these results, the following recommendations will seek to remove barriers to accessing the internet and increase affordability of high-speed services.

I. Financial Assistance Programs

Our results indicate that the availability of low-cost internet plans has a significant impact on a county’s health outcomes. Thus, it is critical that the federal government take action to make high speed internet more affordable.

A key example of how this could be implemented is the 2021 Affordable Connectivity Program. Launched by the federal government during the COVID-19 pandemic, the Affordable Connectivity Program is a benefit program under the Federal Communications Commission that helps households afford broadband internet connectivity needed for work, school, and healthcare. The benefit provides a discount of up to $30 per month for eligible households and $75 per month on households on qualifying Tribal lands. This program was created as part of the Infrastructure Investment and Jobs Act of 2021, permanently extending the previous Emergency Broadband Benefit [17].

This program was able to create real benefits for Americans. The Affordable Connectivity Program, as of February 2024, has been estimated to have helped 1 in 6 households in the United States, or 23 million households, save on their monthly internet bills. This progress was especially beneficial for those who already struggle with structural barriers to access, with 1 in 4 households participating in the program being African American and 1 in four households being Latino [18].

Unfortunately, the Affordable Connectivity Program is currently set to shut down in April of 2024 and has stopped taking new enrollments due to a lack of funding from Congress, threatening to reverse progress made throughout the pandemic and cutting millions of families off of the internet. A 2024 FCC survey revealed that 68% of respondents had "inconsistent connectivity or zero connectivity" prior to enrolling in the Affordable Connectivity Program and 29% of respondents said they would have to drop internet service without the $30 subsidy. Half of the latter respondents would be left with zero internet at all [19]. Given the progress made under the program, we urge the passage of bills such as the Affordable Connectivity Program Extension Act, which aims to invest $7 billion to continue the program [19] and recommend that policymakers work to expand similar programs to increase the affordability of high speed internet.

II. Encouraging Competition Among Internet Service Providers

A major factor behind why broadband costs are high is the nature of the home broadband market. The market has been described as a “localized monopoly”, where companies such as Comcast, Verizon, and AT&T dominate coverage. In fact, FCC data indicates that around 68% of US census blocks have only one provider offering proper broadband service that reaches the 25/3 Mbps baseline, or do not have such a provider at all [20].

This monopolization allows providers to avoid providing low-cost internet and gives them no incentive to improve internet speeds and accessibility. In order to address this issue, it is crucial to examine the reason why the market has developed this way – high costs. Building fiber internet, for example, can cost up to $5,000 per home in rural communities. At a standard broadband price of $60 per month, this high upfront cost makes it so that breaking even for providers could take up to 6 years. Entering an area where a provider is already established is even more expensive, as the cost of connecting to homes is the same but providers get less subscribers, and thus, less revenue [20].

Breaking down this monopoly will prove crucial in improving internet infrastructure, decreasing prices, and expanding access. A 2022 paper by Consumer Reports found that increasing competition among local providers was indeed successful in decreasing prices. People in areas where there were at least three broadband providers available reported paying, on average, about $5 per month less for service than those in areas where there were only one or two providers available [21]. Thus, we recommend the following two policy strategies to increase broadband competition:

1) The federal government should work to support the establishment of regional and municipal broadband strategies by providing funding to decrease barriers to entry in the internet market.
2) Regulatory bodies such as the FCC should put greater scrutiny on currently dominant internet providers, such as Comcast, to prevent anti-competitive business practices and encourage an even playing field for newcomers in the market.

By breaking down the current monopoly in home broadband, the federal government can encourage competition to decrease prices and increase the quality of internet services. These steps forward will increase internet infrastructure in underserved areas by decreasing costs for providers and allow people across the nation to access vital digital health services and online information that will decrease preventable disease and improve health outcomes.

References

[1] *Fourteenth Broadband Deployment Report.* Federal Communications Commission. (2021, January 19). https://www.fcc.gov/reports-research/reports/broadband-progress-reports/fourteenth-broadband-deployment-report

[2] Parys, J. V., & Brown, Z. Y. (2023, August 21). *Broadband internet access and health outcomes: Patient and provider responses in Medicare*. NBER. https://www.nber.org/papers/w31579

[3] Federal Communications Commission. (2023). *National Broadband Map – Broadband Summary by Geography Type [Data set]*. FCC. https://broadbandmap.fcc.gov/data-download/nationwide-data?version=jun2023

[4] Federal Communications Commission. (2023). *National Broadband Map – Provider Summary by Geography Type [Data set]*. FCC. https://broadbandmap.fcc.gov/data-download/nationwide-data?version=jun2023

[5] Federal Communications Commission. (2023). *Urban Rate Survey [Data set]*. FCC. https://us-fcc.app.box.com/v/2023URSBroadbandResults

[6] Federal Communications Commission. (2023). *National Broadband Map – Provider List [Data set]*. FCC. https://broadbandmap.fcc.gov/data-download/nationwide-data?version=jun2023

[7] County Health Rankings & Roadmaps. (2023). *2023 County Health Rankings National Data [Data set]*. University of Wisconsin Population Health Institute. https://www.countyhealthrankings.org/sites/default/files/media/document/2023%20County%20Health%20Rankings%20Data%20-%20v2.xlsx

[8] *Broadband speed: FCC should improve its communication of Advanced Telecommunications Capability assessments*. U.S. Government Accountability Office. (2023, April 25). https://www.gao.gov/products/gao-23-105655

[9] King, J. (2023, November 2). *FCC seeks input on upgrading national broadband speeds. Fierce Telecom*. https://www.fiercetelecom.com/broadband/fcc-seeks-input-upgrading-national-broadband-speeds

[10] Cooper, T. (2023, May 9). *High-income Americans pay less for internet service overall*. BroadbandNow. https://broadbandnow.com/research/broadband-pricing-inequities

[11] Liu, C. (2023, January 31). *The FCC broadband maps: Meet the new maps, same as the old maps*. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2023/01/fcc-broadband-map-has-problems

[12] National Broadband Mapping Coalition. (2022, August 1). *Assessing available databases: FCC Form 477.* Broadband Mapping Coalition. https://broadbandmappingcoalition.org/assessing-available-databases-fcc-form-477/

[13] Cao, M. (n.d.). *More than half of American households used the internet for health-related activities in 2019, NTIA Data Show*. National Telecommunications and Information

Administration.
https://www.ntia.gov/blog/2020/more-half-american-households-used-internet-health-related-activities-2019-ntia-data-show

[14] Lutz, A. (2023, April 24). *How rural hospitals can tackle cost-driven health care delays without reducing prices*. NRHA, National Rural Health Association.
https://www.ruralhealth.us/blogs/ruralhealthvoices/april-2023/how-rural-hospitals-can-tackle-cost-driven-health

[15] *JEC analysis finds opioid epidemic cost U.S. nearly $1.5 trillion in 2020*. U.S. Representative Don Beyer. (2022, September 28).
https://beyer.house.gov/news/documentsingle.aspx?DocumentID=5684#:~:text=Adapting%20a%20methodology%20used%20by,likely%20to%20continue%20to%20increase.

[16] United States Joint Economics Committee. (2019, April 24). *Losing our minds: Brain drain across the United States*.
https://www.jec.senate.gov/public/index.cfm/republicans/2019/4/losing-our-minds-brain-drain-across-the-united-states

[17] Simon, S. (2023, September 19). *Emergency broadband benefit program: Learn about providers and why it's now the ACP program.* Forbes.
https://www.forbes.com/home-improvement/internet/what-is-the-acp-program/#:~:text=When%20did%20the%20ACP%20program,enrollments%20on%20December%2031%2C%202021.

[18] The United States Government. (2024, February 6). *Fact sheet: As Affordable Connectivity Program hits milestone of providing affordable high-speed internet to 23 million households nationwide, Biden-Harris Administration calls on Congress to extend its funding*. The White House.
https://www.whitehouse.gov/briefing-room/statements-releases/2024/02/06/fact-sheet-as-affordable-connectivity-program-hits-milestone-of-providing-affordable-high-speed-internet-to-23-million-households-nationwide-biden-harris-administration-calls-on-congress-t/#:~:text=The%20Affordable%20Connectivity%20Program%2C%20enacted,on%20their%20monthly%20internet%20bills.

[19] Baumgartner, J. (2024, February 29). *Majority of ACP households will see service disruptions if program ends – FCC survey*. Light Reading.
https://www.lightreading.com/digital-divide/majority-of-acp-households-will-see-service-disruptions-if-program-ends-fcc-survey

[20] Thankachan, D. (2021, November 16). *US Business Internet isn't actually a Monopoly. Lightyear.* https://lightyear.ai/blogs/us-business-internet-isnt-actually-a-monopoly

[21] Schwantes, J. (2022, November 17). *Broadband Pricing: What Consumer Reports Learned From 22,000 Internet Bills.* Consumer Reports.
https://advocacy.consumerreports.org/wp-content/uploads/2022/11/FINAL.report-broadband.november-17-2022-2.pdf

[22] Campisi, N. (2023, May 26). *Millions of Americans are still missing out on broadband access and leaving money on the table-here's why.* Forbes.
https://www.forbes.com/advisor/personal-finance/millions-lack-broadband-access/

Appendix

I. Data Cleaning and Merging

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

health_df =
pd.read_csv('/content/drive/MyDrive/LK/MTFC23/Modeling/HealthData.csv')
broadband_df =
pd.read_csv('/content/drive/MyDrive/LK/MTFC23/Modeling/BroadbandData.csv')
price_df =
pd.read_csv('/content/drive/MyDrive/LK/MTFC23/Modeling/PriceData.csv')
provider1_df =
pd.read_csv('/content/drive/MyDrive/LK/MTFC23/Modeling/ProviderData.csv')
provider2_df = provider_df =
pd.read_csv('/content/drive/MyDrive/LK/MTFC23/Modeling/ProviderIDData.csv')

provider1_df = provider1_df[provider1_df['geography_type'] == 'County']
provider1_df = provider1_df[provider1_df['data_type'] == 'Fixed Broadband']
provider1_df = provider1_df.drop(columns=['geography_type', 'geography_desc',
'bus_iv_pct', 'data_type'])
provider2_df = provider2_df.drop(columns='frn')

provider_df = pd.merge(provider1_df, provider2_df, on='provider_id',
how='left')
provider_df = provider_df.drop_duplicates()

provider_df['state_abbr'] = provider_df['geography_desc_full'].str[-2:]

price_df = price_df[price_df['Download Bandwidth Mbps'] >= 25]
price_df = price_df[price_df['Upload Bandwidth Mbps'] >= 3]
price_df = price_df.drop(columns=['Weight', 'Usage Allowance GB', 'Other
Mandatory Charge', 'Surcharge', 'Total Charge', 'Total Charge Spread', 'Model
Data'])

def categorize_speed(row):
    if row['Download Bandwidth Mbps'] < 50:
        return 'Low Speed'
    elif row['Download Bandwidth Mbps'] < 250:
        return 'Medium Speed'
    else:
        return 'High Speed'

price_df['Speed Category'] = price_df.apply(categorize_speed, axis=1)
```

```
aggregated_df = price_df.groupby(['Provider', 'State', 'Technology', 'Speed
Category'])['Monthly Charge'].mean().reset_index()
aggregated_df.rename(columns={'Monthly Charge': 'Average Monthly Charge'},
inplace=True)
aggregated_df = price_df.groupby(['Provider', 'State', 'Technology', 'Speed
Category'])['Monthly Charge'].mean().reset_index()
aggregated_df.rename(columns={'Monthly Charge': 'Average Monthly Charge'},
inplace=True)

pivot_df = aggregated_df.pivot_table(index=['Provider', 'State', 'Technology'],
columns='Speed Category', values='Average Monthly Charge',
aggfunc='first').reset_index()

pivot_df.columns.name = None
pivot_df = pivot_df.rename(columns={'High Speed': 'High Speed Price', 'Medium
Speed': 'Medium Speed Price', Low Speed': 'Low Speed Price'})

pivot_df['State'] = pivot_df['State'].map(state_abbrev)

unique_geography_ids = provider_df['geography_id'].unique()

coverage_df = pd.DataFrame({'geography_id': unique_geography_ids, 'Fiber':
np.zeros(len(unique_geography_ids)), 'Cable':
np.zeros(len(unique_geography_ids)), 'Wireless':
np.zeros(len(unique_geography_ids)), 'Copper':
np.zeros(len(unique_geography_ids))})

low_cost_threshold = 50

tech_map = {'FTTH': 'Fiber', 'Cable': 'Cable', ‘DSL': 'Copper', 'Fixed
wireless': 'Wireless'}

for index, row in pivot_df.iterrows():
    if pd.notnull(row['Low Speed Price']) and row['Low Speed Price'] <
low_cost_threshold:
        service_type = tech_map.get(row['Technology'], None)
        if service_type:
            matches = provider_df[provider_df['state_abbr'] == row['State']]

            for _, match in matches.iterrows():
                geography_id = match['geography_id']
                if geography_id in coverage_df['geography_id'].values:
                    coverage_df.loc[coverage_df['geography_id'] ==
geography_id, service_type] += match['res_st_pct']

health_df['FIPS'] = health_df['FIPS'].astype(str)
health_df['FIPS']= [a.lstrip('0') for a in health_df['FIPS']]

broadband_df = broadband_df[broadband_df['geography_type'] == 'County']
```

```
broadband_df = broadband_df[broadband_df['total_units'] != 0]
broadband_df = broadband_df[broadband_df['area_data_type'] == 'Total']

broadband_df['geography_id'] = broadband_df['geography_id'].astype(str)
broadband_df['geography_id'] = [a.lstrip('0') for a in
broadband_df['geography_id']]

broadband_df.rename(columns={'t1_s3_r':'copper_25_3'}, inplace=True)
broadband_df.rename(columns={'t1_s4_r':'copper_100_20'}, inplace=True)
broadband_df.rename(columns={'t1_s5_r':'copper_250_25'}, inplace=True)
broadband_df.rename(columns={'t2_s3_r':'cable_25_3'}, inplace=True)
broadband_df.rename(columns={'t2_s4_r':'cable_100_20'}, inplace=True)
broadband_df.rename(columns={'t2_s5_r':'cable_250_25'}, inplace=True)
broadband_df.rename(columns={'t3_s3_r':'fiber_25_3'}, inplace=True)
broadband_df.rename(columns={'t3_s4_r':'fiber_100_20'}, inplace=True)
broadband_df.rename(columns={'t3_s5_r':'fiber_250_25'}, inplace=True)
broadband_df.rename(columns={'t4_s3_r':'gso_25_3'}, inplace=True)
broadband_df.rename(columns={'t4_s4_r':'gso_100_20'}, inplace=True)
broadband_df.rename(columns={'t4_s5_r':'gso_250_25'}, inplace=True)
broadband_df.rename(columns={'t5_s3_r':'ngso_25_3'}, inplace=True)
broadband_df.rename(columns={'t5_s4_r':'ngso_100_20'}, inplace=True)
broadband_df.rename(columns={'t5_s5_r':'ngso_250_25'}, inplace=True)
broadband_df.rename(columns={'t6_s3_r':'unlicensedwireless_25_3'},
inplace=True)
broadband_df.rename(columns={'t6_s4_r':'unlicensedwireless_100_20'},
inplace=True)
broadband_df.rename(columns={'t6_s5_r':'unlicensedwireless_250_25'},
inplace=True)
broadband_df.rename(columns={'t7_s3_r':'licensedwireless_25_3'}, inplace=True)
broadband_df.rename(columns={'t7_s4_r':'licensedwireless_100_20'},
inplace=True)
broadband_df.rename(columns={'t7_s5_r':'licensedwireless_250_25'},
inplace=True)
broadband_df.rename(columns={'t8_s3_r':'other_25_3'}, inplace=True)
broadband_df.rename(columns={'t8_s4_r':'other_100_20'}, inplace=True)
broadband_df.rename(columns={'t8_s5_r':'other_250_25'}, inplace=True)
broadband_df.rename(columns={'t1_2_3_4_5_6_7_8_s3_r':'any_25_3'}, inplace=True)
broadband_df.rename(columns={'t1_2_3_4_5_6_7_8_s4_r':'any_100_20'},
inplace=True)
broadband_df.rename(columns={'t1_2_3_4_5_6_7_8_s5_r':'any_250_25'},
inplace=True)
broadband_df.rename(columns={'mobilebb_3g_area_st_pct':'mobile_3g_outdoor'},
inplace=True)
broadband_df.rename(columns={'mobilebb_3g_area_iv_pct':'mobile_3g_vehicle'},
inplace=True)
broadband_df.rename(columns={'mobilebb_5g_spd2_area_st_pct':'mobile_5g_outdoor'
}, inplace=True)
broadband_df.rename(columns={'mobilebb_5g_spd2_area_iv_pct':'mobile_5g_vehicle'
}, inplace=True)
```

```
olumns_to_keep = ['geography_id', 'geography_desc_full']

prefixes = ['unlicensedwireless', 'licensedwireless', 'copper', 'cable',
'fiber']
for col in broadband_df.columns:
    if col.startswith(tuple(prefixes)):
        columns_to_keep.append(col)

broadfilt_df = broadband_df[columns_to_keep]

broadfilt_df['wireless_25_3'] = broadfilt_df[['unlicensedwireless_25_3',
'licensedwireless_25_3']].max(axis=1)
broadfilt_df['wireless_100_20'] = broadfilt_df[['unlicensedwireless_100_20',
'licensedwireless_100_20']].max(axis=1)
broadfilt_df['wireless_250_25'] = broadfilt_df[['unlicensedwireless_250_25',
'licensedwireless_250_25']].max(axis=1)

broadfilt_df.drop(columns=["unlicensedwireless_25_3",
"unlicensedwireless_100_20", "unlicensedwireless_250_25",
"licensedwireless_25_3", "licensedwireless_100_20", "licensedwireless_250_25"
], inplace=True)

broadfilt_df = pd.merge(broadfilt_df, coverage_df, how="inner",
on='geography_id')

broadfilt_df.rename(columns={"Cable": "cable_low_speed_coverage", "Fiber":
"fiber_low_speed_coverage","Copper": "copper_low_speed_coverage", "Wireless":
"wireless_low_speed_coverage"}, inplace=True)

columns_to_keep = ['FIPS','Life Expectancy', 'Child Mortality Rate', '% Adults
with Diabetes', 'HIV Prevalence Rate', '% Food Insecure', 'Drug Overdose
Mortality Rate','% Uninsured Adults', '% Uninsured Children', '% Households
with Broadband Access']

health_df = health_df[columns_to_keep]

final_df = pd.merge(broadfilt_df, health_df, left_on='geography_id',
right_on='FIPS', how="inner")
final_df = final_df.drop(columns='FIPS')
final_df.head()

final_df.to_csv("/content/drive/MyDrive/LK/MTFC23/Modeling/Data2.csv",
index=False)
```

II. Ordinary Least Squares Regression

```
import numpy as np
```

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('/content/drive/MyDrive/LK/MTFC23/Modeling/Data2.csv')

for column in df.columns:
    if df[column].dtype in ['float64', 'int64']:
        df[column].fillna(df[column].mean(), inplace=True)

corr = df.corr()

# Heat Map
plt.figure(figsize=(10, 8))
sns.heatmap(corr, annot=False, fmt=".2f", cmap='coolwarm', square=True,
linewidths=.5)
plt.title('Correlation Heatmap')
plt.show()

import statsmodels.api as sm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

predictors = ['copper_25_3', 'copper_100_20', 'copper_250_25', 'cable_25_3',
'cable_100_20', 'cable_250_25', 'fiber_25_3', 'fiber_100_20', 'fiber_250_25',
'wireless_25_3', 'wireless_250_25', 'wireless_100_20',
'fiber_low_speed_coverage', 'cable_low_speed_coverage',
'wireless_low_speed_coverage', 'copper_low_speed_coverage', '% Households with
Broadband Access']

# Target variables list
targets = ['Life Expectancy', 'Child Mortality Rate', '% Adults with Diabetes',
'HIV Prevalence Rate',
           '% Food Insecure', 'Drug Overdose Mortality Rate', '% Uninsured
Adults', '% Uninsured Children']

# Loop through each target variable, perform regression, and plot the
regression line
for target in targets:
    X = df[predictors]
    y = df[target]
    X = sm.add_constant(X)  # Adding a constant for the intercept
    model = sm.OLS(y, X, missing='drop').fit()  # 'missing=drop' automatically
drops any NaNs

    # Predicting the y values
    df[f'predicted_{target}'] = model.predict(X)
```

```
# Plotting actual vs. predicted values with a regression line
plt.figure(figsize=(10, 6))
sns.regplot(x=y, y=df[f'predicted_{target}'], line_kws={"color": "red"})
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title(f'Actual vs. Predicted {target} with Line of Best Fit')
plt.show()

# Printing the results
print(f"Results for {target}:")
print(model.summary())
```