

Key Predictors of COVID-19 Seropositivity among Adults in New York City; Community Health Survey 2020

Apeksha Mewani^{1,*}, Vincent Jones II², Erin Jacques²

¹Health Equity, Administration, and Technology, Lehman College, New York City, USA

²Health and Human Performance, York College, New York City, USA

*Corresponding author: apeksha.mewani@lehman.cuny.edu

Received February 13, 2025; Revised April 29, 2025; Accepted May 08, 2025

Abstract This study identifies common key predictors of COVID-19 seropositivity by comparing various regression models using a hierarchical regression method among a sample of New York City (NYC) adults, based on the analysis of the New York City Community Health Survey (NYC CHS) 2020 dataset for this. An exploratory approach is used to understand the social, environmental, and individual determinants of health in NYC's population at the peak of the pandemic and the effects on COVID-19 seropositivity. Hierarchical logistic regression was carried out on a sample of 928 participants. The findings suggest that age (65-75 years), race (Black and Hispanic), and the birthplace (US) were significant to the hierarchical regression when only socioeconomic factors were considered. Including health behaviors, tobacco usage behaviors, and physical activity became s. In the full model, BMI, asthma prevalence, and suicidal thoughts significantly correlated with COVID-19 seropositivity. The findings are consistent with public health literature highlighting the importance of healthy behaviors and public health efforts in maintaining overall health and immunity.

Keywords: COVID seropositivity, Social Determinants of Health, Health Education Program, Community Health Surve

Cite This Article: Apeksha Mewani, Vincent Jones II, and Erin Jacques, "Key Predictors of COVID-19 Seropositivity among Adults in the New York City Community Health Survey 2020." *American Journal of Public Health Research*, vol. 13, no. 3 (2025): 103-116. doi: 10.12691/ajphr-13-3-2.

1. Introduction

The novel coronavirus was first identified in Wuhan, China, in December 2019, prompting the World Health Organization to declare a global health emergency by January 30, 2020 [1]. By May 1, 2022, an estimated 10.5 million children worldwide had lost parents or caregivers due to COVID-19, with around 7.5 million orphaned [2]. In the US, 204,000 children and teens lost caregivers [3], and 115,000 healthcare workers died from the virus [4]. The impact of the pandemic continues to affect vulnerable populations, especially those who experienced sensitive life stages [5].

Epidemiologic trends showed how vulnerable groups were negatively impacted during the pandemic. Daily wage workers faced significant challenges, with severe food insecurity, per a study concerning Bangladeshi households, who experienced considerable economic loss due to lockdowns, estimated at US\$64.2 million [6]. In Australia, vulnerable children and young people faced compounded health risks and barriers to accessing support [7]. In contrast, occupational risk disparities highlighted workers' plight in health, tourism, and retail industries [8].

In the US, the COVID-19 pandemic led to significant job losses, disproportionately affecting minorities [9] [10]. Losing jobs often means losing health insurance, creating barriers to treatment. Support programs for employment, childcare, and healthy food access are critical for vulnerable populations. Thus, socio-economic conditions and other determinants, which often influence health, significantly impacted COVID-19 morbidity and mortality, exacerbating health disparities in underserved populations [10]. These factors are crucial when addressing health inequities during emergencies, wherein access to healthcare is essential. Strategies like increasing insurance coverage can help individuals obtain necessary services. Despite initiatives like the Healthy People 2030, which aim to improve access to high-quality healthcare, 14% of US adults, particularly non-white, lower education and income adults, indicated reluctance to seek medical care for COVID-19 symptoms [11]. Economic instability, including unemployment, exacerbates health issues, and is linked to the occurrence of cardiovascular disease and diabetes, which increased the risk of contracting COVID-19 [12]. Stress too negatively affects immune function, especially in pregnancy, highlighting the need for a psychoneuroimmunology approach to address the stress-immune interaction.

Healthy People 2030 aims to enhance community safety and health through policy changes and interventions, including creating sidewalks and bike lanes to promote physical activity. Health literacy is crucial in preventing communicable diseases like COVID-19, helping individuals understand transmission and the importance of following health recommendations during pandemics [10]. The pandemic has highlighted the interconnection between economic, social, and health status. Rezaei et al. [13] emphasize the need for accurate data on health equity and social determinants of health (SDOH) to ensure transparency and accountability in addressing these issues, particularly during crises like COVID-19. SDOH factors, including water sanitation, food security, and access to healthcare, significantly impact health outcomes. Economic strategies like direct payments and food support can help vulnerable populations cope with disparities.

Biological vulnerabilities, like pre-existing conditions and age, further influence COVID-19 outcomes, with significant disparities observed among racial and ethnic minority groups [14]. The CDC reports higher hospitalization among African Americans as compared to their population proportion.

Robertson et al. [14] argue that public health interventions can inadvertently increase health disparities. For example, social distancing may have disproportionately affected certain racial and ethnic groups due to their employment situations and living conditions. Upon analyzing the SDOH's impact through indices focused on social distancing ability, susceptibility to COVID-19 complications, and healthcare access, the researchers suggest that effective public health strategies must consider SDOH to prevent exacerbating existing disparities in exposure and access to treatment.

1.1. Rationale for Selecting Variables in This Paper

The NYC CHS 2020 dataset, with over 200 variables, influenced the rationale for selecting variables in this research. Existing literature on the social, environmental, and individual determinants of health highlights gaps in analyses regarding disease seropositivity and health disparities among NYC adults during the peak of the COVID-19 pandemic. The study employs exploratory analyses aligned with relevant literature to address these gaps.

The variables are categorized into four groups: general health and individual, environmental, and social determinants. The key variables include access to health insurance, medical care over the past year, prescription medication use, prevalence of hypertension, diabetes, and other such diseases, mental health indices, substance use, living arrangements, vaccination status, physical activity, and demographics. COVID-19 seropositivity is the primary outcome measured.

Individual-level factors, such as tobacco and alcohol usage, physical activity and dietary habits, are important indicators of overall health and chronic disease risk and influence long-term health trajectories. Thus, monitoring them offers insights into public health trends. Environmental factors like the locale of residence, household size, and social determinants like education and access to healthcare will help assess health inequities.

Additionally, studying the prevalence of conditions like asthma and diabetes, vaccination rates and psychological distress, will elucidate population health characteristics.

COVID-19 seropositivity was chosen as the primary outcome due to the absence of severity data in the NYC CHS 2020 dataset. While seropositivity and severity are essential to measure disease burden, focusing on the former provides a more comprehensive understanding of the epidemic.

1.2. Frameworks Guiding this Research

The conceptual frameworks described below provide a robust foundation for identifying key predictors of COVID-19 seropositivity among New York City adults by integrating social, environmental, and individual determinants of health. Using the Community Health Survey (NYC CHS) 2020 dataset, the study employs a hierarchical regression method to compare various regression models to explore how these determinants interact and contribute to health outcomes. By adopting an exploratory approach, the research delves into the influences of factors such as access to healthcare, demographic characteristics, and individual health behaviors at the pandemic's peak. This link between conceptual frameworks and empirical analysis provides a comprehensive understanding of the complex influences on seropositivity, ultimately informing and improving public health interventions to address health disparities among NYC's residents.

Solar and Irwin [15] said, "the World Health Organization Commission on Social Determinants of Health published a conceptual framework for action on the social determinants of health" in 2010.

The National Academies of Sciences, Engineering, and Medicine [16], describes the conceptual framework succinctly below:

This framework [displayed below] is divided into structural and intermediary determinants. The structural determinants comprise the societal, economic, and political context in which a person is born and lives, which dictates one's socioeconomic position. One's socioeconomic position then sets the stage for the intermediary determinants (material circumstances, psychosocial circumstances, behavioral and biological factors, and the health system itself) and the likelihood of exposure to health-compromising conditions. Illnesses caused by poor living conditions can then return to the structural determinants if, for example, a person experiences a loss of employment or income, lowering their socioeconomic status [16].

The intermediary and structural factors were incorporated into the framework to bridge the gap between social cohesion and social capital, despite the minimal political gain. This was done to develop approaches to public health and the social determinants of health that were less politically inclined, as well as systems that would promote equity, making it easier for citizens and institutions to cooperate. The effects of these structural and intermediary determinants on equity in health and well-being are shown in the figure's final box to the right. These provide feedback into structural determinants and the positive, negative, or neutral effects

on subsequent generations.

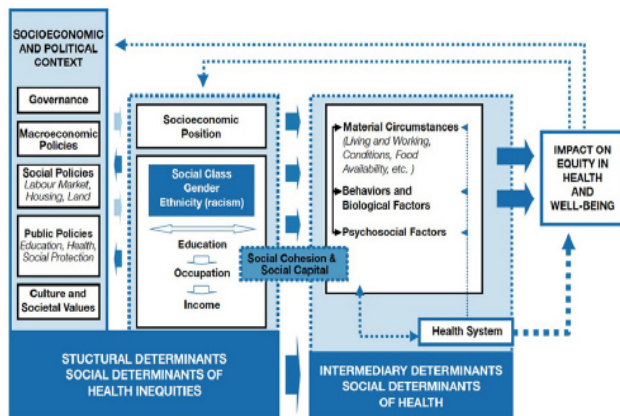


Figure 1. WHO Framework for Educating Health Professionals to Address the Social Determinants of Health

Frieden [17] developed a five-tier pyramid to improve public health (fig 2). The interventions with the highest potential to influence social determinants of health (including reduced poverty, increased education) are at the base of the pyramid. Interventions that are labor-intensive but would benefit many people are at the following levels: fluoridated water treatments, immunization, and vaccination programs. Next comes the direct clinical interventions that prevent specific conditions and significantly impact an individual's health, including cardiovascular diseases. Finally, health education is the most labor-intensive intervention with negligible impact on public health. Frieden [17] states that similar public health frameworks have been developed since 1994, emphasizing the effects of clinical health services and health care delivery. However, in contrast to these frameworks, Frieden's model relies on the social determinants of health.

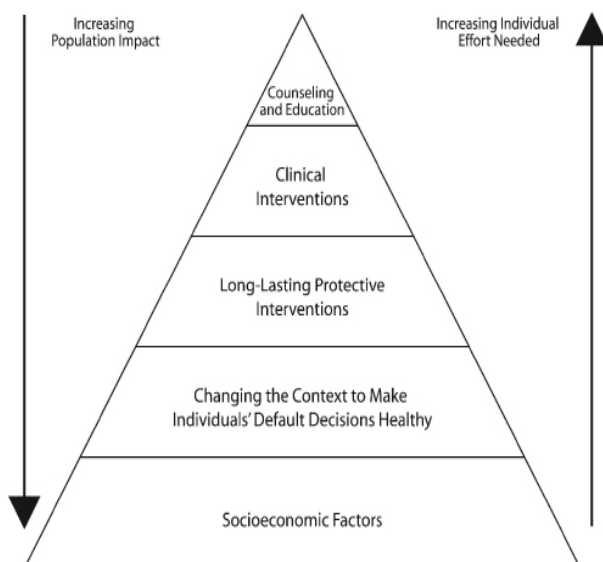


Figure 2. The Frieden Framework

2. Methods

As discussed above, determinants of health can be divided into three major categories: individual

characteristics and behaviors, physical environment, social and economic environment. The order of each variable's entry into the hierarchical analysis is based on the determinants of health theory. The first model includes the socio-economic factors; the second model comprises socio-economic and health behavior factors; the third model includes socio-economic, health behavior, and built environment factors; and the full model includes all the above along with existing co-morbidities and health conditions.

The variables of interest are: access to health insurance; access to a personal care provider; the ability to receive medical care (12 months); prescription medication use; the prevalence of diseases such as hypertension, asthma, and diabetes (either type 1 or 2); Kessler-6 index score for mental health; the prevalence of non-specific psychological distress; use of mental health treatments; tobacco use; alcohol use; home ownership; rent payment delays; the location of residence; flu vaccination status; physical activity engagement (30 days); consumption of fruits and vegetables; consumption of sugar-sweetened beverages; difficulty in performing daily activities; assisted device utilization; engagement in HIV test; the incidence of interpersonal violence; the incidence of suicidal thoughts; and lastly demographics such as age, race, orientation, gender, marital status, education levels, employment status, number of adults in the household, imputed poverty level, citizenship, and BMI. The outcome variable will be COVID-19 seropositivity.

The data for this study were collected from the New York City Community Health Survey. The New York City Department of Health and Mental Hygiene, a part of the Bureau of Epidemiology Services, conducts an annual telephonic survey, namely The New York City Community Health Survey (CHS). The annual cross-sectional study provides substantial insight into the health of New Yorkers with neighborhood, borough, and citywide estimations. The CHS uses a disproportionate stratified random sample to help assure geographic representativeness across the city [18]. "Participation goals are set for 42 United Hospital Fund neighborhoods, defined by contiguous zip codes" (p. 188).

The survey recruits approximately 10,000 randomly selected adults every year from all five boroughs in New York City (Manhattan, Brooklyn, Queens, Bronx, and Staten Island). The data is collected from participants via landline and mobile phone through a computer-assisted telephone interviewing (CATI) system. The target population of the CHS includes adults in non-group quarters aged 18 and above who live in New York City and have a mobile phone or live in a household with a landline telephone. (Before 2009, the CHS only included those living in households with a landline telephone.) In addition, the survey is translated into languages other than English, like Spanish, Russian, Mandarin, and Cantonese. In 2017, Bengali and Haitian Creole were also added as language options.

The data for CHS for COVID-19 includes that of children. Five thousand three hundred five adults were interviewed following this module from March to August 2020. A total of 6,777 New Yorkers were included in the sample and questioned about the experiences of an additional 1,472 children. 75% of homes with minors

were asked to answer the COVID-19 module questions on behalf of the child at home from March through May 2020, if the adult conducting the survey knew enough about the child. Participants under 18 were not included in this study. Participants for the serosurvey were recruited from the ongoing NYC CHS. From June to October 2020, 1074 respondents completed the survey; 497 provided whole blood, and 577 provided only self-reported antibody test results [18] (p.190).

A complete case analysis in the form of a case deletion was used if data was missing. This technique involves discarding cases with missing values and proceeding with the analysis using standard methods. CID links the CHS and Population Health Survey responses to results from the NYC Serosurvey. COVID-19 data may also be related to the Serosurvey dataset using CID. The serosurvey dataset includes the results of SARS-CoV-2 antibody tests for a subset of the adult survey population that agreed to provide blood samples for the purpose. The 2020 serosurvey is linked with the Population Health Survey dataset to run analyses by demographic groups or other variables of interest.

First, the R libraries were loaded in RStudio v2022. Then, the haven and Dplyr libraries were loaded in the R script. The dimensions were assessed after loading the combined dataset into the R script. The SORT command was used to determine the column names (variables). The combined dataset without missing cases has 5305 observations with 202 variables. This dataset is used for RQ1 to determine the characteristics of the population. The FILTER command only added the 'sero1_result' to the new dataset. This removed the cases without serosurvey results, and the latest data set was left with 928 observations and 202 variables. Interpersonal violence variables were combined with the OR clause.

Furthermore, the categorical variables were changed into factors using the MUTATE clause. COVID-19 seropositivity was coded zero as unfavorable and one as positive. All other binomial variables were coded similarly, with 0 being negative and 1 being positive.

2.1. Statistical Analysis

Logistic regression helps analyze effects among variables, where the dependent variable is categorical (e.g., presence or absence) and the explanatory (independent) variables are categorical, numerical, or both. The outcome variable for research question 2, COVID-19 seropositivity (1: positive for COVID-19, 2: negative for COVID-19), is categorical. Hence this is compatible with the analysis.

The analyses investigate which variable has the most significant effects on the seropositivity of COVID-19 using a hierarchical logistic regression model-building method. As discussed in chapter 2, determinants of health can be divided into three major categories: individual characteristics and behaviors, physical environment, and social and economic environment. The order of the variable's entry into the hierarchical analysis is based on the determinants of health theory. The first model includes the socio-economic factors, the second model includes socio-economic and health behavior factors, the third model includes socio-economic, health behavior, and built environment factors, and the full model includes all the

above and existing co-morbidities and health conditions.

The sample size for the hierarchical regression is 928. Before the models were created, all required libraries were installed and added to RStudio v2022 for data transformation: haven and dplyr. All categorical variables were converted into factors. The mutate clause was used in R to perform this conversion and create a subset. Then, the four different models were pre-defined. The logistic regression models were generated, and the summary of the model and coefficients were pulled and assessed. Each variable's effect and significance with z statistics and p values from the regression coefficient table were identified. The model fit (model performance) was assessed using the Akaike Information Criterion (AIC) and the Hosmer-Lemeshow goodness of fit chi-square test.

The random forest model is used to identify influential variables affecting the seropositivity of COVID-19 using the variance importance plot results. For the random forest modeling in the second research question, the data were not randomly split into train and test datasets with a 7:3 ratio since we are not predicting but viewing the variables of most importance. The "randomForest" library was installed and primarily used for this model. First, the random forest equation was created and generated. Then, the variance importance plot was made with only the top ten critical variables according to the random forest model. This ranks the variables by their influence on the model.

3. Results

In model 1 (Table 3), a logistic regression was performed to ascertain the effects of socio-economic factors such as age, race, marital status, access to healthcare and insurance, among others, on the likelihood of testing positive /negative for COVID-19 seropositivity. The logistic regression model was statistically significant, $X^2(34, N = 928) = 67.24, p = 0.001$. The Hosmer and Lemeshow test indicated a good fit ($p = 0.20$). The model explained 7.16% (McFadden R^2) of the variance in seropositivity and correctly classified 79.63% of cases. Age group 65-75 ($p = 0.04$), Black ($p = 0.01$) and Hispanic ($p = 0.02$) race, and the US as the birthplace ($p = 0.04$) were statistically significant factors. Those between the ages of 65-75 were 0.38 times as likely than 18-24-year-olds to test positive for COVID-19 seropositivity (OR=0.38, 95% CI [0.15,0.95]). The Blacks (OR=1.99, 95% CI [1.19,3.3]) and Hispanics (OR=1.82, 95% CI [1.12,2.98]) were almost twice as likely as whites to be tested positive for COVID-19 seropositivity. Being born in the US, was also associated with testing negative for COVID-19 seropositivity (OR=0.66, 95% CI [0.44,0.98]).

In model 2 (Table 2) a logistic regression was performed to ascertain the effects of socio-economic factors with the health behavior factors on the likelihood of testing positive/negative for COVID-19 seropositivity. The logistic regression model was statistically significant, $X^2(42, N = 928) = 87.12, p < 0.001$. The Hosmer and Lemeshow test indicated a good fit ($p = 0.78$). The model explained 9.28% (McFadden R^2) of the variance in seropositivity and correctly classified 79.63% of cases. With the inclusion of health behaviors in this model, the Black race ($p = 0.04$), being a tobacco smoker ($p = 0.03$),

and exercising or engaging in physical activity ($p < 0.01$) were statistically significant. Those who are smokers were 0.46 times as likely as those who never smoked to be tested positive for COVID-19 seropositivity (OR=0.46, 95% CI [0.22,0.89]). Additionally, those who exercised recently were 0.53 times as likely to be tested positive for COVID-19 seropositivity (OR=0.53, 95% CI [0.35,0.81]).

Model 3 also entails a logistic regression to ascertain the effects of socio-economic and health behavior factors and built environment factors on the likelihood of testing positive or negative for COVID-19 seropositivity. The logistic regression model was statistically significant, X^2 (47, $N = 928$) = 92.62, $p < 0.001$ (Table 5). The Hosmer and Lemeshow test indicated a good fit ($p=0.42$). The model explained 9.87% (McFadden R^2) of the variance in seropositivity and correctly classified 81.14% of cases. With the inclusion of built environment factors such as borough of residence and number of adults in the household in this model, the same variables; Black race ($p=0.04$), being a current tobacco smoker ($p=0.03$) and exercising or engaging in physical activity ($p < 0.01$) were still statistically significant.

In the last model (Table 6), which is the complete model, a logistic regression was performed to ascertain the effects of socio-economic, health behavior, built environment factors, with the inclusion of comorbidities and health conditions, on the likelihood of testing positive or negative for COVID-19 seropositivity. The logistic regression model was statistically significant, X^2 (61, $N = 928$) = 116.78, $p < 0.001$. The Hosmer and Lemeshow test indicated a good fit ($p=0.90$). The model explained 12.45% (McFadden R^2) of the variance in seropositivity and correctly classified 80.60% of cases. With the inclusion of existing comorbidities and health conditions in this whole model, exercising or engaging in physical activity ($p < 0.01$), having asthma currently or in the past ($p=0.04$), BMI ($p=0.01$), and experiencing suicidal

thoughts recently ($p=0.04$) were found to be statistically significant factors. Those with or had asthma are 0.57 times as likely to be tested positive for COVID-19 seropositivity (OR=0.57, 95% CI [0.33,0.96]). Every point increase of BMI increases the chance of testing positive for COVID-19 seropositivity by 4% (OR=1.04, 95% CI [1.01,1.08]). Individuals who experienced suicidal thoughts recently were also 2.65 times more likely to be tested positive for COVID-19 seropositivity (OR=2.65, 95% CI [1.02,6.55]).

After controlling for demographical and socio-economic characteristics and introducing health behavior characteristics in model 2, the explanatory variance increased by 2.12%, with both physical activity and smoking behavior reaching significant levels. The predictive power of exercise behavior was higher, indicating that less exercise was significantly associated with greater chances of seropositivity. After incorporating the built environment factors in model 3, including socio-economic and health behavior characteristics, the total explanatory variance rose only 0.59%. In model 4, with the addition of comorbidities, the variance was 12.45%, increasing by 2.58%. The comparison of all models and change in significance of variables is represented in Table 3. Goodness of fit testing was performed for the regression analysis using Hosmer-Lemeshow test. These values indicate that these four models and data fit was good. Additionally, this showed that using regression models with these variables was appropriate while predicting the outcome.

When using the random forest method, the variance importance plot (Figure 3) suggested ten variables that had the highest effect on COVID-19 seropositivity: BMI, Kessler-6 index score for mental health, consumption of fruits and vegetables, imputed poverty level, age group, race, the borough of residence, marital status, and the number of adults in the household.

Table 1. Results of Logistic Regression Model with Socioeconomic Factors

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer-Lemeshow χ^2 (p)	Pseudo R^2 (McFadden)
Model 1- Socio-Economic Factors	16	67.24 (0.001)	940.85	10.98 (0.20)	7.16%
Model Summary					
	Predictor	Estimate (SE)	z (p)	OR (95% CI)	
	(Intercept)	-1.42 (0.81)	-1.75 (0.08)	0.24 (0.05, 1.14)	0.24
Access to Healthcare					
	Has insurance	-0.27 (0.35)	-0.77 (0.44)	0.77 (0.39, 1.53)	0.77
	Has personal doctor	-0.09 (0.22)	-0.41 (0.68)	0.91 (0.59, 1.42)	0.91
	Did not get care recently	0.07 (0.23)	0.31 (0.76)	1.08 (0.67, 1.68)	1.08
	Has regular prescription	0.33 (0.21)	1.55 (0.12)	1.39 (0.92, 2.11)	1.39
	Got mental health care	-0.4 (0.24)	-1.66 (0.1)	0.67 (0.42, 1.07)	0.67
Neighborhood Environment					
	Has home ownership	-0.01 (0.22)	-0.04 (0.97)	0.99 (0.65, 1.51)	0.99
	Recently delayed rent pay	0.34 (0.23)	1.48 (0.14)	1.4 (0.89, 2.18)	1.4
Demographics					
Age	Reference: 18-24				
A	25-29	-0.54 (0.46)	-1.16 (0.24)	0.58 (0.23, 1.44)	0.58
B	30-44	-0.48 (0.37)	-1.28 (0.2)	0.62 (0.3, 1.31)	0.62
C	45-64	-0.55 (0.39)	-1.43 (0.15)	0.57 (0.27, 1.24)	0.57
D	65-75	-0.97 (0.47)	-2.08 (0.04*)	0.38 (0.15, 0.95)	0.38
E	75+	-1.21 (0.65)	-1.85 (0.06)	0.3 (0.08, 1.03)	0.3

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer-Lemeshow χ^2 (p)	Pseudo R2 (McFadden)
	Cisgender woman	-0.09 (0.18)	-0.52 (0.6)	0.91 (0.64, 1.3)	0.91
Race	Reference: White				
A	Black	0.69 (0.26)	2.64 (0.01*)	1.99 (1.19, 3.3)	1.99
B	Hispanic	0.6 (0.25)	2.4 (0.02*)	1.82 (1.12, 2.98)	1.82
C	Asian/PI	0.05 (0.35)	0.14 (0.89)	1.05 (0.52, 2.04)	1.05
D	Other	0.6 (0.52)	1.15 (0.25)	1.82 (0.61, 4.81)	1.82
	Born in US	-0.42 (0.2)	-2.05 (0.04*)	0.66 (0.44, 0.98)	0.66
Marital Status	Reference: Married				
A	Divorced	0.19 (0.23)	0.83 (0.41)	1.21 (0.77, 1.91)	1.21
B	Widowed	0.24 (0.34)	0.72 (0.47)	1.27 (0.64, 2.44)	1.27
C	Separated	-0.48 (0.49)	-0.99 (0.32)	0.62 (0.22, 1.54)	0.62
D	Never Married	-0.09 (0.29)	-0.33 (0.74)	0.91 (0.52, 1.59)	0.91
Orientation	Reference: Gay/Lesbian				
A	Straight	0.52 (0.51)	1.03 (0.31)	1.68 (0.67, 5.15)	1.68
B	Bisexual	0.86 (0.62)	1.37 (0.17)	2.35 (0.72, 8.61)	2.35
C	Other	1.17 (0.68)	1.71 (0.09)	3.23 (0.87, 13.16)	3.23
Education	Reference: Less than HS				
A	HS Grad	0.2 (0.34)	0.57 (0.57)	1.22 (0.63, 2.4)	1.22
B	Some College	0.06 (0.35)	0.16 (0.87)	1.06 (0.53, 2.14)	1.06
C	College Grad	-0.06 (0.34)	-0.17 (0.86)	0.94 (0.49, 1.85)	0.94
Employment	Reference: Employed				
A	Unemployed	-0.15 (0.26)	-0.58 (0.57)	0.86 (0.51, 1.42)	0.86
B	Not in Labor Force	0.02 (0.24)	0.07 (0.94)	1.02 (0.63, 1.64)	1.02
Poverty Group	Reference: <100% poverty				
A	100%-<200% poverty	0.2 (0.3)	0.67 (0.5)	1.22 (0.68, 2.2)	1.22
B	200%-<400% poverty	0.16 (0.32)	0.51 (0.61)	1.18 (0.63, 2.19)	1.18
C	400%-<600% poverty	0.14 (0.32)	0.44 (0.66)	1.15 (0.62, 2.16)	1.15
D	>=600% poverty	-0.1 (0.34)	-0.29 (0.77)	0.91 (0.46, 1.78)	0.91

Table 2. Results of Logistic Regression Model with Socioeconomic and Health Behavior Factors

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer-Lemeshow χ^2 (p)	Pseudo R2 (McFadden)
Model 2- Socioeconomic and Health Behavior Factors	23	87.12 (< 0.001)	936.98	4.76 (0.78)	9.28%
Model Summary					
	Predictor	Estimate (SE)	z (p)	OR (95% CI)	
	(Intercept)	-1.1 (0.86)	-1.27 (0.2)	0.33 (0.06, 1.74)	0.33
Access to Healthcare					
	Has insurance	-0.26 (0.36)	-0.73 (0.46)	0.77 (0.39, 1.57)	0.77
	Has personal doctor	-0.06 (0.23)	-0.27 (0.79)	0.94 (0.61, 1.47)	0.94
	Did not get care recently	0.09 (0.24)	0.37 (0.71)	1.09 (0.68, 1.72)	1.09
	Has regular prescription	0.3 (0.22)	1.38 (0.17)	1.35 (0.88, 2.06)	1.35
	Got mental health care	-0.38 (0.24)	-1.57 (0.12)	0.68 (0.42, 1.09)	0.68
Neighborhood Environment					
	Has home ownership	0 (0.22)	0.01 (0.99)	1 (0.65, 1.54)	1 (0)
	Recently delayed rent pay	0.29 (0.23)	1.25 (0.21)	1.34 (0.84, 2.11)	1.34
Demographics					
Age	Reference: 18-24				
a	25-29	-0.48 (0.47)	-1.03 (0.31)	0.62 (0.24, 1.55)	0.62
b	30-44	-0.47 (0.39)	-1.21 (0.23)	0.62 (0.29, 1.36)	0.62
c	45-64	-0.54 (0.4)	-1.36 (0.17)	0.58 (0.27, 1.29)	0.58
d	65-75	-0.96 (0.48)	-2 (0.05)	0.38 (0.15, 0.98)	0.38
e	75+	-1.22 (0.67)	-1.81 (0.07)	0.3 (0.07, 1.06)	0.3
	Cisgender woman	0 (0.19)	-0.01 (0.99)	1 (0.69, 1.44)	1 (0)
Race	Reference: White				
a	Black	0.55 (0.27)	2.01 (0.04*)	1.73 (1.01, 2.96)	1.73
b	Hispanic	0.42 (0.26)	1.63 (0.1)	1.53 (0.92, 2.55)	1.53
c	Asian/PI	0.04 (0.35)	0.11 (0.91)	1.04 (0.51, 2.04)	1.04
d	Other	0.56 (0.53)	1.05 (0.29)	1.75 (0.57, 4.73)	1.75

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer-Lemeshow χ^2 (p)	Pseudo R2 (McFadden)
	Born in US	-0.36 (0.21)	-1.71 (0.09)	0.7 (0.46, 1.06)	0.7
Marital Status	Reference: Married				
a	Divorced	0.16 (0.23)	0.7 (0.48)	1.18 (0.75, 1.87)	1.18
b	Widowed	0.28 (0.35)	0.8 (0.43)	1.32 (0.65, 2.6)	1.32
c	Separated	-0.46 (0.49)	-0.94 (0.35)	0.63 (0.22, 1.59)	0.63
d	Never Married	-0.19 (0.29)	-0.66 (0.51)	0.82 (0.46, 1.45)	0.82
Orientation	Reference: Gay/Lesbian				
a	Straight	0.68 (0.52)	1.32 (0.19)	1.98 (0.78, 6.13)	1.98
b	Bisexual	1.02 (0.63)	1.6 (0.11)	2.76 (0.83, 10.26)	2.76
c	Other	1.16 (0.7)	1.64 (0.1)	3.17 (0.82, 13.35)	3.17
Education	Reference: Less than HS				
a	HS Grad	0.16 (0.35)	0.44 (0.66)	1.17 (0.59, 2.36)	1.17
b	Some College	0.05 (0.36)	0.15 (0.88)	1.06 (0.52, 2.18)	1.06
c	College Grad	-0.02 (0.35)	-0.07 (0.94)	0.98 (0.49, 1.97)	0.98
Employment	Reference: Employed				
a	Unemployed	-0.1 (0.27)	-0.36 (0.72)	0.91 (0.53, 1.52)	0.91
b	Not in Labor Force	0.04 (0.25)	0.16 (0.87)	1.04 (0.63, 1.7)	1.04
Poverty Group	Reference: <100% poverty				
a	100%-<200% poverty	0.26 (0.31)	0.84 (0.4)	1.29 (0.71, 2.36)	1.29
b	200%-<400% poverty	0.23 (0.32)	0.72 (0.47)	1.26 (0.67, 2.36)	1.26
c	400%-<600% poverty	0.18 (0.33)	0.54 (0.59)	1.19 (0.63, 2.26)	1.19
d	>=600% poverty	-0.04 (0.35)	-0.12 (0.9)	0.96 (0.48, 1.91)	0.96
Substance Use					
Tobacco Use	Reference: Never smoked				
a	Current	-0.77 (0.35)	-2.19 (0.03*)	0.46 (0.22, 0.89)	0.46
b	Former	-0.06 (0.22)	-0.3 (0.77)	0.94 (0.61, 1.43)	0.94
	Heavy drinker	-0.01 (0.34)	-0.03 (0.98)	0.99 (0.49, 1.88)	0.99
Physical Activity					
	Recently exercised	-0.63 (0.21)	-3.01 (<0.01*)	0.53 (0.35, 0.81)	0.53
	Had HIV test	0.27 (0.23)	1.17 (0.24)	1.31 (0.84, 2.07)	1.31
Immunization					
	Recently received flu shot	0.08 (0.18)	0.45 (0.65)	1.09 (0.76, 1.56)	1.09
Nutrition					
	Vegetables consumed	-0.1 (0.05)	-1.77 (0.08)	0.91 (0.81, 1.01)	0.91
	Drinks soda every day	0.22 (0.23)	0.97 (0.33)	1.24 (0.79, 1.93)	1.24

Table 3. Results of Logistic Regression Model with Socioeconomic, Health Behavior, and Built Environment Factors

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer-Lemeshow χ^2 (p)	Pseudo R2 (McFadden)
Model 3- Socioeconomic, Health Behaviors, and Built Environment factors	25	92.62 (< 0.001)	941.48	8.09 (0.42)	9.87%
Model Summary					
	Predictor	Estimate (SE)	z (p)	OR (95% CI)	
	(Intercept)	-1.62 (0.91)	-1.79 (0.07)	0.2 (0.03, 1.13)	0.2
Access to HealthCare					
	Has insurance	-0.28 (0.36)	-0.78 (0.44)	0.76 (0.38, 1.55)	0.76
	Has personal doctor	-0.01 (0.23)	-0.04 (0.97)	0.99 (0.64, 1.56)	0.99
	Did not get care recently	0.08 (0.24)	0.35 (0.73)	1.09 (0.67, 1.72)	1.09
	Has regular prescription	0.28 (0.22)	1.29 (0.2)	1.32 (0.87, 2.04)	1.32
	Got mental health care	-0.37 (0.24)	-1.54 (0.12)	0.69 (0.42, 1.1)	0.69
Neighborhood Environment					
	Has home ownership	-0.02 (0.23)	-0.07 (0.95)	0.98 (0.63, 1.53)	0.98
	Recently delayed rent pay	0.3 (0.23)	1.3 (0.19)	1.36 (0.85, 2.14)	1.36
Demographics					
Age	Reference: 18-24				
a	25-29	-0.43 (0.47)	-0.91 (0.36)	0.65 (0.25, 1.64)	0.65
b	30-44	-0.3 (0.4)	-0.76 (0.45)	0.74 (0.34, 1.64)	0.74
c	45-64	-0.42 (0.4)	-1.05 (0.29)	0.65 (0.3, 1.46)	0.65

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer- Lemeshow χ^2 (p)	Pseudo R2 (McFadden)
d	65-75	-0.79 (0.49)	-1.6 (0.11)	0.45 (0.17, 1.19)	0.45
e	75+	-1 (0.68)	-1.48 (0.14)	0.37 (0.09, 1.33)	0.37
	Cisgender woman	0 (0.19)	-0.03 (0.98)	1 (0.68, 1.44)	1 (0)
Race	Reference: White				
a	Black	0.56 (0.28)	2.03 (0.04*)	1.75 (1.02, 3.01)	1.75
b	Hispanic	0.42 (0.27)	1.56 (0.12)	1.53 (0.89, 2.6)	1.53
c	Asian/PI	0.01 (0.36)	0.02 (0.98)	1.01 (0.49, 2)	1.01
d	Other	0.57 (0.53)	1.08 (0.28)	1.78 (0.58, 4.82)	1.78
	Born in US	-0.33 (0.21)	-1.57 (0.12)	0.72 (0.47, 1.09)	0.72
Marital Status	Reference: Married				
a	Divorced	0.09 (0.24)	0.38 (0.7)	1.1 (0.69, 1.75)	1.1
b	Widowed	0.22 (0.35)	0.61 (0.54)	1.24 (0.61, 2.46)	1.24
c	Separated	-0.4 (0.5)	-0.81 (0.42)	0.67 (0.24, 1.71)	0.67
d	Never Married	-0.17 (0.3)	-0.56 (0.57)	0.85 (0.47, 1.51)	0.85
Orientation	Reference: Gay/Lesbian				
a	Straight	0.58 (0.52)	1.11 (0.27)	1.79 (0.7, 5.56)	1.79
b	Bisexual	0.93 (0.63)	1.47 (0.14)	2.55 (0.76, 9.49)	2.55
c	Other	1.01 (0.71)	1.43 (0.15)	2.74 (0.7, 11.62)	2.74
Education	Reference: Less than HS				
a	HS Grad	0.18 (0.36)	0.5 (0.62)	1.2 (0.6, 2.43)	1.2
b	Some College	0.05 (0.37)	0.15 (0.88)	1.06 (0.52, 2.19)	1.06
c	College Grad	0.01 (0.36)	0.04 (0.97)	1.01 (0.51, 2.07)	1.01
Employment	Reference: Employed				
a	Unemployed	-0.06 (0.27)	-0.23 (0.82)	0.94 (0.55, 1.57)	0.94
b	Not in Labor Force	0.04 (0.25)	0.17 (0.86)	1.05 (0.63, 1.71)	1.05
Poverty Group	Reference: <100% poverty				
a	100%-<200% poverty	0.25 (0.31)	0.81 (0.42)	1.28 (0.7, 2.36)	1.28
b	200%-<400% poverty	0.2 (0.32)	0.62 (0.54)	1.22 (0.65, 2.3)	1.22
c	400%-<600% poverty	0.19 (0.33)	0.57 (0.57)	1.2 (0.63, 2.3)	1.2
d	>=600% poverty	-0.05 (0.36)	-0.14 (0.89)	0.95 (0.48, 1.92)	0.95
Substance Use					
Tobacco Use	Reference: Never smoked				
a	Current	-0.75 (0.36)	-2.12 (0.03*)	0.47 (0.22, 0.91)	0.47
b	Former	-0.06 (0.22)	-0.29 (0.77)	0.94 (0.61, 1.44)	0.94
	Heavy drinker	-0.02 (0.34)	-0.05 (0.96)	0.98 (0.49, 1.87)	0.98
Physical Activity					
	Recently exercised	-0.67 (0.21)	-3.18 (<0.01*)	0.51 (0.34, 0.78)	0.51
	Had HIV test	0.27 (0.23)	1.15 (0.25)	1.31 (0.83, 2.07)	1.31
Immunization					
	Recently received flu shot	0.1 (0.19)	0.53 (0.59)	1.1 (0.77, 1.59)	1.1
Nutrition					
	Vegetables consumed	-0.1 (0.05)	-1.76 (0.08)	0.91 (0.81, 1.01)	0.91
	Drinks soda every day	0.2 (0.23)	0.87 (0.39)	1.22 (0.77, 1.89)	1.22
Built Environment					
Borough	Reference: Bronx				
a	Brooklyn	0.2 (0.26)	0.75 (0.45)	1.22 (0.73, 2.04)	1.22
b	Manhattan	0.26 (0.29)	0.91 (0.36)	1.3 (0.74, 2.29)	1.3
c	Queens	0.09 (0.28)	0.32 (0.75)	1.09 (0.63, 1.89)	1.09
d	Staten Island	-0.22 (0.49)	-0.44 (0.66)	0.81 (0.29, 2)	0.81
	# of adults in household	0.17 (0.09)	1.95 (0.05)	1.19 (1, 1.41)	1.19

Table 4. Results of the Full Model

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer- Lemeshow χ^2 (p)	Pseudo R2 (McFadden)
Full Model- with Cormorbidities/health condition factors included	36	116.78 (<0.001)	945.32	3.47 (0.90)	12.45%
Model Summary					
	Predictor	Estimate (SE)	z (p)	OR (95% CI)	
	(Intercept)	-2.56 (1.03)	-2.48 (0.01*)	0.08 (0.01, 0.56)	0.08
Access to Healthcare					
	Has insurance	-0.21 (0.37)	-0.57 (0.57)	0.81 (0.4, 1.69)	0.81
	Has personal doctor	-0.04 (0.24)	-0.17 (0.86)	0.96 (0.61, 1.54)	0.96
	Did not get care recently	0.16 (0.25)	0.62 (0.54)	1.17 (0.7, 1.91)	1.17
	Has regular prescription	0.31 (0.23)	1.34 (0.18)	1.36 (0.87, 2.16)	1.36
	Got mental health care	-0.25 (0.27)	-0.93 (0.35)	0.78 (0.45, 1.31)	0.78
Neighborhood Environment					
	Has home ownership	0.04 (0.23)	0.16 (0.87)	1.04 (0.65, 1.64)	1.04
	Recently delayed rent pay	0.35 (0.25)	1.43 (0.15)	1.42 (0.87, 2.29)	1.42
Demographics					
Age	Reference: 18-24				
a	25-29	-0.58 (0.49)	-1.19 (0.23)	0.56 (0.21, 1.45)	0.56
b	30-44	-0.44 (0.41)	-1.07 (0.29)	0.64 (0.29, 1.46)	0.64
c	45-64	-0.59 (0.43)	-1.38 (0.17)	0.55 (0.24, 1.3)	0.55
d	65-75	-1.01 (0.52)	-1.96 (0.05)	0.36 (0.13, 1)	0.36
e	75+	-1.29 (0.7)	-1.84 (0.07)	0.28 (0.07, 1.05)	0.28
	Cisgender woman	-0.06 (0.2)	-0.32 (0.75)	0.94 (0.64, 1.38)	0.94
Race	Reference: White				
a	Black	0.48 (0.29)	1.66 (0.1)	1.62 (0.91, 2.85)	1.62
b	Hispanic	0.36 (0.28)	1.26 (0.21)	1.43 (0.82, 2.5)	1.43
c	Asian/PI	0.1 (0.36)	0.27 (0.79)	1.1 (0.53, 2.22)	1.1
d	Other	0.51 (0.55)	0.92 (0.36)	1.66 (0.53, 4.66)	1.66
	Born in US	-0.32 (0.22)	-1.45 (0.15)	0.73 (0.47, 1.12)	0.73
Marital Status	Reference: Married				
a	Divorced	0.09 (0.24)	0.36 (0.72)	1.09 (0.68, 1.77)	1.09
b	Widowed	0.26 (0.36)	0.7 (0.48)	1.29 (0.62, 2.61)	1.29
c	Separated	-0.42 (0.52)	-0.82 (0.41)	0.65 (0.22, 1.73)	0.65
d	Never Married	-0.1 (0.31)	-0.34 (0.74)	0.9 (0.49, 1.64)	0.9
Orientation	Reference: Gay/Lesbian				
a	Straight	0.46 (0.54)	0.86 (0.39)	1.58 (0.6, 5.06)	1.58
b	Bisexual	0.84 (0.65)	1.29 (0.2)	2.33 (0.66, 8.98)	2.33
c	Other	0.94 (0.73)	1.28 (0.2)	2.55 (0.62, 11.28)	2.55
Education	Reference: Less than HS				
a	HS Grad	0.13 (0.37)	0.34 (0.73)	1.14 (0.55, 2.36)	1.14
b	Some College	-0.04 (0.38)	-0.09 (0.93)	0.97 (0.46, 2.05)	0.97
c	College Grad	-0.07 (0.37)	-0.19 (0.85)	0.93 (0.46, 1.95)	0.93
Employment	Reference: Employed				
a	Unemployed	-0.1 (0.28)	-0.37 (0.71)	0.9 (0.52, 1.54)	0.9
b	Not in Labor Force	0.23 (0.28)	0.83 (0.41)	1.26 (0.73, 2.15)	1.26
Poverty Group	Reference: <100% poverty				
a	100%-<200% poverty	0.18 (0.32)	0.57 (0.57)	1.2 (0.64, 2.24)	1.2
b	200%-<400% poverty	0.08 (0.33)	0.25 (0.8)	1.09 (0.56, 2.09)	1.09
c	400%-<600% poverty	0.09 (0.34)	0.26 (0.8)	1.09 (0.56, 2.13)	1.09
d	>=600% poverty	-0.13 (0.37)	-0.35 (0.72)	0.88 (0.43, 1.81)	0.88
Substance Use					
Tobacco Use	Reference: Never smoked				
a	Current	-0.71 (0.36)	-1.95 (0.05)	0.49 (0.23, 0.98)	0.49
b	Former	-0.05 (0.23)	-0.22 (0.82)	0.95 (0.61, 1.47)	0.95
	Heavy drinker	0.03 (0.35)	0.07 (0.94)	1.03 (0.51, 1.97)	1.03
Physical Activity					
	Recently exercised	-0.69 (0.22)	-3.12 (<0.01*)	0.5 (0.33, 0.78)	0.5

Type	Number of Predictors	Model Significance χ^2 (p)	AIC	Hosmer- Lemeshow χ^2 (p)	Pseudo R2 (McFadden)
	Had HIV test	0.27 (0.24)	1.14 (0.26)	1.31 (0.82, 2.12)	1.31
Immunization					
	Recently received flu shot	0.16 (0.19)	0.86 (0.39)	1.18 (0.81, 1.72)	1.18
Nutrition					
	Vegetables consumed	-0.09 (0.06)	-1.63 (0.1)	0.91 (0.82, 1.01)	0.91
	Drinks soda every day	0.16 (0.24)	0.68 (0.5)	1.17 (0.73, 1.85)	1.17
Built Environment					
Borough	Reference: Bronx				
a	Brooklyn	0.31 (0.27)	1.16 (0.25)	1.37 (0.81, 2.33)	1.37
b	Manhattan	0.39 (0.3)	1.31 (0.19)	1.48 (0.82, 2.68)	1.48
c	Queens	0.12 (0.29)	0.44 (0.66)	1.13 (0.65, 1.99)	1.13
d	Staten Island	-0.25 (0.5)	-0.49 (0.62)	0.78 (0.28, 1.99)	0.78
	# of adults in household	0.17 (0.09)	1.94 (0.05)	1.19 (1, 1.41)	1.19
Physical Health Status					
General Health	Reference: Excellent				
	Very Good	-0.04 (0.26)	-0.14 (0.89)	0.96 (0.58, 1.61)	0.96
	Good	0.26 (0.27)	0.94 (0.35)	1.29 (0.76, 2.22)	1.29
	Fair	-0.08 (0.37)	-0.23 (0.82)	0.92 (0.44, 1.91)	0.92
	Poor	-0.05 (0.6)	-0.07 (0.94)	0.96 (0.27, 3)	0.96
	Has high blood pressure	0.01 (0.23)	0.04 (0.97)	1.01 (0.64, 1.6)	1.01
	Has diabetes	-0.15 (0.33)	-0.47 (0.64)	0.86 (0.44, 1.61)	0.86
	Has/had asthma	-0.55 (0.27)	-2.04 (0.04*)	0.57 (0.33, 0.96)	0.57
	Bmi	0.04 (0.02)	2.87 (<0.01*)	1.04 (1.01, 1.08)	1.04
	Use assistive device	-0.19 (0.37)	-0.51 (0.61)	0.83 (0.39, 1.69)	0.83
	Has difficulty in daily activities	-0.44 (0.34)	-1.31 (0.19)	0.64 (0.33, 1.22)	0.64
Mental Health Status					
	Kessler-6 index score	-0.01 (0.03)	-0.16 (0.87)	0.99 (0.94, 1.06)	0.99
	Recently had non-specific psychological distress	-0.43 (0.51)	-0.83 (0.4)	0.65 (0.23, 1.75)	0.65
	Recently thought suicide	0.98 (0.47)	2.08 (0.04*)	2.65 (1.02, 6.55)	2.65
	Experienced domestic violence	-0.08 (0.34)	-0.24 (0.81)	0.92 (0.45, 1.76)	0.92

Table 5. Odds Ratios and Significance of Predictors of Hierarchical Logistic Regression

Block	Category	Predictor	Model 1	Model 2	Model 3	Model 4
Socioeconomic Factors	Access to Healthcare	(Intercept)	0.24	0.33	0.2	0.08*
		Has insurance	0.77	0.77	0.76	0.81
		Has personal doctor	0.91	0.94	0.99	0.96
		Did not get care recently	1.08	1.09	1.09	1.17
		Has regular prescription	1.39	1.35	1.32	1.36
		Got mental health care	0.67	0.68	0.69	0.78
	Neighborhood Environment	Has home ownership	0.99	1	0.98	1.04
		Recently delayed rent pay	1.4	1.34	1.36	1.42
	Demographics	Age				
		25-29	0.58	0.62	0.65	0.56
		30-44	0.62	0.62	0.74	0.64
		45-64	0.57	0.58	0.65	0.55
		65-75	0.38	0.38	0.45	0.36
		75+	0.3	0.3	0.37	0.28
		Cisgender woman	0.91	1	1	0.94
		Race				
		Black	1.99*	1.73*	1.75*	1.62
		Hispanic	1.82	1.53	1.53	1.43
		Asian/PI	1.05	1.04	1.01	1.1
		Other	1.82	1.75	1.78	1.66
		Born in US	0.66	0.7	0.72	0.73
		Marital Status				
		Divorced	1.21	1.18	1.1	1.09
		Widowed	1.27	1.32	1.24	1.29

Block	Category	Predictor		Model 1	Model 2	Model 3	Model 4
			Separated	0.62	0.63	0.67	0.65
			Never Married	0.91	0.82	0.85	0.9
		Orientation					
			Straight	1.68	1.98	1.79	1.58
			Bisexual	2.35	2.76	2.55	2.33
			Other	3.23	3.17	2.74	2.55
		Education					
			HS Grad	1.22	1.17	1.2	1.14
			Some College	1.06	1.06	1.06	0.97
			College Grad	0.94	0.98	1.01	0.93
		Employment					
			Unemployed	0.86	0.91	0.94	0.9
			Not in Labor Force	1.02	1.04	1.05	1.26
		Poverty Group					
			100%-<200% poverty	1.22	1.29	1.28	1.2
			200%-<400% poverty	1.18	1.26	1.22	1.09
			400%-<600% poverty	1.15	1.19	1.2	1.09
			>=600% poverty	0.91	0.96	0.95	0.88
Health Behaviors	Substance Use	Tobacco Use					
			Current		0.46*	0.47*	0.49
			Former		0.94	0.94	0.95
		Heavy drinker			0.99	0.98	1.03
	Physical Activity	Recently exercised			0.53**	0.51**	0.5**
		Had HIV test			1.31	1.31	1.31
	Immunization	Recently received flu shot			1.09	1.1	1.18
	Nutrition	Vegetables consumed			0.91	0.91	0.91
Drinks soda every day			1.24	1.22	1.17		
Environment	Built Environment	Borough					
			Brooklyn			1.22	1.37
			Manhattan			1.3	1.48
			Queens			1.09	1.13
			Staten Island			0.81	0.78
		# of adults in household				1.19	1.19
Health Conditions	Physical Health Status	General Health					
			Very Good				0.96
			Good				1.29
			Fair				0.92
			Poor				0.96
		Has high blood pressure					1.01
		Has diabetes					0.86
		Has/had asthma					0.57*
		Bmi					1.04**
		Use assistive device					0.83
		Has difficulty in daily activities					0.64
	Mental Health Status	Kessler-6 index score					0.99
		Recently had non-specific psychological distress					0.65
		Recently thought suicide					2.65*
		Experienced domestic violence					0.92
* Significance at 0.05 level							
** Significance at 0.01 level							

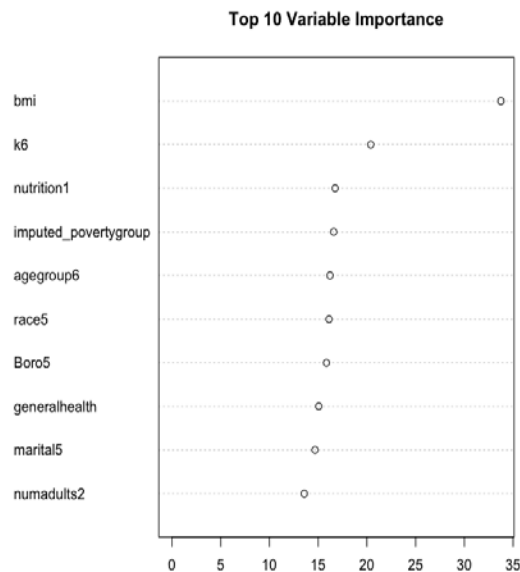


Figure 3. Variance Importance Plot for the Random Forest Logistic Regression Model

4. Discussion

The statistical analyses, which took the exploratory route, entailed creating four logistic regression models in a hierarchical regression method to gain a comprehensive understanding of how such varied determinants of health affect COVID-19 seropositivity in the New York City population.

The complete model found significant relationships between COVID-19 seropositivity and asthma, BMI, physical activity, and the incidence of suicidal thoughts. Those with or had asthma are 0.57 times as likely to be tested positive for COVID-19 seropositivity (OR=0.57, 95% CI [0.33,0.96]). Every point increase in BMI increases the chance of testing positive for COVID-19 seropositivity by 4% (OR=1.04, 95% CI [1.01,1.08]). Individuals who experienced suicidal thoughts were also 2.65 times more likely to be tested positive for COVID-19 seropositivity (OR=2.65, 95% CI [1.02,6.55]). Exercise is a protective factor, whereas suicidal thoughts, high BMI, higher number of adults in a household, and asthma are found to be risk factors.

The finding that physical activity is associated with lower COVID-19 seropositivity in this NYC sample is consistent with public health literature highlighting the importance of regular exercise in maintaining overall health and immunity. A recent study analyzed data from 48,440 adults in the UK and found that those physically active before the pandemic were less likely to be hospitalized or die from COVID-19, as it may improve immune function and reduce the severity of symptoms [19]. Similarly, the association between BMI and COVID-19 seropositivity is consistent with the known risk factors, including obesity, leading to severe outcomes [20]. Asthma is generally a risk factor for severe respiratory infections, hence the findings here are consistent. Individuals with asthma may be more likely to practice social distancing and other protective measures, which could explain this association. Ko et al. [8] suggest that asthma patients contracting SARS-CoV-2 are more likely

to have severe outcomes and require intensive care. We also know from the literature that the odds of COVID-19 seropositivity were significantly higher among individuals living in households with more than four people than those living in smaller households. This finding runs true in this NYC sample as well. The inferences are generally consistent with the literature, which identified several risk factors for COVID-19, including age, underlying health conditions, and social determinants of health [10].

The finding that Black individuals have a higher risk of COVID-19 seropositivity in models 1, 2, and 3 is consistent with the public health literature. The group has been disproportionately affected by COVID-19 in many settings, including New York City in 2020 [10]. Concurrently, the associations between BMI, physical activity, difficulty in doing daily activities, and the prevalence of asthma with COVID-19 seropositivity are consistent with previous research on the risk factors [20]. Lippi et al. [21] found that smoking is associated with an increased risk of contracting COVID-19, as well as more severe outcomes and a higher mortality rate. Including the incidence of suicidal behaviors as a predictor of COVID-19 seropositivity is novel and warrants further investigation.

The Hosmer-Lemeshow test results indicate good model fit, and the variable selection across the different models provides some confidence in the results. The high p-values for the Hosmer-Lemeshow tests suggest a good fit for all models, and the AIC value indicates that this model is a good fit for the data. The results from this method provide further evidence for the several known risk factors for COVID-19 seropositivity and highlight the potential of machine learning methods in identifying complex and nonlinear relationships between predictors and outcomes.

The backward and forward stepwise regression models were deemed to have the best performance. While the backward stepwise model had slightly lower AIC than the forward stepwise model, the forward stepwise model had a higher p-value in the Hosmer-Lemeshow test and, thus, showed lower error between the predicted model and the actual values.

However, a model's performance may vary depending on the specific research question and dataset. Using multiple regression models with different variable selection techniques allowed for a comprehensive examination of the data and helped identify consistent predictors of COVID-19 seropositivity.

4.1. Implications

To date, no exploratory study has been conducted that comprehensively examines the impact of various determinants of health on COVID-19 seropositivity in the NYC population. Specifically, there is minimal research investigating the effects of approximately 40 types of determinants of health on the prevalence of COVID-19 antibodies in this population. This analysis helps us gain a more nuanced understanding of the determinants of health and their effects during the biggest pandemic we have lived through. Gaining a comprehensive view of the factors that were found to be shared among the various regression models; the prevalence of asthma, BMI, physical activity levels, the incidence of suicidal thoughts, and tobacco use, provides insights into the key factors that drive transmission and spread of the virus and should guide the development of targeted interventions and policies to improve health outcomes in vulnerable populations. For instance, policies could be implemented to provide targeted support and resources for individuals and households with a high prevalence of asthma or larger household sizes, such as access to medication, home-based oxygen therapy, or financial assistance for housing.

Public health policies promote healthy lifestyles and behaviors by providing education and resources to improve physical activity levels and reduce tobacco use. Such policies could include initiatives that increase access to affordable nutritious food options, create safe spaces for physical activity, and promote smoking cessation programs. Furthermore, policies could be developed to address the incidence of suicidal thoughts, encourage access to mental health services, and improve HIV testing rates to reduce the burden of comorbidities associated with COVID-19. Identifying critical predictors of COVID-19 seropositivity in an exploratory study offers valuable insights into the potential focus areas for public health policies and interventions to reduce disparities and improve health outcomes during future pandemics.

Identifying these critical predictors of COVID-19 seropositivity also offers opportunities to generate new research questions and explore the relationships between these factors and COVID-19 outcomes. Some possible future research questions and ideas include: A) Can interventions targeting the prevention and management of asthma help reduce the risk of COVID-19 infection and severity in such cases? B) What is the relationship between suicidal thoughts and COVID-19 outcomes, and what interventions could be implemented to support mental health and reduce the risk of such thoughts during a pandemic? C) What factors contribute to lower rates of HIV testing among specific populations, and the interventions needed to increase such testing rates and reduce the burden of comorbidities associated with the two diseases? By exploring these research questions and ideas, we can better understand the complex relationships

between various determinants of health and COVID-19 outcomes, thus developing effective public health interventions to reduce the burden of the pandemic on vulnerable populations.

It is also important to note a novel methodology used in the research question of this research study, that is, using variance importance plots generated by the random forest method, the most critical predictors contributing to prediction accuracy can be identified, and findings in public health research can be validated. Comparing results with logistic regression can help identify discrepancies or biases. This approach provides a comprehensive understanding of factors contributing to negative and positive health outcomes and helps prioritize interventions and policies to improve population health. Overall, this novel approach of comparing results from the random forest and logistic regression methods using variance importance plots can contribute to public health by providing new insights and more accurate predictions of health outcomes.

4.2. Limitations and Future Directions

This study has several limitations related to the self-reported nature of the NYC Community Health Survey (NYC CHS). This study's cross-sectional design inherently limits the ability to establish causal relationships between the identified predictors and COVID-19 seropositivity. The data represents a single time point, making it difficult to assess changes in exposure, health behaviors, or seropositivity status over time. Cross-sectional design also hinders establishing causal relationships and tracking changes over time. For example, the 2020 survey did not adequately capture the evolving nature of COVID-19, focusing only on seropositivity rather than severity, which was a significant limitation.

Additionally, the reliance on self-reported data introduces potential biases, including recall bias and social desirability bias, which may lead to misclassification or underreporting of key health behaviors and conditions. These factors could influence the observed associations, potentially overestimating or underestimating the strength of relationships between socioeconomic, behavioral, and health-related factors and seropositivity outcomes. Future studies utilizing longitudinal designs and objective health measures would be valuable in confirming these findings and better understanding the directionality of these associations. Misinterpretation of survey questions and incomplete responses further complicate data accuracy.

While this study provides valuable insights into COVID-19 seropositivity predictors among NYC adults, its generalizability to the broader NYC population and similar urban settings should be considered cautiously. The NYC CHS employs a stratified random-digit-dialing method, which may underrepresent certain demographic groups, particularly individuals without stable housing, those with limited access to telecommunications, and non-English speakers who may face language barriers despite survey translation efforts. Additionally, the study sample may not fully capture the diversity of experiences across all boroughs, particularly in neighborhoods with historically lower survey response rates. Although the findings align with broader public health research on health

disparities, caution is warranted when extrapolating results to populations outside NYC, as regional differences in healthcare access, socioeconomic conditions, and public health policies may influence seropositivity trends. Future research incorporating more diverse sampling methods and longitudinal designs would enhance the applicability of these findings to other urban populations.

The NYC CHS uses a random-digit-dialing methodology, which could introduce sampling bias and affect voluntary survey participation. Assessing severity can vary significantly among individuals, thus complicating data standardization. The overrepresentation of white males in the NYC CHS sample may have skewed results and affected the accuracy and generalizability of findings related to health behaviors. The presence or absence of specific demographic groups may have influenced observation patterns in the data.

Future research should address this study's limitations by incorporating longitudinal designs that track individuals over time to establish better causal relationships between socioeconomic, behavioral, and health-related factors and disease seropositivity. Longitudinal studies help determine whether specific risk factors precede infection or result from broader health and social conditions. Additionally, integrating objective health data, such as clinical records or biomarker assessments, could reduce reliance on self-reported measures and improve data accuracy. Expanding research to include more comprehensive data on disease severity, rather than just seropositivity, would provide a clearer picture of the health impact of a future pandemic across different demographic and socioeconomic groups. Finally, conducting similar studies in diverse urban and rural settings could enhance the generalizability of findings and inform targeted public health interventions.

ACKNOWLEDGEMENTS

We want to thank the Bureau of Epidemiology and DOHMH in NYC for their support with data access.

References

- [1] CDC, 2020.
- [2] Bellandi D., "Estimate: 10.5 Million Children Lost a Parent, Caregiver to COVID-19," *JAMA*, 328(15). 1490. 2022.
- [3] Rady Children's Hospital. COVID collaborative for children [Internet]. San Diego. Rady Children's Hospital. [Cited 2023 March 17] Available from: <https://www.rchsd.org/health-safety/health-alerts/covid-collaborative-for-children/>.
- [4] ICN. ICN says 115,000 healthcare worker deaths from COVID-19 exposes collective failure of leaders to protect global workforce [Internet]. Place unknown; Oct 2021. [Cited 2023 March 17] Available from: <https://www.icn.ch/news/icn-says-115000-healthcare-worker-deaths-covid-19-exposes-collective-failure-leaders-protect>.
- [5] Coker T.R., Cheng T.L., Ybarra M. "Addressing the Long-term Effects of the COVID-19 Pandemic on Children and Families: A Report from the National Academies of Sciences, Engineering, and Medicine," *JAMA*. March 2023, [Online].
- [6] Mottaleb, K.A., Mainuddin, M., and Sonobe, T., "Covid-19 induced economic loss and ensuring food security for vulnerable groups: Policy implications from Bangladesh," *PLOS ONE*, 15(10), Oct. 2020.
- [7] Jones, B., Woolfenden, S., Pengilly, S., Breen, C., Cohn, R., Biviano, L., Johns, A., Worth, A., Lamb, R., Lingam, R., Silove, N., Marks, S., Tzioumi, D., and Zwi, K., "COVID-19 pandemic: The impact on vulnerable children and young people in Australia," *Journal of Paediatrics and Child Health*, 56(12), 1851–1855, Sep. 2020.
- [8] Koh, D., "Occupational risks for COVID-19 infection," *Occupational Medicine*, 70(1), 3–5, January 2020.
- [9] Couch, K.A., Fairlie, R.W., and Xu, H., "Early evidence of the impacts of COVID-19 on minority unemployment," *Journal of Public Economics*, 192, 104287, 2020.
- [10] Singu, S., Acharya, A., Challagundla, K., and Byrareddy, S.N., "Impact of social determinants of health on the emerging COVID-19 pandemic in the United States," *Frontiers in Public Health*, 8, [Online] Available on: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2020.00406>.
- [11] Witters, D. In U.S., 14% with likely COVID-19 to avoid care due to cost [Internet]. Gallup; April 28, 2020 [Cited November 29, 2022]. Available from: <https://news.gallup.com/poll/309224/avoid-care-likely-covid-due-cost.aspx>.
- [12] Abrams, E.M., and Szefer, S.J., "COVID-19 and the impact of Social Determinants of Health," *The Lancet Respiratory Medicine*, 8(7), 659–661, Jul 2020.
- [13] Rezaei, N., Moghaddam, S.S., Farzadfar, F., and Larijani, B., "Social Determinants of health inequity in Iran: A narrative review," *Journal of Diabetes and Metabolic Disorders*, 22, 5-12, Jun 2023.
- [14] Robertson, M.M., Shamsunder, M.G., Brazier, E., Mantravadi, M., Zimba, R., Rane, M.S., Westmoreland, D.A., Parcesep, A.M., Maroko, A.R., Kulkarni, S.G., Grov, C., and Nash, D. "Racial/ethnic disparities in exposure, disease susceptibility, and clinical outcomes during COVID-19 pandemic in National Cohort of adults, United States," *Emerging Infectious Diseases*, 28(11), 2171–2180, Nov 2022.
- [15] Solar, O., and A. Irwin., "A conceptual framework for action on the social determinants of health. Social determinants of health discussion paper 2 (policy and practice)" Geneva, Switzerland: WHO; 2010 [cited: September 22, 2016] Available from: http://www.who.int/sdhconference/resources/ConceptualframeworkforactiononSDH_eng.pdf.
- [16] Committee on Educating Health Professionals to Address the Social Determinants of Health, *A framework for educating health professionals to address the social determinants of health*. The National Academies Press: National Academies of Sciences, Engineering, and Medicine, Institute of Medicine, Board on Global Health. Oct 14, 2016.
- [17] Frieden, T. R., "A framework for public health action: The health impact pyramid," *American Journal of Public Health* 100(4):590-595, Apr, 2010.
- [18] Parrott, J.C., Maleki, A.N., Vassor, V.E., Osahan, S., Hsin, Y., Sanderson, M., Fernandez, S., Levanon Seligson, A., Hughes, S., Wu, J., DeVito, A.K., LaVoie, S.P., Rakeman, J.L., Gould, L.H., and Aloy, K.A., "Prevalence of SARS-COV-2 antibodies in New York City adults, June–October 2020: A population-based survey," *The Journal of Infectious Diseases*, 224(2), 188–195, Jul, 2021.
- [19] Sallis, J.F., Adlakha, D., Oyeyemi, A., Salvo, D., and Anjana, R.M., "Adapting to the COVID-19 Pandemic: Physical Activity, Situational Awareness, and Resilience," *Translational Journal of the American College of Sports Medicine*, 6(12), 251-258, 2021.
- [20] Mänty, M., Heinonen, O.J., Viljanen, A., Pajala, S., Koskenvuo, M., Kaprio, J., and Rantanen, T., "Body mass index and disability in adulthood: A 20-year panel study," *American Journal of Public Health*, 101(2), 260-266, 2011.
- [21] Lippi, G., Henry, B.M., and Sanchis-Gomar, F., "Active smoking and COVID-19: a double-edged sword," *European Journal of Internal Medicine*, 77, 123-124, Jul, 2020.

