The affordability of safety: COVID and Affordable Housing in Gwinnett County

Caleb Case

DATA 512 Final Assignment

Dec 12th, 2021

# Introduction

The COVID-19 pandemic has highlighted and exacerbated many of the underlying inequalities and injustices in America over the past two years. Key among those is the unavailability and unaffordability of housing. The amount of people paying rent paycheck to paycheck became apparent as soon as a drop in unemployment occurred, and national eviction moratoriums had to be put in place to prevent mass eviction and homelessness.

Less clear has been the impact of housing on COVID-19. When households must be larger and denser to afford rent, quarantining and social distancing becomes more difficult. When workers are unable to lose income without becoming unable to pay rent, they may have to make tough decisions after a potential exposure or showing symptoms. If more affordable housing does help slow the spread of disease, this should inform policy makers and add one more social benefit to the creation of more affordable living situations.

This paper examines the impacts of affordable housing availability on the spread of COVID-19, focusing specifically on Gwinnett County in Georgia as it compares to the rest of the Atlanta metro area. While there is insufficient evidence to establish a relationship between low-income housing units and COVID spread, there is indication that larger household size may increase exposure rates. This could warrant further study into the relation of policy, household size and disease spread.

# Background

## Affordable Housing

There is significant study relating to housing, evictions, and COVID-19. However, the majority of the literature focuses on how COVID has impacted housing conditions, and less common are studies of how COVID transmission has been impacted by housing.

There is some indication that “Counties with a higher percentage of households with poor housing had higher incidence of, and mortality associated with, COVID-19.” (Ahmad, 2020) In this study, poor housing included overcrowding and high housing cost, but also incomplete facilities. Affordable housing is included in this definition and correlated with both COVID spread and mortality. However, it is not separated form other characteristics such as incomplete kitchen and toilet facilities, and the effect of affordable housing specifically is not clear.

Eviction, one outcome of a lack of affordable housing especially when combined with economic disruptions, has been found to increase the spread of COVID-19. (Benfer, 2021) Even more concerningly, “Disproportionate rates of both COVID-19 and eviction in communities of color compound negative health effects” (Benfer, 2021) However, the authors suggest eviction moratoriums and other supportive measures as a solution, which would likely mitigate the negative effects of evictions, but would not necessarily increase the supply of affordable housing, and would not help to treat any disparities caused by the lack thereof.

This research supports this paper’s hypothesis that a lack of affordable housing could increase COVID transmission rates. It also adds additional reasons affordable housing would aid in the pandemic, such as mitigating the costs of eviction and the disproportionate effects on at-risk groups. However, it also suggests that there may be some more specific measures that are truly important, such as household size and living conditions, that affordable housing just serves as an indicator of.

## COVID Transmission Models

A large amount of work has been done in attempts to model COVID transmission. One such robust model is presented by Yang and Wang, (Yang & Wang, 2020). They model COVID transmission in a series of differential equations, breaking populations into exposed, infected, and hospitalized people and tracking transmission rates and transition rates between each group and the environment. However, they still model actual transmission between groups as a single rate or interaction, without the differentiation between exposure and actual transmission that this paper’s model includes.

They also conclude that their “approach based on different transmission rates in different time periods has a better performance than that based on the standard approach of using uniform, constant transmission rates throughout the entire time domain.” (Yang & Wang, 2020) This paper’s model borrows this same method, of refitting the rates on regular intervals to capture changes in behavior that happen throughout the progression of the pandemic.

# Methodology

## Data Collection and Processing

Multiple datasets were combined for this analysis. The full list of data sources can be found in the data section below. Daily case data was gathered from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, representing daily cumulative case counts by county.

Mask compliance data was used from a New York Times survey in July 2020. The survey asked, “How often do you wear a mask in public when you expect to be within six feet of another person?” and participants responded one of “Never”, “Rarely”, “Sometimes”, “Frequently”, “Always”. This data was assigned values of 0%, 25%, 50%, 75% and 100% compliance respectively to consolidate it into a single estimated rate of mask usage.

Low Income housing data came from the HUD National Low Income Housing Tax Credit (LIHTC) Database. This data relates to units designated as Low Income by the department of Housing and Urban Development, which landlords receive a tax credit for maintaining. Specifically, the number of such tax-credit low-income units in each county was used.

Demographic data came from the US Census Bureau. The mean household income, household size and total population was used form this 2019 census data.

Data was joined together at the county level, and a series of additional calculations was performed on the case count data. New Cases were calculated as the different in total case count form day to day. This displayed a large amount of weekly seasonality, likely due to reporting and testing patterns, so a rolling seven-day average was used for further calculations. The number of estimated active cases was calculated as the number of new cases in the past 14 days, representing an average 14-day active period of infections in line with CDC guidance. Finally, the estimated vulnerable population, those still able to be infected, was taken as the total population minus the total case count. This analysis assumes that people cannot get COVID twice.

## Transmission Model

A simplified transmission model was used that allowed for two free parameters: transmission chance and exposure rate. Transmission chance represents the average chance of a transmission happening when an infected person and vulnerable person come into contact. This would take into account effects such as natural immunity, mask wearing, ventilation, and other factors that would make it more or less likely for an interaction between two people to result in a transmission. The other factor is the exposure rate, representing the number of people per day, infected or not, that a vulnerable person might be exposed to. This would take into account any efforts at social distancing, adherence to gathering bans, workplace attendance, and any other factors that would change the daily number of exposures.

Using these two parameters, the number of new cases is modeled. The exposure rate and proportion of the current population that is actively infected is used to model the expected number of exposures to an infected person that a vulnerable person would experience in a day. The transmission chance is then used to determine the chance that aa transmission occurs, by taking the additive inverse of the chance that no transmissions occur. This gives us the model equation

Where and are new and active cases, and are vulnerable and total population, is the transmission rate and is the exposure rate.

This model is fit to the new case data for each county and each month, using a gradient descent method to find the lease-squares fit. The parameters are bound to their domains, being [0,1] for the transmission chance and greater than 0 for the exposure rate.

Note that this model is used in an interpretive rather than predictive way, with the purpose being to extract the parameters for analysis, not to predict future cases. Because of this, the model is only run at a one-step prediction: the actual active cases are always used as the input.

## Parameter Analysis

The application of the model above provides two data points per county per month: the transmission chance and exposure rate. The following analysis is then focused on searching for links between these parameters and various demographic and housing data.

The data set consists of one data point per county per month, for a total of 220 rows with seven features each. Two features are treated as dependent: transmission chance and exposure rate. The other five features are used as independent variables to assess their relations to the dependent variables in two separate regressions.

Some collinearity is seen within the features (Figure 1). Low Income units per Household is the most impacted feature, with a 0.59 Pearson Correlation Coefficient with Mask Compliance and a -0.62 correlation with Mean Income.

A primary concern was that any false result that could indicate a negative relationship between affordable housing and covid spread could unfairly stigmatize those in low-income situations and could become a false reason to avoid constructing more affordable supply. Some values that could possible cause this spurious relationship were included, such as mean income and household size, but the risk is still prevalent and any results should not be taken to imply causation.

Graphical user interface

Description automatically generated

Figure 1

### Transmission Chance

Transmission Chance is evaluated with a Generalized Linear Model using a Negative Binomial distribution and a log link function. This method was used to better fit the domain of the transmission chance, which is bound from 0 to 1. While the Negative Binomial model is positive, it is not bound at the top by 1. However, in evaluating results, no model prediction exceeded 0.11, so the model may still be a reasonable fit for this data set.

The residuals displayed strong deviance from normal for the edge cases (Figure 2), so Robust Covariance was used for determining coefficient significance.

Chart, line chart

Description automatically generated

Figure 2

### Exposure Rate

Exposure Rate was evaluated with a Generalized Linear Model using a Poisson distribution and a log link function. This method was used to better fit the domain of the exposure rate, which is positive and unbounded. Residuals were distinctly non-normal (Figure 3), so robust covariance was used for evaluating parameter significance.

Chart, line chart

Description automatically generated

Figure 3

# Findings

## Transmission Model

The transmission model proved to be a reasonable method of modeling the case counts. On Gwinnett County it found an R-Squared value of 0.897, and for all of Atlanta found an even higher value of 0.958. These indicate that the model accounts for around or over 90% of the variance in new case counts. The average error was 18% of the mean new cases for Gwinnett and 11% for the entire Atlanta area.

The extracted parameters fit a few commonsense checks. Both transmission and exposure have their highest values at the start of the pandemic, before any measures were put into place. Exposure rate also shows more change than transmission chance. This fits expectations for Georgia; with no mask mandates in place, but multiple periods of business closure, it seems logical that the exposure rate would be more easily modified than the transmission chance.

The model aids in highlighting some if the differences between the new case curves in Atlanta and Gwinnett. Overall, Gwinnett generally followed the same patterns as the larger metro area. However, it can be seen to have a shallower beginning to the second wave in Fall 2020, and an overall lower peak in this wave. Chart, histogram

Description automatically generated

Chart

Description automatically generated

Figure 4

The difference in curve is visible in the parameters, but this highlights how sensitive the model curve is to these changes. In October, Atlanta had just 0.08 percentage points higher transmission rate and 0.09 more daily exposures, yet saw 15% more new cases per capita than Gwinnett. Because of the exponential nature of the transmission when the parameters get high, small changes in parameters are associated with large changes in case count and curve steepness.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Figure 5

## Parameter Analysis

### Transmission Chance

The demographic data was evaluated as significant in the transmission chance regression by testing at the p < 0.05 level, indicating that under the null hypothesis of no relation, these results would occur less than 5% of the time. At this level, three features were determined to be related to transmission chance: Mask Compliance, Household Size, and Income.

Text

Description automatically generated

Table 1

Its important to note the meaning of the coefficients. Because this was done using a Negative Binomial model, coefficients can be interpreted as changes in the Odds Ratio of the Transmission Chance. In this setting, the Odds Ratio would represent the amount of transmission occurring divided by the amount not occurring in a set of exposures. The mean transmission rate in all samples was approximately 3%, meaning the average odds ratio is .031. At small values like this, the odds ratio changes similarly to the likelihood, so the exponentiation of coefficients can be roughly interpreted as changes in likelihood.

The findings that Mask Compliance and Income are negatively correlated with transmission chance fit the experiment hypothesis. However, the result that larger household size is also linked to lower transmission chance is unexpected.

#### Mask Compliance

Mask compliance was found significant at the p ~ 0 level. The coefficient of -2.784 translates to a change of .06%. A 1% increase in mask usage would relate to a ~ .094% decrease in transmission chance.

#### Household Size

Household Size was found significant at the p ~ 0 level. The coefficient of –0.3954 translates to a change of 0.67%. A 0.1 person increase in average household size would relate to a ~0.033% decrease in transmission chance.

#### Mean Income

Mean Income was found significant at the p = 0.018 level. The coefficient of -2.15e-6 translates to a change of 0.999998%. A $10,000 increase in mask usage would relate to a ~ 0.02% decrease in transmission chance.

### Exposure Rate

The demographic data was evaluated as significant in the exposure rate regression by testing at the p < 0.05 level, indicating that under the null hypothesis of no relation, these results would occur less than 5% of the time. At this level, one feature was determined to be related to exposure rate: Active Cases. Additionally, the Household Size was significant at the p <= 0.1 level.

In the Poisson regression, the exponent of the coefficients can be interpreted as the percentage change in the dependent variable for each percent change in the independent variable. At the mean Exposure Rate of 2.8 people per day, a 1% change is approximately equivalent to 0.28 people per day.

#### Active Cases

Active Cases was found significant at the p = 0.005 level. The coefficient of -2.74e-5 translates to a change of 0.000027%. An additional 1000 active cases would reduce the Exposure Rate by around 0.000756 people per day. So while the relationship was found to be significant, the effect is minimal and generally not material in this context.

#### Household Size

Household Size was found significant at the p = 0.098 level. The coefficient of 0.2041 translates to a change of 22.6%. An additional 1 average person per household would increase the Exposure Rate by around 6.3 people per day. While the significance of this relationship is less certain, the impact is dramatic.

## Impact on Gwinnett

Gwinnett County has so far experienced materially higher rates of COVID than the greater Atlanta area. Is has 0.0036 more daily cases per capita on average, and at the last available date of October 31st, 2021, had 5.988 more total cases per capita. Based on the model, this is due to the high exposure rate in Gwinnett. The county had 0.0063 more daily exposures on average, but 0.017 percentage points lower average transmission chance.

Overall, a large amount of the difference in Exposure Rate between Gwinnett and Atlanta was attributable to the observed measures, but a much smaller proportion of the transmission rate was. Gwinnett has 3% points higher mask compliance, $6,243 less in mean income, and 0.29 more average persons per household. These variables and the difference in active cases explained 92.6% of the difference in average exposure rate, but only 17.6% of the difference in transmission chance.

The observed higher household density and lower income in Gwinnett relate to its higher exposure rate and ultimately high case count, even though a high mask compliance was connected to the lower transmission chances that Gwinnett experienced.

# Discussion

The analysis was structured intentionally to facilitate insight into human behavior and attempt to arrive at verifiable and actionable results. With the goal of creating a product that users could evaluate against their own background knowledge, the model was created to produce tangible parameters such as transmission chance and exposure rate, that could be explained in simple language. The focus of the analysis was geared towards providing insight to policy makers

The results are not entirely as expected but do suggest some interesting avenues of further study. The transmission model appears to be effective and could lead to more behaviorally focused analysis of the pandemic. However, some verification of the parameter interpretations would be required. The analysis of low-income housing on the model parameters failed to find any relation between low-income units and either transmission or exposure, but did suggest that household size, a related measurement, could impact both.

## Transmission Model

The transmission model was seen to create strong fits to the data, which is an indication that it may be a useful model. The value in this model is that it reduces case growth to more human factors, where transmission chance and especially exposure rate are more easily attributable to human behavior than other methods of measuring infection spread, such as the generic environment to human rate used in (Yang & Wang, 2020).

There are two primary benefits to this. The first is that this focuses results on more human behaviors and adds granularity. In this study, Mask Compliance was found to have a strong negative correlation with transmission chance, but no relation to exposure rate. This relation is supported by the real-world interpretation: its logical that mask usage could lower transmission chances and would be harder to explain why it could impact exposure rates. It can also help exposure relationships that may have been hidden before. In this study, household size was found to have a negative correlation with transmission chance and a positive one with exposure rate. It’s possible that when looking at just case growth, the opposite effects would cancel out and no relationship would be detected.

The second benefit is that this model could help to inform the best methods to reduce spread. Depending on the current rates, number of active cases and ratio of vulnerable population, changes in transmission chance and exposure rate will have different effects on new case count. In some situations, reducing exposure rate may be very impactful, while in others it may be less useful than reducing transmission chance. This could help policy makers determine the best usage of resources, such as when to focus on mask compliance or when to focus on reducing gathering sizes.

The most beneficial path of further research would be to establish stronger connections between the model parameters and the real-world interactions they represent. For example, transmission chance is assigned on a purely theoretical basis in this analysis. But if it could be compared to some empirical measures of transmission chance, the power of the model in terms of its actual parameters, and not just as a predictor of cases, could be evaluated. Further research could also go towards evaluating causality in some of the discovered relationships and evaluating the usefulness of the model with longer prediction windows.

## Affordable Housing

The analysis failed to find any relationship between the number of low-income units available and either the transmission chance or exposure rate. However, household size was found to be related to both transmission chance and possibly exposure rate, and median income was linked to transmission chance.

The relation to income is unsurprising. Income provides myriad of possible effect mechanisms, from better access to healthcare, larger homes, better education, etc. There is also a notable negative correlation between mean-income and low-income housing availability. This warrants further research into the relation and direction of that relationship.

Household size was found to correlate negatively with transmission chance and positively with exposure rate. The relationship to transmission chance is unexpected. Some possible mechanisms may be correlations with unmeasured demographics, such as larger households generally having more children who are less susceptible to infection. A possible direct mechanism could be mental and physical health benefits from companionship. If the mechanism were to be determined through further research, it could provide another lever for policy to pull to counteract covid spread.

The positive relation to exposure rate is more logical. More people living together naturally leads to more exposures to each other, and likely leads to more exposure outside of the household if each person has their own network of contacts. This could inform decisions around regulation and housing development if further research confirms the effects. Encouraging smaller households through more affordable options, small constructions and subsidized family planning could all provide additional benefit through the effect on pandemic risk.

While no relation to affordable housing was found in this study, relationships to income and housing characteristics were both found. There are enough potential limitations, discussed in the next section, to warrant further research.

# Limitations

## Active Case Assumption

Possibly the largest assumption made in the study is that cases remain active for 14 days, and that this value is constant throughout the pandemic. This number was chosen based on CDC guidance and is likely conservative due to this. It’s also possible that the period of cases being active may change based on medical technology, hospitalization rates and demographics, none of which were included in the analysis. A further avenue of research would be to allow the active case period to be a further variable in the model, instead of an assumption

## Parameter Transition Points

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated Another weakness of the model was the arbitrary determination to re-evaluate every month. This was chosen for convenience without any major justification and led to some notable spikes in modeled cases at the start of some months (Figure 6 and 7). A better approach would be to evaluate the best points individually, or to estimate a new parameter for each day using a rolling window of values.

Figure 6

Figure 7

Figure 8

## Data

Some data points were not completely in line with the intended purpose. The number of Low-Income housing units was based on the HUD Tax Credit program. While these are certainly good representations of affordable housing, it’s a narrow definition. Any housing taking advantage of other tax cuts or just kept low due to market rates is not included, and it’s unclear how the HUD LI Unit count may relate to this. If HUD Low-Income units are not a good representation of affordable housing availability, then the analysis is flawed.

The Mean Income is also not the best metric for county wealth. The mean income for Atlanta was nearly $100,000 per year, but it seems probable that most households in Atlanta earn less than this and it’s inflated by a small number of super-earners. Median income or a spread of quartile incomes may be a better metric.

## Regression Assumptions

Some of the assumptions underlying the analysis regressions are not clearly met or even likely broken. Both the Negative Binomial and Poisson regression assume that the data is independent, but because the data points were a set of time-series data, it’s highly possible that many of them were dependent on the previous month’s values. Both regressions also assume that the dependent variable is integer values, which is not true in the used context. They were chosen because they more closely fit the domain than a linear regression, but at the cost of breaking this assumption.

## Conclusions and Causality

No causality was determined in this experimental setup. However, due to the nature of the data, and the fact that all the demographic data was gathered before COVID existed, it is natural for the reader and author to assume causation from demographic data to the pandemic parameters. However, even if causation as established, the best policy solution is not readily apparent from the result. If higher density households create more COVID spread, does that mean we should attempt to split large households? Or does it mean that they represent at-risk populations who should receive additional resources? Opposite policy conclusions could be drawn from the same results.

## Timing

Due to data gathering methods, the timing of various datasets is not consistent. Census data and HUD data are form 2019, while case count and mask compliance data begin in 2020. If there were major demographic changes in 2019 and 2020, the data may not be accurately compared.

## Scope

The data was restricted to the Atlanta area to allow a focused case-study on Gwinnett County. While this allowed for a detailed demonstration of the impacts of the analyzed variables, it also limited the dataset and variety of data available. The pandemic curve in Gwinnett was largely similar to all of Atlanta, so the observed variation in case counts, modeled parameters and demographic variables were all relatively small. Using national data could strengthen the analysis and provide more interesting comparisons between more varied areas.

# Conclusion

The goals of this project were to create a model of COVID spread that used explainable behavioral parameters, and to analyze the impact of affordable housing availability on these parameters. Data was gathered from the New York Times, US Census, US HUD, and John Hopkins University to help follow this plan. The project focused on Gwinnett County, using the larger Atlanta metro area for comparison.

The model generated had some noticeable shortcomings but ultimately appeared to be a good fit, closely modeling actual case-counts and providing results that met basic logical requirements. The resulting Transmission Chance and Exposure Rate parameters from the model were deemed acceptable to use in the analysis portion of the project, and the model itself holds the possibility of becoming a valuable behavioral tool with some enhancements and verifications.

The hypothesis was that the availability of affordable housing would slow COVID spread, especially exposure rates as it helped people stay off the streets and experience less pressure to work. However, no link between the prevalence of HUD Tax-Credit Low Income units and COVID spread was found. There was some evidence that mean income is negatively correlated to transmission chance, and that household size is negatively correlated to transmission and positively correlated to exposure rate.

The existence of the found relationships between income and housing characteristics, as well as the shortcomings inherent in the gathered data, leave the question open on the impact of affordable housing on COVID spread. Further research is warranted.

# Works Cited

Ahmad, K. e. (2020). Association of poor housing conditions with COVID-19 incidence and mortality across US counties. *Plos One*. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241327

Benfer, E. A. (2021). Eviction, Health Inequity, and the Spread of COVID-19: Housing Policy as a Primary Pandemic Mitigation Strategy. *J Urban Health*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7790520/

Yang, C., & Wang, J. (2020). Modeling the transmission of COVID-19 in the US – A case study. *Science Direct*. Retrieved from https://www.sciencedirect.com/science/article/pii/S246804272030110X

# Data Sources

## Case Data

Case data was obtained from a [Kaggle project](https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university/version/377?select=RAW_us_confirmed_cases.csv) that provides access to the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. It was accessed on 2021-11-1. (CC BY 4.0 license)

## Mask Compliance Data

A [New York Times and Dynata Estimate](https://github.com/nytimes/covid-19-data/tree/master/mask-use) based on survey data from July 2nd to 14th, 2020. The data is based on 250,000 online survey responses and contains estimates down to the county level of mask usage while in public. Specifically, each participant was asked: How often do you wear a mask in public when you expect to be within six feet of another person? This was accessed on 2021-11-01. [License](https://github.com/nytimes/covid-19-data/blob/master/LICENSE)

## Income Data

The US Census Bureau [Data Table Tool](https://data.census.gov/cedsci/table?q=Income&g=0400000US13%240500000&tid=ACSST1Y2019.S1901&hidePreview=true&tp=true) with 2019 census data on household income levels. Accessed on 11-28-2021 [License](https://www.census.gov/data/software/x13as/disclaimer.html)

## Population Data

The US Census Bureau [Data Table Tool](https://data.census.gov/cedsci/table?q=Population&g=0400000US13%240500000&tid=ACSDP1Y2019.DP05&hidePreview=true&tp=true) with 2019 census data on population. Accessed on 11-28-2021 [License](https://www.census.gov/data/software/x13as/disclaimer.html)

## Low Income Unit Data

The Office of Policy Development and Research in the Department of Housing offers this [Low-Income Housing Tax Credit (LIHTC](https://www.huduser.gov/portal/datasets/lihtc/property.html#data)) dataset. No license is found, but “HUD hopes to enable researchers to learn more about the effects of the tax credit program.”