Association of Neighborhood-Level Factors and COVID-19 Infection Patterns in Philadelphia Using Spatial Regression

Mary Regina Boland, MA, MPhil, PhD, FAMIA[1], Jessica Liu[1], Cecilia Balocchi, PhD, Jessica Meeker, MPH[1], Ray Bai, PhD[2], Ian Mellis, PhD[3], Danielle L. Mowery, PhD, MS, MS[1], Daniel Herman, MD, PhD[3]
[1]Department of Biostatistics, Epidemiology & Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [2]Department of Statistics, University of South Carolina, SC, USA; [3]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract
As of August 2020, there were ~6 million COVID-19 cases in the United States of America, resulting in ~200,000 deaths. Informatics approaches are needed to better understand the role of individual and community risk factors for COVID-19. We developed an informatics method to integrate SARS-CoV-2 data with multiple neighborhood-level factors from the American Community Survey and opendataphilly.org. We assessed the spatial association between neighborhood-level factors and the frequency of SARS-CoV-2 positivity, separately across all patients and across asymptomatic patients. We found that neighborhoods with higher proportions of individuals with a high-school degree and/or who were identified as Hispanic/Latinx were more likely to have higher SARS-CoV-2 positivity rates, after adjusting for other neighborhood covariates. Patients from neighborhoods with higher proportions of individuals receiving public assistance and/or identified as White were less likely to test positive for SARS-CoV-2. Our approach and its findings could inform future public health efforts.

1. Introduction

1.1 Background on COVID-19

A novel strain of coronavirus (COVID-19) was discovered in Wuhan, China in December 2019. Initially, COVID-19 began as an epidemic in that local region, resulting in strict lock-downs of approximately 500 million people in China. Since then COVID-19 has since spread across the globe, impacting every continent including Antarctica (as of December 2020). As of August 25, 2020, there had been over 23 million confirmed cases and 800,000 deaths worldwide[1].

1.2 Importance of Identifying Neighborhood-Level Factors in COVID-19

Since March 2020, a plethora of COVID-19 related research papers have correlated social determinants of health with COVID-19 spread. Factors such as poverty and income can affect an individual's ability to effectively distance physically from others (colloquially as 'social distancing'). In addition, socioeconomic status is closely associated with job type [2]. Occupations, such as grocery store workers and pharmacy store workers are at increased risk of COVID-19 [3], and a large proportion of these jobs are held by those of lower socioeconomic status and/or immigrant populations [4]. It follows that neighborhood-level characteristics (e.g., income, poverty) of certain census tracts (i.e., neighborhoods) within a large diverse city such as Philadelphia, could be associated with different SARS-CoV-2 positivity rates. However, it is unclear at this present time what neighborhood-level factors are the most closely linked with COVID-19 spread and their relative importance to each other (e.g., is education or poverty or income more informative?). What is lacking is sufficient understanding of neighborhood-level characteristics to tailor public health messaging and other interventions to reduce the SARS-CoV-2 transmission across communities that may be otherwise high-risk for spread (e.g., certain low-income communities).

1.3 Significant Variability Exists in Neighborhood Characteristics in Philadelphia

Neighborhoods are defined in this paper as census tracts. Great variability exists across Philadelphia neighborhoods and it is truly a 'tale of two cities'. One is rich and the other poor in material resources. These neighborhood (i.e., census-tract) level associations within Philadelphia are directly applicable to healthcare. One example of neighborhood-level or census tract-level differences within Philadelphia is income with low-income neighborhoods having been linked with increases in fire-related injuries [5]. There also exists considerable disparity in terms of the number of primary care providers. The average ratio of adults per primary care provider was 1,073 across all census tracts within the city of Philadelphia [6]. However, some census tracts had only 105 while others had 10,321 [6]. This disparity affects physical availability access to care. In addition, we have shown that neighborhood-level factors in Philadelphia affect C-section rates among pregnant women [7]. Overall, significant variability exists within the city of Philadelphia in terms of the neighborhood-level characteristics (i.e., census tract-level characteristics). This is also true of many other urban centers (e.g., New York City, San Francisco) and therefore our approach should generalize to other urban centers.

1.4 Need for Informatics Approaches to Address these Gaps

Biomedical informatics methods can be developed to identify neighborhood-level factors that are associated with COVID-19 infection. If these factors were known, it could assist in effective containment of COVID-19 or other pandemics in future [8]. SARS-CoV-2 testing is being offered in a variety of settings. Some patients are received through an outpatient office or via drive through testing sites. Others are received via the in-patient care system (e.g., Emergency Departments). The testing orders, results, and associated clinical information such as clinical symptomology are stored and distributed in a complex web. For this study, we focused on testing being performed in one academic medical center on specimens collected from a variety of sites and patients with a variety of indications, all aggregated in one integrated electronic health record (EHR) system. Effective informatics solutions are needed to identify relevant patient cohorts (e.g., asymptomatic vs. symptomatic) and to understand associated neighborhood-level factors for these different cohorts.

1.5 Purpose of Study

The goals of this study are two-fold: 1.) develop an informatics approach to integrate multiple, potentially relevant neighborhood-level factors (e.g., income, poverty, education) from American Community Survey (ACS) and other Philadelphia-specific open datasets (e.g., violence, housing quality) from OpenDataPhilly.org with SARS-CoV-2 testing data; and 2.) determine what neighborhood-level factors are associated spatially with COVID-19 spread within the city of Philadelphia. These neighborhood-level factors are available at the census-tract level. We envision that this information could be used for future city and institutional planning to more readily identify COVID-19 'hotspots' within the city and the risk factors associated with these hotspots (e.g., poverty). Moreover, for healthcare institutions, these information should be useful for institutional planning purposes to understand what types of communities are most at risk to enable adaptation of testing and contact tracing strategies for those communities.

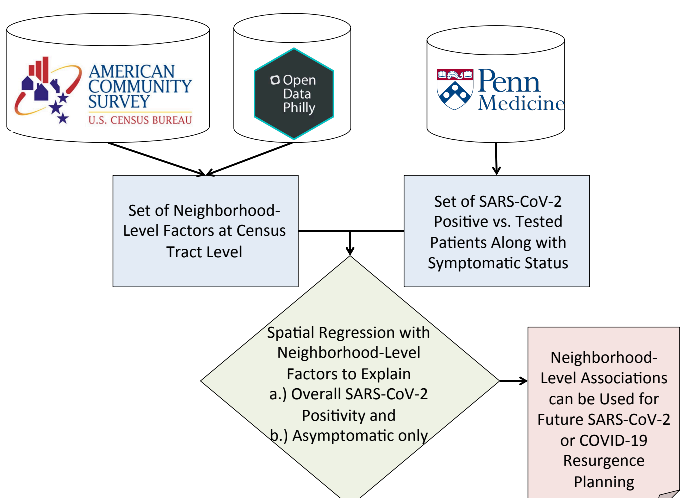

Figure 1. Overview of Informatics Approach to Aggregate a Set of Neighborhood-Level Covariates from Public Sources and Link Using Spatial Regression with SARS-CoV-2 Positivity Status for All Patients Tested and All Asymptomatic Patients Tested.

2. Materials and Methods

Our informatics approach involves assembling and integrating data SARS-CoV-2 testing data with various sources on neighborhood-level factors. We obtained clinical data of SARS-CoV-2 positivity rates broken down by symptomatic status (i.e., overall vs. asymptomatic patients only). We then built a spatial regression model to assess the role of each neighborhood-level factor on SARS-CoV-2 positivity rates within the city of Philadelphia. Our overall strategy is visualized in **Figure 1.** This study was performed as a quality assurance and quality improvement project within the Department of Pathology and Laboratory Medicine in the University of Pennsylvania.

2.1 Obtaining Neighborhood-Level Descriptors

Table 1 includes the neighborhood-level factors included in our model along with each source. All of our 'neighborhoods' consist of census tracts. Therefore, for the purpose of this study, a neighborhood is a census tract. We chose census-tract characteristics to inform our models regarding the neighborhood and communities living situations within that census tract.

2.1.1 American Community Survey

We queried the United States Census Bureau website for data on neighborhood-level factors (at the census-tract level) for inclusion in our models. We then queried the US Census query interface called CEDSCI (Center for Enterprise Dissemination Services and Consumer Innovation) accessible at: https://data.census.gov/cedsci/. We downloaded all data for 2010-2017 and only included 2017 data in our modeling of the neighborhood-level associations with COVID-19, to focus on the most recent survey data. Specific data file names are given in **Table 1**. Although, these data are available for all US census tracts, we restricted our datasets to Philadelphia-only census tracts to study the relationship between neighborhood-level factors and SARS-CoV-2 positivity rates within this most immediate urban catchment area for Penn Medicine.

2.1.2 OpenDataPhilly

We incorporated OpenDataPhilly data accessible at https://www.opendataphilly.org/ for information on housing quality. This housing dataset contains information on buildings and units that have been cited by law enforcement and/or city regulations and inspections officials for violations to housing quality laws. We integrated general violations data from the Licenses and Inspections violations, obtainable at: https://www.opendataphilly.org/dataset/licenses-and-inspections-violations. We also only used the general violations category to ensure that we had data across as many Philadelphia census tracts as possible for the year 2017. For information on the violent and non-violent crime rates, we included data previously obtained by Dr. Balocchi [9] that was obtained from the Philadelphia Police Department, available on https://www.opendataphilly.org/dataset/crime-incidents. To be consistent with the ACS data, we only included data from 2017.

Table 1. Sources and Data Files for Neighborhood-Level Factors Included in Our Spatial Regression Model

Neighborhood-Level Factors	Source	Data File
Prop. of women aged 15-50 years in each census tract below 100 percent poverty level	ACS	S1301
Prop. of women aged 15-50 years in each census tract that graduated high school (including equivalency)	ACS	S1301
Prop. of women aged 16-50 years in each census tract that are in the labor force	ACS	S1301
Prop. of women aged 15-50 years in each census tract that received public assistance income in the past 12 months	ACS	S1301
Prop. of occupied housing units in each census tract that are owner-occupied	ACS	S2502
Prop. of occupied housing units in each census tract that are renter-occupied	ACS	S2502
Median family income (dollars)	ACS	S1903
Prop. of each census tract that identifies as Asian Alone	ACS	B01001D
Prop. of each census tract that identifies as Black or African-American	ACS	B01001B
Prop. of each census tract that identifies as Hispanic or Latinx	ACS	B01001I
Prop. of each census tract that identifies as White Alone	ACS	B01001A
Housing Violations	OpenDataPhilly	
Violent Crime Rate	OpenDataPhilly	
Non-Violent Crime Rate	OpenDataPhilly	

2.2 SARS-CoV-2 Positive Patients Broken Down by Symptomatic Status

We identified whether patients being tested for SARS-CoV-2 were symptomatic based on documentation in the test's ordering questions, as stored in EHR. This allowed us to investigate all patients grouped together regardless of symptomatic status (comparison 1) and also only asymptomatic patients (comparison 2). Asymptomatic patients receiving a SARS-CoV-2 test did not report any specific signs or symptoms of COVID-19 (e.g., fever, cough, loss of taste or smell) and were either tested due to presumed or confirmed exposure (e.g., living with a COVID-19 positive individual), screening prior to clinical care (e.g. surgical procedure, hospital admission, pregnancy labor & delivery), the result of contact tracing efforts. We were able to perform this stratification because we had access to documentation of symptomatic status.

2.3 Statistical Analysis: Spatial Regression Model of Effect of Neighborhood-Level Factors on SARS-CoV-2 Positivity Status

First, we performed a log-transformation of housing violation data, income, and both violent and non-violent crime to normalize them. Next, we evaluated two models: one to predict SARS-CoV-2 positivity rate at the neighborhood-level for all patients and the second model to predict SARS-CoV-2 positivity rate among asymptomatic patients only. All analyses were performed using R statistical software (version 3.6.1). We used the R spautolm function from spdep package to perform spatial autoregression, fit using Maximum Likelihood estimation. This method is optimized for sparse matrices, as observed in our dataset. Our outcome was the % SARS-CoV-2 positive out of total conclusive test results (i.e., Number of Positive / (Number of Positive + Number of Negative)) for each census tract. After univariable results were obtained, we then explored multi-predictor models and included all nominally significant ($p <= 0.05$) results into a single spatial regression model. This allowed us to assess the importance of each nominally significant neighborhood-level predictor while accounting for the other factors in an adjusted model.

3. Results

3.1 Penn Medicine Cohort of Individuals Tested for SARS-CoV-2

Our population consists of 46,001 unique individuals tested for SARS-CoV-2 between March 3, 2020 and June 9, 2020. We geocoded patient addresses using ArcGIS and mapped the corresponding latitude and longitude coordinates to census tracts within the city of Philadelphia. We observed 19,281 distinct patients lived within the city of Philadelphia and 26,824 distinct test results that could be linked to census tract information for analysis. Several patients were tested multiple times.

Most of our patients had a documented symptomatic status (i.e., asymptomatic or symptomatic) allowing us to stratify our population and perform subanalyses among asymptomatic individuals only. **Table 2** shows the positivity rates by symptomatic status, revealing a rate of 10.1% SARS-CoV-2 positivity among asymptomatic individuals as compared to 20.7% positivity rate for all patients during this time frame. For simplicity, we excluded rare inconclusive and non-discrete results from our statistical analysis.

Table 2. SARS-CoV-2 Positivity Rates by Symptomatic Status, Penn Medicine March-June 2020

SARS-CoV-2 Test Result	All Patients	Known Asymptomatic
Positive	4781 (20.73%)	680 (10.14%)
Negative	18278 (79.27%)	6023 (89.86%)
Total Positive or Negative Result	23059	6703
Total Completed Results	26824	7395

3.2 Neighborhood-Level Associations with SARS-CoV-2 Positivity Status: Overall

3.2.1 Single Predictor Spatial Regression Model

We performed single predictor or univariable analysis to obtain results for each variable and its relationship with SARS-CoV-2 positivity (**Table 3).** All neighborhood-level predictors were associated ($p < 0.001$) with the exception of four, the proportion of those identifying as Asian living in a census-tract, the non-violent crime rate, and the proportion of occupied housing units that was either renter occupied or owner occupied. Interestingly, three neighborhood-level factors were associated with *lower* than expected SARS-CoV-2 positivity: median family income, proportion of those identifying as White living in a census-tract, proportion of women aged 16-50 that participated in the labor force (i.e., women who were working). In other words as these three factors (i.e., income, White, labor force participation) increased in a given neighborhood the rate of SARS-CoV-2 positivity went down.

3.2.2 Multi-Predictor Spatial Regression Model

We next constructed a multi-predictor spatial regression model using all 10 statistically significant neighborhood-level factors from the single predictor models (**Table 4).** We report the goodness-of-fit statistics for our multi-predictor model, including lambda=-0.0707, Hessian Standard Error of lambda=0.0842, log-likelihood=372.12 and log likelihood ratio of 0.6665 with a corresponding p-value=0.4143. Because the p-value of the log-likelihood ratio >=0.05, we can conclude that spatial clustering of SARS-CoV-2 positivity rates that may appear in our data has been accounted for by the factors included in this multi-predictor model with results shown in **Table 4**[10]. We visualized the SARS-CoV-2 positivity rates in Philadelphia by census tract (**Figure 2)** and compared this with one of the nominally significant variables that is potentially actionable, the proportion of women 15-50 receiving public assistance (**Figure 3).**

Table 3. Results of Single Predictor Spatial Regression Analyses of Neighborhood-Level Factors on Neighborhood-Level SARS-CoV-2 Positivity Rate for All Patients

Short Name	Neighborhood-Level Factors	Odds Ratio	P-value
Income	Median family **income** (dollars) (log-transformed variable)	0.9144	<0.001
Education	Prop. of women aged 15-50 years in each census tract that **graduated high school** (including equivalency)	1.4471	<0.001
White	Prop. of each census tract that identifies as **White** Alone	0.8601	<0.001
Violent Crime	**Violent Crime Rate** (log-transformed variable)	1.0463	<0.001
Black	Prop. of each census tract that identifies as **Black or African-American** Alone	1.1117	<0.001
Poverty	Prop. of women aged 15-50 years in each census tract below 100 percent **poverty** level	1.2243	<0.001
Public Assistance	Prop. of women aged 15-50 years in each census tract that **received public assistance** income in the past 12 months	1.6357	<0.001
Housing Violations	**Housing Violations** (log-transformed variable)	1.0302	<0.001
Hispanic	Prop. of each census tract that identifies as **Hispanic or Latinx**	1.1853	<0.001
Labor Force	Prop. of women aged 16-50 years in each census tract that are **in the labor force**	0.8162	<0.001
Asian	Prop. of each census tract that identifies as **Asian** Alone	0.96	0.5582
Renter	Prop. of occupied housing units in each census tract that are **renter**-occupied	0.9841	0.5879
Owner	Prop. of occupied housing units in each census tract that are **owner**-occupied	1.0161	0.5879
Non-Violent Crime	**Non-Violent Crime** Rate (log-transformed variable)	1.0017	0.8439

Table 4. Results of Multi-Predictor Spatial Regression Model of Neighborhood-Level Factors on Neighborhood-Level SARS-CoV-2 Positivity Rate for All Patients

Short Name	Neighborhood-Level Factors	Adj. Odds Ratio	P-value
Education	Prop. of women aged 15-50 years in each census tract that **graduated high school** (including equivalency)	1.2114	<0.001
White	Prop. of each census tract that identifies as **White** Alone	0.8283	0.0011
Hispanic	Prop. of each census tract that identifies as **Hispanic or Latinx**	1.1302	0.0027
Public Assistance	Prop. of women aged 15-50 years in each census tract that **received public assistance** income in the past 12 months	0.8136	0.0406
Labor Force	Prop. of women aged 16-50 years in each census tract that are **in the labor force**	1.0942	0.1045
Black	Prop. of each census tract that identifies as **Black or African-American** Alone	0.9421	0.2708
Income	Median family **income** (dollars) (log-transformed variable)	0.9891	0.5592
Housing Violations	**Housing Violations** (log-transformed variable)	1.0027	0.7263
Violent Crime	**Violent Crime Rate** (log-transformed variable)	0.9989	0.9116
Poverty	Prop. of women aged 15-50 years in each census tract below 100 percent **poverty** level	1.0012	0.9816

Four neighborhood-level factors appeared nominally associated ($p < 0.05$) in our multi-predictor model. Two variables were associated with *increased* SARS-CoV-2 positivity rates at the neighborhood-level: the proportion of women 15-50 having at least a high school education and the proportion of individuals identifying as Hispanic/Latinx in that neighborhood. Two neighborhood-level factors were associated with *decreased* SARS-CoV-2 positivity rates: the proportion of individuals identifying as White alone in a given census-tract and the proportion of women aged 15-50 years old receiving public assistance. Of note, according to the Census Bureau public assistance includes both cash and non-cash benefits (e.g., Temporary Assistance for Needy Families or TANF, Supplemental Nutrition Assistance Program or SNAP) to low-income families or individuals [11]. Importantly, the proportion of a census tract that identifies, as Black or African American was no longer associated with SARS-CoV-2 positivity in the multi-predictor model, suggesting other correlated neighborhood-level factors were more informative in SARS-CoV-2 positivity rates.

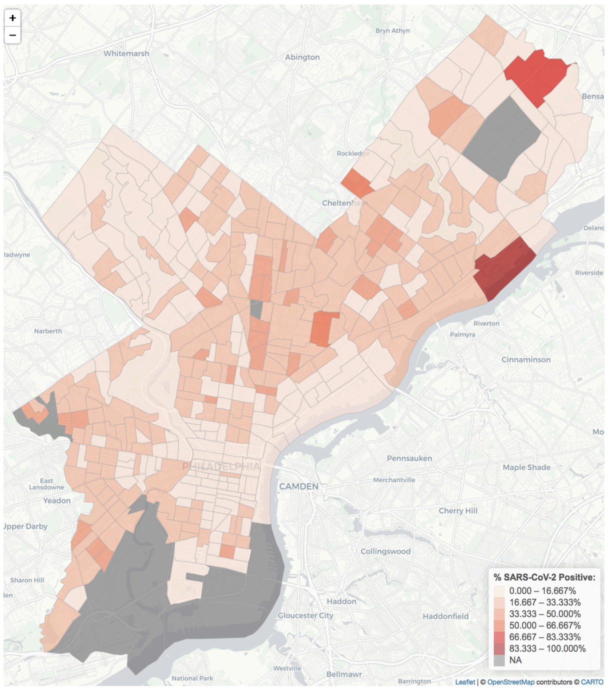

Figure 2. SARS-CoV-2 Positivity Rates (March-June 2020)

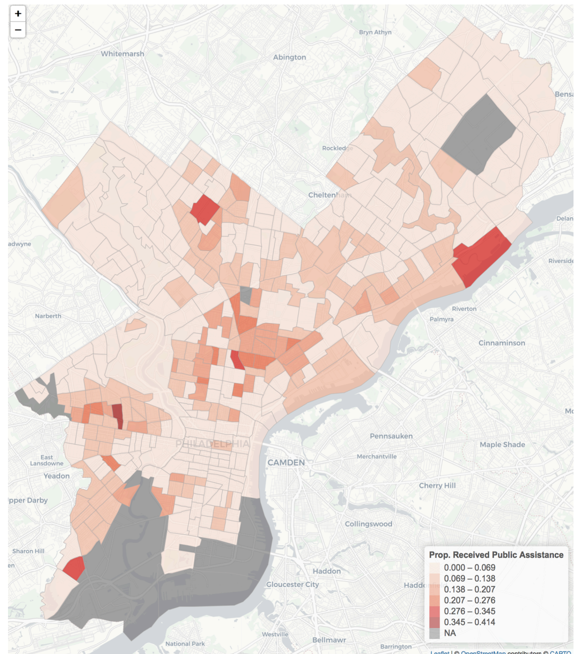

Figure 3. Proportion of Women 15-50 Receiving Public Assistance in 2017 in Philadelphia

3.3 Neighborhood-Level Associations with SARS-CoV-2 Positivity Status: Asymptomatic Only

We conducted a sub-analysis among only patients that were identified as being asymptomatic at the time of testing. We had 7,395 COVID-19 test results that were from asymptomatic patients who live within the city of Philadelphia (**Table 2**) and we focused the patients with confirmed positive or negative COVID-19 test results. The asymptomatic population has a lower overall positivity rate, as expected.

3.3.1 Single Predictor Spatial Regression Model

We performed single predictor (univariable) regression to characterize each variable's relationship with SARS-CoV-2 positivity among asymptomatic patients. We found that eight neighborhood-level factors were associated ($p < 0.05$) with positivity (**Table 5**). Two of these neighborhood-level factors *lower* SARS-CoV-2 positivity rates, including the proportion of the census tract that identified as being of White race and the median family income in a census tract. The remaining six associations were positively correlated, indicating that they were associated with *higher* SARS-CoV-2 positivity rates. Those neighborhood characteristics included, the proportion of individuals identifying as being of Black or African American race, the violent crime rate, the number of housing violations and the proportion of women 15-50 that graduated high school and the proportion of women 15-50 that received public assistance. As in the overall model (which included those that were symptomatic or asymptomatic), we observed that the proportion of housing units that were either renter occupied or owner occupied was not significantly correlated with SARS-CoV-2 positivity rates.

Table 5. Results of Single Predictor Spatial Regression Analysis of Neighborhood-Level Factors on Neighborhood-Level SARS-CoV-2 Positivity Rate: Asymptomatic Only

Short Name	Neighborhood-Level Factors	Odds Ratio	P-value
Black	Prop. of each census tract that identifies as **Black or African-American** Alone	1.0831	<0.001
White	Prop. of each census tract that identifies as **White** Alone	0.9177	<0.001
Income	Median family **income** (dollars) (log-transformed variable)	0.9574	<0.001
Violent Crime	**Violent Crime Rate** (log-transformed variable)	1.0261	<0.001
Housing Violations	**Housing Violations** (log-transformed variable)	1.0216	<0.001
Education	Prop. of women aged 15-50 years in each census tract that **graduated high school** (including equivalency)	1.1716	<0.001
Poverty	Prop. of women aged 15-50 years in each census tract below 100 percent **poverty** level	1.1000	0.0033
Public Assistance	Prop. of women aged 15-50 years in each census tract that **received public assistance** income in the past 12 months	1.2549	0.0108
Non-Violent Crime	**Non-Violent Crime** Rate (log-transformed variable)	1.0138	0.1176
Asian	Prop. of each census tract that identifies as **Asian** Alone	0.9007	0.1274
Labor Force	Prop. of women aged 16-50 years in each census tract that are **in the labor force**	0.9370	0.1371
Owner	Prop. of occupied housing units in each census tract that are **owner**-occupied	0.9554	0.1425
Renter	Prop. of occupied housing units in each census tract that are **renter**-occupied	1.0467	0.1425
Hispanic	Prop. of each census tract that identifies as **Hispanic or Latinx**	0.9928	0.8288

3.3.2 Multi-Predictor Spatial Regression Model

We included all 8 nominally statistically significant neighborhood-level factors from our asymptomatic only univariable models into a multi-predictor model (**Table 6).** Metrics denoting the goodness-of-fit for the model, include lambda=-0.0493, Hessian Standard Error of lambda=0.0860, log-likelihood=291.73 and log likelihood ratio of 0.3234 with a corresponding p-value=0.5695. Because the p-value of the log-likelihood ratio >=0.05, we can conclude that any spatial clustering of SARS-CoV-2 positivity rates that may appear in our data has been accounted for by the factors included in this multi-predictor model [10]. In this model, we observed no neighborhood-level factors that were significantly associated with SARS-CoV-2 positivity. The factor with the lowest p-value (p = 0.05) was the proportion of those identifying as being Black or African American in a census tract. Importantly, because the log likelihood ratio was 0.3234 with a corresponding p-value=0.5695, our model does remove any spatial variability in the SARS-CoV-2 positivity rates (if present) among asymptomatic individuals.

Table 6. Results of Multi-Predictor Spatial Regression Model of Neighborhood-Level Factors on Neighborhood-Level SARS-CoV-2 Positivity Rate: Asymptomatic only

Short Name	Neighborhood-Level Factors	Adj. Odds Ratio	P-value
Black	Prop. of each census tract that identifies as **Black or African-American** Alone	1.1032	0.0505
Poverty	Prop. of women aged 15-50 years in each census tract below 100 percent **poverty** level	1.0653	0.2209
Education	Prop. of women aged 15-50 years in each census tract that **graduated high school** (including equivalency)	1.0586	0.3684
White	Prop. of each census tract that identifies as **White** Alone	1.0499	0.4336
Public Assistance	Prop. of women aged 15-50 years in each census tract that **received public assistance** income in the past 12 months	0.9338	0.5776
Housing Violations	**Housing Violations** (log-transformed variable)	1.0052	0.5809
Violent Crime	**Violent Crime Rate** (log-transformed variable)	1.0040	0.7369
Income	Median family **income** (dollars) (log-transformed variable)	0.9993	0.9769

4. Discussion

4.1 Method Enables Rapid Identification of Neighborhood-Level Factors Linked to SARS-CoV-2 Positivity

Our informatics approach enabled us to identify four neighborhood-level factors associated with the frequency of SARS-CoV-2 test positivity, after adjusting for other correlated neighborhood-level covariates. Two neighborhood-level factors were positively correlated with increased SARS-CoV-2 positivity rates in our adjusted model; conversely, two other factors were negatively correlated with SARS-CoV-2 positivity rates. We discuss the negative or protective associations in section 4.2 and focus here on the neighborhood-level variables that are *positively* associated with SARS-CoV-2 test positivity. The two variables that were associated with higher SARS-CoV-2 positivity rates at the neighborhood-level were: the proportion of women 15-50 having at least a high school education, which is a proxy for education status, and the proportion of individuals identifying as Hispanic/Latinx in that neighborhood (**Table 4**). Patients from neighborhoods with increased proportions of individuals with at least a high school degree appeared more likely to be SARS-CoV-2 positive (adjusted odds ratio [aOR] = 1.2). Neighborhoods with increased proportions of individuals identifying as Hispanic or Latinx showed slightly higher SARS-CoV-2 positivity rates in the adjusted model (aOR=1.130).

It is extremely challenging to translate these apparent associations into causal relationships, because of the complex ascertainment schemas in these data, including the associations between neighborhood factors and symptom status. That said, a study conducted in the Washington DC area also found higher rates of SARS-CoV-2 positivity among Hispanic and Latinx populations [12]. Although generalizations tend to be overly simplistic [13], potential explanations include reduced rates of healthcare insurance and utilization among Hispanic populations [12]. Another factor worth considering is reduced ability to physically distance due to living in tightly crowded communities or inability to stop working due to economic constraints [14]. Finally, there is the added complexity of undocumented immigrants, who within the past year were disenrolled from public assistance programs (i.e., SNAP) [14].

We found that the proportion identifying as Black or African American in a given census tract did not appear independently associated with SARS-CoV-2 positivity in the multivariable model (**Table 4**). However, it was associated in the univariable model (**Table 3**), suggesting that some other neighborhood-level factors were correlated and potentially more informative for SARS-CoV-2 positivity rates. Of note, many other studies that observed associations between Black/African American communities and SARS-CoV-2 positivity [15] did not adjust for other correlated factors that may either increase risk (e.g., poverty) or decrease risk (e.g., public assistance). In addition, our adjusted model including other highly correlated factors, including the proportion of individuals identifying as White alone, which was observed to be associated with *lower* rates of SARS-CoV-2 positivity among neighborhoods (aOR = 0.83).

4.2 Our Method Can Identify Neighborhood-Level Factors Associated with Lower SARS-CoV-2 Positivity Rates in Large Metropolitan Areas

Our method assesses the role of a number of neighborhood-level factors that could impact SARS-CoV-2 positivity rates. We first examined each of these factors at the single variable level to identify those factors that are associated in some way with SARS-CoV-2 positivity. We then included all of the significant variables in a multivariable model to identify a set of variables that appear independently associated with test positivity. We found that two neighborhood-level factors were associated with *lower* SARS-CoV-2 positivity rates. One of these associations has been well established, namely that neighborhoods with a higher proportion of individuals identifying as White is associated with a decreased SARS-CoV-2 positivity rate. Cordes et Castro's also observed an association with reduced SARS-CoV-2 positivity rates among neighborhoods with higher proportions of White individuals analysis in their New York City population [15]. However, their analysis was only conducted at the univariable level and they did not conduct a multivariable model to adjust for other factors. Our multivariable model suggests that race may be a stronger predictor than other associated factors, including median family income and poverty status. Further research is warranted on this topic.

Importantly, our multivariable model found *lower* SARS-CoV-2 positivity rates at the neighborhood-level were associated with higher proportions of women aged 15-50 years old receiving public assistance (aOR=0.8). This is a particularly intriguing finding, because it hints at a potential public health intervention. Individuals receiving these public assistance programs are impoverished, but receipt of government assistantship could enable these individuals to avoid environments that put them at higher risk of infection. This association held even when adjusting for other correlated variables (e.g., poverty, income, education, and various race/ethnicity factors). This finding supports the importance and impact of government and institutional planning initiatives to increase public assistance to at-risk communities.

4.3 Our Approach Can Be Useful for Public Health Intervention Design and Implementation

As we progress through the fall and winter, with further spikes in SARS-CoV-2 positivity rates, public health interventions to mitigate the spread of COVID-19 are crucial. The ability to identify COVID-19 hotspots, both in Philadelphia and across the USA, is critical for the design, development, implementation, and evaluation of these interventions. While contact tracing is an important tool in containing the spread of COVID-19 [16-18], some people remain unable to effectively quarantine or isolate. Therefore, programs to protect people and support contact tracing must be considered. For these programs to be effective, it will be invaluable to understand which areas in a city such as Philadelphia should be targeted. Furthermore, these data can inform where additional testing sites are needed to improve access for those communities that are most at risk.

4.4 Restricting Analysis to Asymptomatic Patients Helps Tease Out Causation from Amongst Associations

Our institution collects information on symptomatology at the time of testing, so we were able to subset our tested population into those with and without reported symptoms. This allowed us to perform two overall sets of analyses, the first involved all tested patients and the second was among only those who were asymptomatic state. Importantly, there are differences in the demographics factors between patients being tested while symptomatic or asymptomatic. Symptomatic patients are coming for care from the communities in the catchment area. On the other hand, asymptomatic patients are enriched for patients receiving specialty care or elective procedures. The asymptomatic population tends to be wealthier and better engaged in the healthcare system. We found 8 neighborhood-level factors appeared associated with SARS-CoV-2 positivity rates among asymptomatic patients. However, when we included all of these factors in a multivariable model there was insufficient evidence of association of any single factor with positivity (**Table 6**). This lack of associations was likely in part due to insufficient power from the lower sample size of asymptomatic subjects and the lower diversity in this population.

4.5 Limitations and Future Work

The limitations of this study include that the data is from March to June of 2020. Inclusion of more recent data would likely provide a richer study of the association of neighborhood-level factors and SARS-CoV-2 infection rates. We used neighborhood-level factors as surrogates for underlying patient factors and thus they cannot fully capture all of the relevant contextual factors for any single individual. This study is also constrained to the neighborhood-level factors collected by the ACS, OpenDataPhilly and other governmental agencies. Other important neighborhood-level factors may have been overlooked; thereby, limiting our results. In addition, our analysis is static and therefore does not account for changes in these factors and their associations over time. Neighborhood-level factors are available at the year-level and do not account for month over month changes in neighborhoods that could be occurring (therefore there is a temporal lag between neighborhood-level factors as reported by the Census Bureau and the true neighborhood-level characteristics). However, this does not limit our approach and those factors, once available, could be easily plugged into our models. Future work will explore the role of the joint relationship between individual-level factors that may contribute to SARS-CoV-2 positivity along with neighborhood-level factors. Finally, as this study was observational, the causal relationships underlying the observed associations could not be directly identified. We performed some multi-variable analyses and investigated the relationship between neighborhood level factors while adjusting for other important neighborhood level factors. This allowed us to identify neighborhood level factors that were strongly associated with SARS-CoV-2 positivity rates in our population while adjusting for other important neighborhood level factors. However, we did not probe any individual factor sufficiently to definitively establish or refute its association with SARS-CoV-2 positivity, especially among other populations where neighborhood-level characteristics may differ.

5. Conclusion

We have developed an informatics approach for integrating multiple neighborhood-level factors from the American Community Survey and opendataphilly.org with SARS-CoV-2 test results. We used these combined set of neighborhood-level factors to assess the spatial association between neighborhood-level factors and SARS-CoV-2 positivity rates amongst all tested patients and asymptomatic tested patients in one large, urban academic healthcare system. We observed that neighborhoods where either the fraction of the population that had graduated highschool education status or the proportion identified as Hispanic/Latinx was higher were associated with higher SARS-CoV-2 positivity rates; conversely, neighborhoods with higher proportions of those identifying as White or receiving public assistance were associated with lower SARS-CoV-2 positivity rates. Overall, we envision that our approach and its results could be used to inform future government and institutional programs by helping to identify causal risk factors (e.g., poverty) underlying COVID-19 'hotspots' and potential interventions (e.g., public assistance programs such as food stamps) that appear to mitigate SARS-CoV-2 spread.

Acknowledgments: We thank the University of Pennsylvania Perelman School of Medicine for generous startup funds and Department of Pathology and Laboratory Medicine for supporting this quality improvement project.

References

1. WHO. WHO Coronavirus Dashboard. <https://covid19whoint/>. 2020;Accessed on August 25, 2020.
2. Wright AL, Sonin K, Driscoll J, Wilson J. Poverty and economic dislocation reduce compliance with covid-19 shelter-in-place protocols. University of Chicago, Becker Friedman Institute for Economics Working Paper. 2020(2020-40).
3. Koh D. Occupational risks for COVID-19 infection. Occupational medicine (Oxford, England). 2020;70(1):3.
4. Gelatt J. Immigrant Workers: Vital to the US COVID-19 Response, Disproportionately Vulnerable. Washington, DC: Migration Policy Institute. 2020.
5. Shai D. Income, housing, and fire injuries: a census tract analysis. Public health reports. 2006;121(2):149-54.
6. Brown EJ, Polsky D, Barbu CM, Seymour JW, Grande D. Racial disparities in geographic access to primary care in Philadelphia. Health Affairs. 2016;35(8):1374-81.
7. Meeker J, Boland MR, editors. The association between neighborhood level exposures and progression to labor. APHA's 2020 VIRTUAL Annual Meeting and Expo (Oct 24-28); 2020: American Public Health Association.
8. Moore JH, Barnett I, Boland MR, Chen Y, Demiris G, Gonzalez-Hernandez G, et al. Ideas for how informaticians can get involved with COVID-19 research. Springer; 2020.
9. Balocchi C, Jensen ST. Spatial modeling of trends in crime over time in Philadelphia. The Annals of Applied Statistics. 2019;13(4):2235-59.
10. Brunsdon C, Comber L. Chapter 12 CALIBRATING SPATIAL REGRESSION MODELS IN R. Code for An Introduction to Spatial Analysis and Mapping in R 2nd edition: bookdown.org; 2019.
11. CensusBureau. Program Income and Public Assistance. 2020;<https://www.census.gov/topics/income-poverty/public-assistance.html -:~:text=Public%20assistance%20includes%20cash%20and,low%2Dincome%20families%20or%20individuals.&text=This%20report%20presents%20data%20on,ACS)%201%2Dyear%20estimates.>(Accessed in August 2020).
12. Martinez DA, Hinson JS, Klein EY, Irvin NA, Saheed M, Page KR, et al. SARS-CoV-2 positivity rate for Latinos in the Baltimore–Washington, DC Region. JAMA. 2020;324(4):392-5.
13. Weinick RM, Jacobs EA, Stone LC, Ortega AN, Burstin H. Hispanic healthcare disparities: challenging the myth of a monolithic Hispanic population. Medical care. 2004:313-20.
14. Page KR, Venkataramani M, Beyrer C, Polk S. Undocumented U.S. Immigrants and Covid-19. New England Journal of Medicine. 2020;382(21):e62.
15. Cordes J, Castro MC. Spatial analysis of COVID-19 clusters and contextual factors in New York City. Spatial and Spatio-temporal Epidemiology. 2020;34:100355.
16. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health. 2020.
17. Cho H, Ippolito D, Yu YW. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv preprint arXiv:200311511. 2020.
18. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. Swiss medical weekly. 2020;150(11-12):w20225-.