Final

Distribution Agreement

Signature:

________________________________ _______________

Allie Vega Date

Investigating Census Tract Level Disparities and Race as a Pre-Cursor to Invasive Group B *Streptococcus* Infections in Adults 18+ in the 20-county Metropolitan Statistical Area of Georgia using a Neighborhood Deprivation Index

By

Allie Vega

Master of Public Health

Epidemiology

Social Determinants of Health Certificate

____
Scott Fridkin, MD
Committee Chair

____
Michael Kramer, PhD
Committee Member

____
Stepy Thomas, MSPH
Committee Member

Investigating Census Tract Level Disparities and Race as a Pre-Cursor to Invasive Group B *Streptococcus* Infections in Adults 18+ in the 20-county Metropolitan Statistical Area of Georgia using a Neighborhood Deprivation Index

By

Allie Vega

Bachelor of Science
George Mason University
2018

Thesis Committee Chair: Scott Fridkin, MD

An abstract of
A thesis submitted to the faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in
Epidemiology
2022

ABSTRACT

Investigating Census Tract Level Disparities and Race as a Pre-Cursor to Invasive Group B *Streptococcus* Infections in Adults 18+ in the 20-county Metropolitan Statistical Area of Georgia using a Neighborhood Deprivation Index
By Allie Vega

Invasive Group B Streptococcus (GBS) infections continue to increase as a threat of morbidity and mortality to non-pregnant adults. Risk factors for invasive GBS infections include older age, diabetes, and renal failure. Investigation of incidence rates have found black persons as having a higher incidence of disease than other races. There is an interest to understand how race interacts with disparities that may lead to invasive GBS incidence. Literature produced regarding social determinants such as socioeconomic status that lead to adverse health outcomes stress on the necessity to correctly analyze these variables to understand the holistic effect of depravation on health. The purpose of this study was to analyze invasive GBS data derived from the Georgia Emerging Infections Program and utilize a principal component analysis with a Neighborhood Deprivation Index to understand if there was a relationship between deprivation at the census tract level, and if this relationship was present, if it was modified by race. A retrospective cohort study with correlating ACS 5-year estimates census tract data to determine rates of incidence was performed using Georgia Emerging Infection Program MSA database information collected from Active Bacterial Core Surveillance Case Report Forms. Principal component analysis was used to create a Neighborhood Deprivation Index for the census tracts in the 20-county MSA area. Poisson regression was utilized for model development. The magnitude of the Neighborhood Deprivation Index differed by race and was found to be statistically significantly associated with increased rates of invasive GBS. The data from this study demonstrated that there is a statistically significant relationship between the Neighborhood Deprivation Index (NDI) and increase in invasive GBS incidence rates, and that race significantly modifies this relationship among non-Hispanic White race, Asian race, Pacific Islander / Hawaiian race, and American Indian / Alaska Native race.

Investigating Census Tract Level Disparities and Race as a Pre-Cursor to Invasive Group B *Streptococcus* Infections in Adults 18+ in the 20-county Metropolitan Statistical Atlanta area using a Neighborhood Deprivation Index

By

Allie Vega

Bachelor of Science
George Mason University
2018

Thesis Committee Chair: Scott Fridkin, MD

A thesis submitted to the faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in
Epidemiology
2022

ACKNOWLDEGEMENTS

First and foremost, I would like to express my gratitude and thanks to Dr. Scott Fridkin, Dr. Michael Kramer, and Stepy Thomas for their utmost support and insight throughout the process of my thesis. I would also like to recognize and thank the Department of Epidemiology at the Rollins School of Public Health for the educational experience I’ve had to equip me with the integral skills and resources to complete this thesis.

I would like to extend a special thank you to my colleagues at the Georgia Emerging Infections Program, especially the ABCs team, for your encouragement, input, and constant support. In addition, a special thank you to my wonderful friends at Rollins. I am so excited to see where your public health paths take you.

Finally, I could not have done this without the constant love and support from my mother, Lisa Vega, and my sister, Abigail Vega. To my father, Albert Vega, I know you have guided me through this process in your own way, and that you will continue to watch over me. I dedicate this thesis to you all.

DISCLAIMER

The primary dataset used in this project was collected by the Georgia Emerging Infections Program (GAEIP). The GAEIP was not involved in the analyses presented in this thesis.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER I: BACKGROUND

Social Determinants of Health & The Neighborhood Deprivation Index

Health disparities are defined as adverse effects and greater obstacles to health, overall improvement, and well-being that individuals experience as a result of their racial or ethnic identity, socioeconomic status, gender, age, sexual orientation or gender identity, disabilities, geographic location, or other characteristics historically linked to systematic racism, discrimination, or exclusion, that could otherwise be considered avoidable[9]. The relationship between these variables and health outcomes is quite complex in nature and predominately affects Black and Hispanic populations[13]. Proper measurement or analysis of these inequities and the ability to measure deprivation against privilege is essential to understanding the relationship between disparity and health outcome in vulnerable populations. While there has been a larger inclusion of these parameters in research as of late, there is a delay in the ability to properly quantify deprivation parameters in order to understand how socioeconomic and health disparities act as a root cause of predisposition to poor health outcomes[14], emphasizing the importance of selecting an appropriate model and method to analysis to understanding health outcomes of interest.

Recent research regarding social determinants of health, such as health disparities, systemic racism, and outcomes of interest as a result of these determinants, has been largely framed by hypotheses involving conceptual models, many of which examine effects of genetic differences, socioeconomic differences, environmental impact, or behavioral differences to guide the foundation of research to better understand the effects of Social Determinants of Health on adverse health outcomes[10, 11]. Successful models examining structural racism and its impact on health outcomes have used the following parameters collectively: residential housing patterns to examine occurrences of racial segregation within residences with redlining occurrences (mortgage lending discrimination based on racial neighborhood composition)[15], proportions of black and white residents in each state who: were registered to vote, voted, employed, in executive or managerial positions, in professional specialties, who had obtained a bachelors level degree or higher at >25 years old, who were incarcerated, felony disenfranchised, or serving a death sentence[17]. The Reactions to Race variables from the Behavioral Risk Factor Surveillance Survey analyzed perceived racism in different settings and also surveyed concurrent health behaviors[7.] Several other models incorporated cross sectional surveys performed to analyze scores of responses based on unfair treatment at work, housing, or police encounters[14]. Based on prior examples, health deprivation and health effects resulting from social injustice are quantifiably immeasurable by observing measures such as “race”, or “ethnicity” alone, and it is necessary to encompass the entirety of what may contribute to health disparities in analyses to determine what may predispose individuals to adverse health outcomes.

A large indicator of effects on health regarding behavioral and environmental social determinants are those that cluster at the neighborhood level [16, 25]. Many models examining the structural and environmental differences that may present the relationship between health disparities on health outcomes use data derived from the US Census bureau[12]. These effects can be defined using census indicators of deprivation, such as poverty (defined at the state level), unemployment, education levels, income, and median rent[18]. Previous literature shows that living with deprivation indicators such as these has been associated with risk associated behaviors such as gambling, alcohol consumption while pregnant, as well as health outcomes such as cardiovascular disease, AIDS incidence, excess mortality[4], obesity, and diabetes[20]. Census data analyzed individually is limiting for several reasons. There are commonly high correlations between variables, and frequently, variables come with their own definitions, such as poverty. Therefore, research using census tract data has predominately been inconsistent with variable choice and is often limited because of choice and availability of data[4]. Therefore, it is necessary to construct an appropriate analytical approach that incorporates a multilevel deprivation index to appropriately assess the relationship between deprivation and a health outcome of interest[4].

A multilevel deprivation index analysis with principal component analysis satisfies the necessity to properly quantify data to analyze complex socioeconomic disparities at the census level. *Messer, et. al.* Developed a reproducible neighborhood deprivation index that utilizes consistently available data from the U.S Census to investigate what specific locations may be affected by more adverse health outcomes.

Group B Streptococcus

An adverse health outcome of specific interest is the incidence of invasive bacterial infections caused by Group B *Streptococcus* (*Streptococcus* agalactiae). Invasive bacterial infections occur when a bacterial invades and proliferates in parts of the body that are normally sterile, such as blood, cerebrospinal fluid, or in bone[23, 24]. Despite advances in treatment, invasive Group B *Streptococcus* (GBS) infections are still associated with considerable morbidity and mortality[26]. Invasive GBS has historically presented itself as a leading neonatal sepsis pathogen[3], however, invasive GBS disease in nonpregnant adults accounts for a new significant disease burden[2]. GBS presents predominately among older persons, black persons, and adults with diabetes[2], as non-pregnant adults now account for 90% of cases of invasive GBS (CDC). Rates of invasive GBS in non-pregnant adults have increased from 4.2 to 7.2 cases per 100,000 nonpregnant adults over the last 2 decades[6], and to 9.13 in 2021[27]. Incidence of GBS differs by racial groups in both adults and infants; black populations have double the incidence compared to their white counterparts [2, 22, 21] . According to previous literature, race is a probable factor leading to socioeconomic disparities that have an impact on invasive disease disparities[21].

Invasive GBS Analysis and Surveillance

Invasive Group B *Streptococcus* became a pathogen of interest to the Centers for Disease Control in the 1970's when it surpassed Staphylococcus aureus as the most prevalent cause of neonatal sepsis[1]. The Centers for Disease Control and Prevention's response to the several gaps in knowledge of severe increase in invasive bacterial disease burden, including Group B *Streptococcus* (GBS), was to establish a system for active, population-based for the surveillance of invasive bacterial infections, and thus The Emerging Infections Program, and the Active Bacterial Core Surveillance (ABCs) program was created in 10 states[1]. The Georgia Emerging Infections Program (GA EIP) is one of these states, and exists through a collaborative partnership between the Georgia Department of Public Health, the Atlanta Veterans Affairs Health System, the Emory University School of Medicine, and the Centers for Disease Control and Prevention. ABCs performs active laboratory-based and population-based surveillance on 6 different bacterial pathogens that cause meningitis, pneumonia, and bacteremia, including invasive GBS. The ABCs team identifies cases defined as isolation of [Group B *Streptococcus*] from a usually sterile site in a resident of the Georgia defined surveillance areas[1], known as the 20-county Metropolitan Statistical Area (MSA) of Atlanta. GBS case ascertainment occurs through regular queries of all laboratories caring for residents in the surveillance area, with detailed medical record reviews to collect clinical and demographic data.

Previous methods to analyze the relationship between race, socioeconomic parameters, and invasive bacterial infections were performed with a variety of approaches. Examples include using US census data at the state and county level to retrieve individual SES variables such as poverty, crowding, and unemployment, analyzed with several different with Poisson regression models per variable[8], multivariable logistic regression based on phone survey responses to calculate population attributable risk to determine racial differences in prevalence[19], and Chi-square (Fisher's exact) linearity tests investigating the linear relationship between invasive GBS incidence to race (black, white, other).

CHAPTER II

Investigating Census Tract Level Disparities and Race as a Pre-Cursor to Invasive Group B *Streptococcus* Infections in Adults 18+ in the 20-county Metropolitan Statistical Area of Georgia using a Neighborhood Deprivation Index

Allie Vega, Scott Fridkin, Michael Kramer, and Stepy Thomas

ABSTRACT

Invasive Group B Streptococcus (GBS) infections continue to increase as a threat of morbidity and mortality to adults. Risk factors for invasive GBS infections include older age, diabetes, and renal failure. Investigation of incidence rates have found black persons as having a higher incidence of disease than other races. There is an interest to understand how race interacts with disparities that may lead to invasive GBS incidence. Literature produced regarding social determinants such as socioeconomic status that lead to adverse health outcomes stress on the necessity to correctly analyze these variables to understand the holistic effect of depravation on health. The purpose of this study was to analyze invasive GBS data derived from the Georgia Emerging Infections Program and utilize a principal component analysis with a Neighborhood Deprivation Index to understand if there was a relationship between deprivation at the census tract level, and if this relationship was present, if it was modified by race. A retrospective cohort study with correlating ACS 5-year estimates census tract data to determine rates of incidence was performed using Georgia Emerging Infection Program MSA database information collected from Active Bacterial Core surveillance Case Report Forms. Principal component analysis was used to create a Neighborhood Deprivation Index for the census tracts in the 20-county MSA area.

Poisson regression was utilized for model development. The magnitude of the Neighborhood Deprivation Index differed by race and was found to be statistically significantly associated with increased rates of invasive GBS. The data from this study demonstrated that there is a statistically significant relationship between the Neighborhood Deprivation Index (NDI) and increase in invasive GBS incidence rates, and that race significantly modifies this relationship among non-Hispanic White race, Asian race, Pacific Islander / Hawaiian race, and American Indian / Alaska Native race.

INTRODUCTION

Invasive Group B *Streptococcus* (GBS) has transitioned from a pathogen that predominately caused sepsis in newborns, to a pathogen that now causes high rates of mortality and morbidity among non-pregnant adults (≥18)[2, 5, 26]. Known risk factors for invasive GBS among non-pregnant adults include diabetes, renal failure, and older age[3]. Higher incidence among black persons has also been reported[2,3]. While risk factors and racial disparities among incident invasive GBS cases have been established by looking at state level and county level data, little has been investigated on the effect of socioeconomic disparities and the relationship they have with existing racial disparities at the census level.

To effectively evaluate the relationship of racial or socioeconomic disparities with adverse health outcomes, advanced social epidemiology techniques must be utilized in order to fully encompass the presence or absence of a negative effect on health outcomes. One weakness epidemiological research has encountered with trying to assess this relationship in the past, is the lack of consistency in variables, especially at the census level, or lack of quality of measurement in the variables chosen to analyze[4]. A reproducible method within social epidemiology that can be used to measure disparity is one that uses a deprivation index based on neighborhood level parameters such as poverty or other factors that have notoriously been linked to socioeconomic disparities[5].

The Neighborhood Deprivation Index proposed by *Messer et al.* is a method that involves creating a disparity index based on census level data to introduce the factor of socioeconomic deprivation into a study, which can then be analyzed with modifying variables or used in linear regression to observe a potential relationship. The NDI includes census tract data of those who lack a high school level education, are unemployed, households who make less than $35,000/year, who are living in overcrowded housing, those who receive public assistance, individuals who are living below the poverty line, households with a single parent head, and households with a single female head[5].

The goal of this analysis is to determine if there is a relationship between neighborhood level deprivation based on census tract data and the incident count of invasive GBS per census tract, and if this relationship is at all modified by the different races presented in the study.

METHODS

Study Design

This retrospective cohort study of residents of GA EIPs MSA utilized a principal component analysis to create a composite indicator of deprivation (Neighborhood Deprivation Index) from multiple variables derived from ACS 5-Year Estimates of Census Tract data, to test the hypothesis that greater Neighborhood Deprivation was correlated with higher invasive Group B *Streptococcus* incidence, and that this relationship may have been modified by race. This study design was based on an NDI development model by *Messer, et. al.* and an NDI model created by *Dr. Michael Kramer*, at the Rollins School of Public Health, Emory University (2022).

Data Source

Data from active population-based and laboratory-based surveillance collected by the Georgia Emerging Infections Program's Active Bacterial Core Surveillance team was used in this study. Case were included from January 1st, 2017, to December 31st, 2019, for residents of the 20-County Metropolitan Statistical Area of Atlanta. The 20 county MSA area includes the following counties: Barrow, Bartow, Carroll, Cherokee, Clayton, Cobb, Coweta, Dekalb, Douglas, Fayette, Forsyth, Fulton, Gwinnett, Henry, Newton, Paulding, Pickens, Rockdale, Spalding, and Walton County. Data was provided by the GA EIP and included of all positive cases of invasive GBS over the specified time period.

An incident case of invasive GBS was defined as "Group B *Streptococcus* isolated from a normally sterile site, such as blood, cerebrospinal fluid, pleural fluid, peritoneal fluid, pericardial fluid, bone, joint/synovial fluid, or internal body site (e.g., lymph node, brain), or pathogen-specific nucleic acid must be detected in a specimen obtained from a normally sterile body site, using a validated molecular test, and case patient must be a resident of one of the defined surveillance areas" (CDC, 2022). Active Bacterial Core Surveillance Case Report Forms were used to document cases. Case Report Forms assigned each case a "State ID", and captured information on cases including full residential address, county, census tract, birthdate, age, sex, race, ethnicity, culture date, sterile site isolation, treatment hospital, dates hospitalized, ICU status, transfer status, height, weight, residence at time of initial culture, type of insurance, patient outcome (survival or death), discharge location, pregnancy and fetal status, types of infection present with organism, underlying conditions, and substance use, which translated into variables for analysis.

The total number of incident cases that were within the MSA area and that occurred between the selected time period were *n = 1271*. Cases with full residential location information were geocoded using 2010 census data via Georgia EIP. Cases who lacked census tract information and who were not geocoded were excluded from the study, *n = 19*. Geocoded cases had a match rate of 98.5%. A total of *n = 1251* patients were eligible for inclusion in the study based on geocoding. Patients that lacked race or ethnicity information (both race and ethnicity were coded as 'unknown') were also excluded from the study, *n = 19,* and 1 recurrent case was excluded. A total of *n = 1232* included for analysis. Incidence cases that fit inclusion criteria into the study were cases defined as an adult (≥18 years of age), and had a culture date between January 1st, 2017, and December 31st, 2019.

Census tract data from American Community Survey 5-year estimates (2015-2019) was downloaded and included as a population base to calculate incidence rates over the time period of interest. Census tract data included overall population per tract, and population stratified by race. Census tract data was merged into the GAEIP ABCs SAS dataset by case census tract from Case Report Form input.

Principal Component Analysis & Neighborhood Deprivation Index

Principal Component Analysis was utilized in SAS to create a Neighborhood Deprivation index from the following variables, derived from Census American Community Survey 5-year Estimates (2015-2019) for the census tracts in the 20-county MSA area. These variables were identified by *Messer et al (2006)* as census level indicators of socioeconomic status that represent social conditions that have an association with health outcomes.

Variable	Description[1]
pctNOHS	% of adults over 25 without a high school degree or GED equivalent. Scaled as a fraction from 0-1.
pctUNEMP	% of people over 16 who are in the labor force, who are also employed. Scaled as fraction from 0 to 1.
pctHHINCsub35k	% of households with total income below $35,000/year. Scaled as fraction from 0 to 1.
pctOVERCROWDHH	% of households with > 1 person per bedroom on average. Scaled as fraction from 0 to 1.
pctOCCUPATION_MANAGE	% employed people with jobs in administration and management. Scaled as a fraction from 0 to 1.
pctPUBLICASST	% of households receiving public assistance (TANF, WIC, SNAP, etc) of some kind. Scaled as a fraction from 0 to 1.
pctPOV	% of people living below the federal poverty line. Scaled as a fraction from 0 to 1.
pctFHH	% of households with children under 18 with a single parent head. Scaled as a fraction from 0 to 1.
pctSPHH	% of households with children under 18 with single parent head. Scaled as a fraction from 0 to 1.

[1]*Dr. Michael Kramer*, Rollins School of Public Health, Emory University, calculated initial percentages for variables from ACS 5-year estimates using R Studio.

All 8 variables were standardized in SAS. Variables were centered with *mean = 0* and scaled by *std = 1*. Correlation matrices were calculated from the variables to visually examine correlations between the standardized versions of the variables (Figure 1). Eigenvectors and eigenvalues were calculated from PCA (factor analysis) analysis using PROC PRINCOMP in SAS in order to analyze how much variation is contained by each principal component (Figure 2). Principal Component 1 contained 54% of the total variation from the 8 SES variables and had an eigenvalue of 4.28 with a difference of 3.017 from the following eigenvalue for Component 2. Positive loading coefficients were calculated from the eigenvector that represent the strength of association between variables and the principal component and formulated to create the Neighborhood Deprivation Index that described a level of deprivation vs affluence, where more positive values indicate a higher level of deprivation, and more negative levels indicate a higher level of affluency.

Poisson Regression was utilized to analyze the relationship between incidence case counts among the 627 census tracts eligible for inclusion in the MSA area, the Neighborhood Deprivation Index, and the modifying relationship of race. Two models were developed. The first model accounted for the relationship of NDI and case counts. The interaction model included a multi-level "race" variable, which was re-coded from the original data, accounting for all races & ethnicities collected by case report forms and were coded as follows: 0 = non-Hispanic White (reference group), 1 = non-Hispanic Black, 2 = Asian, 3 = American Indian & Alaskan, 4 = Pacific Islander & Hawaiian, and 5 = Hispanic/Latino. The relationship between this race variable and the neighborhood deprivation index was analyzed independently, and with a race*ndi interaction term. The significance of both the independent term, and the interaction term's relationship with case counts was tested with a Type 3 test, and stratum-specific estimates were produced. The population base created for each case was the total population of the case's identified race or ethnicity from their census tract, and a logarithmic offset was created from this total population. All statistical tests were performed using Wald Chi-Square parameter estimates, and *p-values* of <0.05 were considered statistically significant. Data were analyzed using SAS 9.4.

RESULTS

Descriptive Analysis

Overall, 1232 incident cases of invasive Group B *Streptococcus* were included in the study. Of those cases, 40.18% (503) were identified as black, 53.22% (664) were identified as white, 5.35% (67) were identified as Hispanic/Latino, 0.4% (5) were identified as American Indian, 1.36% (17) were identified as Asian, and 0.32% (4) were identified as Pacific Islander. 61 cases who accounted for the Hispanic/Latino portion of the cases were documented as unknown race. 54.22% (668) of the cases were male, and 45.78% (564) were female.

Residential characteristics of cases consisted of predominately private resident types, with >93.6% of all races being documented as private. 2.1% (14) of white cases and 6.4% (32) of black cases were residing in a long-term care facility at the time of culture. Cases per county are displayed in Table 1. Clayton, Coweta, and Fulton County had the most diverse representation of cases. Most counties lacked in cases of multiple races other than black and white. Case counts per census tract had a mean of 2.8 and a standard deviation of 1.6, with a minimum of 1 and a max of 7.

Private insurance was predominately held by all race groups except for black cases. The majority of black cases held Medicare as a form of insurance. Medicaid was held by 15.9% (106) white cases, while the coverage doubled in black cases, at 30.8% (155). 7.9% (53) of white cases were uninsured, and 9.9% (50) of black cases were uninsured. 11.8% (2) of Asian cases, 50% (2) Pacific Islander/Hawaiian cases, and 28.4% (19) of Hispanic/Latino cases held Medicaid as their primary form of insurance, and 38% (22) of Hispanic/Latino cases were uninsured.

Clinical characteristics of cases were that ≥94% of cases across all racial groups required hospitalization. 28.3% (179) of white cases, 33.2% (161) of black cases, 31.3% (5) of Asian cases, 60% (3) of American Indian cases, and 23.8% of (15) Hispanic/Latino cases were admitted to the ICU. Death was an outcome of 4.7% (31) white cases, 5.57% (28) black cases, 20% (1) of American Indian cases, and 8.9% (6) of Hispanic/Latino cases. Diabetes occurred in at least 52.2% of cases across all racial groups, and obesity occurred in 27.8% (184) of white cases, 24.5% (123) of black cases, 17.6% (3) of Asian cases, 80% (4) of American Indian cases, and 59.7% (40) of Hispanic/Latino cases. Additional demographic information can be found in Table 1.

Principal Component Analysis & Neighborhood Deprivation Index

The final formula for the NDI from the first principal component derived from the principal component analysis is shown below:

$$
\begin{aligned}
NDI = \; & 0.345836(pctNOHS) + 0.318857(pctUNEMP) + 0.444923(pctHHINCsub35k) \\
& + 0.261649(pctOVERCROWDHH) - 0.3511148(pctOCCUPATION_{MANAGE}) \\
& + 0.268991(pctPUBLICASST) + 0.434745(pctPOV) + 0.357278(pctFHH)
\end{aligned}
$$

The first principal component was chosen due to its ability to explain the most variation of the census tract deprivation indicators. Because the loading values produced from the eigenvectors and the principal component analysis were positive, except for the census variable that would indicate affluency (occupations in management), the interpretation of the variables are that as they increase, the principal component gets larger, therefore as those values increase, deprivation increases. The overall distribution for the calculated Neighborhood Deprivation Index for the displayed as a right-skewed curve (Figure 3).

Poisson models were created to assess the relationship between the NDI, race, and the count of cases occurring per census tract. Race variables were included in the full Poisson model as interaction terms, and alone in the reduced model to assess effect modification. The models accounted for all races in the study, as they account for all cases in the population in which rates were calculated from. A logarithmic offset was created from the total population of each census tract in which cases occurred.

Results from the Poisson models indicated that the change in the Neighborhood Deprivation Index associated with an increase in incidence rates of invasive GBS was significant, and that interaction between race and the Neighborhood Deprivation Index was present and significant. A 1-unit change in the NDI was associated with a 9.8% increase in the incidence rate of invasive GBS alone in the model (IDR = 1.0979, CI: 1.08 – 1.12). When introducing race into the model, Type 3 joint tests confirmed that race had a significant interaction with the relationship of NDI and invasive GBS incidence rates (NDI, Race, NDI*Race $p < 0.0001$). Contrast estimate statements were produced with White race as the reference group. The effect of the relative change in incidence rates for each 1-unit change in the NDI among Whites was 21.9% (IDR = 1.2187, CI: 1.19 – 1.25). There was no effect on the relative change in incidence rates for each 1-unit change in the NDI among Black race and Hispanic ethnicity compared to their white counterparts (Black IDR = 0.999, CI: 0.9716 – 1.03), (Hispanic IDR = 0.997, CI: 0.88 – 1.12). The effect of the relative change in incidence rates for each 1-unit change in the NDI increased by 45% among the Asian race group (IDR = 1.45, CI: 1.15 – 1.83). Among the American Indian / Alaska native race group, this finding suggests that deprivation is protective against invasive disease; with a decrease in incidence by 17.5% (IDR = 0.825, CI: 0.34 – 1.98), although the extremely wide 95% CI highlights how imprecise this estimate is. Among the Pacific Islander / Hawaiian race group, the effect of the relative change in incidence rates for each 1-unit change in the NDI increased by 21.9% (IDR = 1.219, CI: 1.187 – 1.25). See Table 2.

DISCUSSION

The key takeaway from this study is that there is an indication from the neighborhood deprivation index that socioeconomic deprivation does have a relationship with incidence of invasive GBS disease, and that this relationship is modified among racial groups. There is a strong association with increasing depravity and increase in incidence rate of invasive GBS among non-Hispanic Whites, Asians, and Pacific Islander/Hawaiian race. There was no association with increase in NDI and incidence rates for non-Hispanic Black and Hispanic populations. The foundation of this relationship is important for Social Determinants of Health-based research because the goal of this research is to determine what societal factors may cause or lead to socioeconomic disparities. Within the realm of invasive GBS, other co-morbidities can be analyzed to determine their role within socioeconomic disparities, and if structural components such as food deserts, which have been investigated for their linkage to obesity, or lack of access to healthcare, are what is causing this relationship. It is important that this is emphasized as the purpose of this research.

Principal component analysis with standardized census tract variables was chosen for this study because of the ability of eigenvector, eigenvalue, and principal component generation to explain more variation in data than any census tract variable analyzed independently. Because analyses in the past involving invasive bacterial disease outcomes have analyzed the relationship between socioeconomic variables and disease incidence as independent variables, there was a strong appeal to perform a holistic analysis that accounted for the correlation and natural variation between census tract deprivation variables that independent regression alone does not have the ability to account for. Additionally, because of the previous literature published on invasive Group B *Streptococcus* that indicated there was prevalent disparities among black Americans who faced incidence of invasive GBS, there was an interest to analyze the effect of race on the relationship between deprivation and invasive GBS incidence.

Strengths

This study used data collected from the GA EIP ABCs program. The precise methodology of ABCs case ascertainment and acquisition ensures robust incident rates and case level information. ABCs obtains cases via ongoing active population-based surveillance with robust audits to ensure accurate case acquisition. Case report forms are completed through thorough medical record review by trained surveillance epidemiologists with regular quality control.

Additionally, the principal component analysis that was performed, and the Poisson models utilized to examine the relationship derived from U.S. Census data, are easily replicable. This analysis investigating the relationship between severe, rare diseases could be performed in a different geographical location, or even using a different State's EIP ABCs data. Not only are the data uniformly consistent between EIP locations due to set case definitions and ascertainment methodology, but the designated variables that are pulled from the census bureau that indicate deprivation ensure consistency and can easily be compared between sites, or different metropolitan areas. Census data, if used consistently, is also usually very representative of true living conditions amongst most residents, allowing a strong potential relationship to be formed between socioeconomic variables and an adverse health outcome of interest.

Limitations

There were notable limitations within the study. It was hypothesized that race potentially modified the relationship between invasive GBS case incidence and higher indications of deprivation, and a large percentage of those who were placed in long term care facilities were black (66.7%). ABCs defines a residence of a long-term care facility for any patient that has spent ≥14 days in the facility. therefore, long-term care facility residence indicators may offset the presence of disparity, as a resident may be placed in a LTCF that is not representative of their true residential census tract/neighborhood. If a patient originally resided in a census tract with a high amount of deprivation but was placed on assistive equipment such as a ventilator, the patient would be required to be transferred to a long-term care facility that supported residents on ventilators (or other assistive equipment), which may be located within census tracts that did not have high amounts of deprivation.

Additionally, the short period over which data was selected to use from may have limited the study. Derived from reviewing previous literature, studies performed with ABCs invasive GBS or other bacterial pathogen data typically analyzed data where incident cases occurred over a period of anywhere from 7 to 17 years, and studies with longer time constraints were studies that typically found high amounts of incident invasive GBS case disparity among black Americans. This could have been especially restrictive to this study because the population of interest was already limited to one geographical area, so a longer time period is likely necessary to encompass the full effect of race's relationship with census-level deprivation. Even though the data produced results that the modifying relationship was still significant among black persons, it was expected that the relationship, based on previous literature, would have been more profound. There were under 20 cases of Asian, Pacific Islander/Hawaiian, and American Indian Cases, which made the study less representative of the actual population of the MSA area, which could have introduced selection bias.

Additionally, comorbidities were not analyzed in the study. Because of the nature of the study, and the high dependence on census tract data to analyze socioeconomic and race variables, disease incidence other than invasive GBS was unable to be considered (no census denominator data for diabetes, drug use etc.). A case-control study could be used in the future for patients who were hospitalized at the same time as incident invasive GBS cases to have a stronger foundation for case rates and incidence on other conditions.

REFERENCES

1. Schuchat A, Skoff T, Zell E, Farley, MM, et al. Active Bacterial Core Surveillance of the Emerging Infections Program Network. *Emerging infectious diseases.* 2001;7. 92-9. 10.3201/eid0701.700092.
2. Skoff TH, Farley MM, Petit S, et al. Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990-2007. *Clin Infect Dis*. 2009;1;49(1):85-92. doi: 10.1086/599369. PMID: 19480572.
3. Francois Watkins LK, McGee L, Schrag SJ, et al. Epidemiology of Invasive Group B Streptococcal Infections Among Nonpregnant Adults in the United States, 2008-2016. *JAMA internal medicine*. 2019;*179*(4), 479–488. https://doi.org/10.1001/jamainternmed.2018.7269
4. Messer LC, Laraia BA, Kaufman JS et al. The Development of a Standardized Neighborhood Deprivation Index. J *Urban Health* 83. 2006;1041–1062. https://doi.org/10.1007/s11524-006-9094-x
5. Raabe VN, Shane AL. Group B *Streptococcus* (*Streptococcus agalactiae*). Microbiol Spectr. 2019;7(2):10.1128/microbiolspec.GPP3-0007-2018. doi: 10.1128/microbiolspec.GPP3-0007-2018. PMID: 30900541; PMCID: PMC6432937.
6. Farley MM, Larry JS. Group B Streptococcal Disease in Nonpregnant Adults, *Clinical Infectious Diseases*. 2001;33(51) 556-561, https://doi.org/10.1086/322696
7. Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System, Reactions to Race Variables. 2014. https://www.cdc.gov/brfss/annual_data/2014/pdf/2014calculated_variables_version8.pdf

8. Spicer JO, Thomas S, Holst A, Baughman W, Farley MM. Socioeconomic and racial disparities of pediatric invasive pneumococcal disease after the introduction of the 7-valent pneumococcal conjugate vaccine. Pediatr Infect Dis J. 2014;33(2):158-64. doi: 10.1097/INF.0000000000000025. PMID: 24418837.
9. Healthy people 2020. Office of Disease Prevention and Health Promotion. Accessed March 15th, 2022.
10. Krieger, Nancy. Theories for social epidemiology in the 21st century: an ecosocial perspective, *International Journal of Epidemiology*. 2001;30(4) 668–677.
11. Diez Roux AV. Conceptual approaches to the study of health disparities. *Annu Rev Public Health*. 2012;33:41-58. doi:10.1146/annurev-publhealth-031811-124534
12. Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am J Epidemiol*. 2007;165(8):858-867. doi:10.1093/aje/kwm040
13. Chantarat T, Van Riper DC, Hardeman RR. The intricacy of structural racism measurement: A pilot development of a latent-class multidimensional measure. *EClinicalMedicine*. 2021;40:101092. Published 2021 Aug 13. doi:10.1016/j.eclinm.2021.101092
14. Groos M, Wallace M, Hardeman R, Theall KP. Measuring inequity: a systematic review of methods used to quantify structural racism. *Journal of Health Disparities Research and Practice*. 2018;11(2) 13. Available at: https://digitalscholarship.unlv.edu/jhdrp/vol11/iss2/13

15. Jacoby SF, Dong B., Beard, JH, et al. The enduring impact of historical and structural racism on urban violence in Philadelphia. *Soc Sci Med*. 2017; doi:10.1016/j.socscimed.2017.05.038

16. Sampson RJ, Morenoff JD, Gannon-Rowley T. Assessing Neighborhood effects: social processes and new directions in research. *Annu Rev Sociology*. 2002;28:443–478.

17. Lukachko A, Hatzenbuehler ML, Keyes KM. Structural racism and myocardial infarction in the United States. *Soc Sci Med*. 2014;103:42-50. doi:10.1016/j.socscimed.2013.07.021

18. Pearl M, Braveman P, Abrams B. The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *Am J Public Health*. 2001;91:1808–1814.

19. Flory JH. Joffe M, Fishman NO, Edelstein PH, Metlay JP. Socioeconomic risk factors for bacteraemic pneumococcal pneumonia in adults. *Epidemiol. Infect*. (2009). doi:10.1017/S0950268808001489

20. Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. BMC Med. 2017;19;15(1):131. doi: 10.1186/s12916-017-0901-x. PMID: 28720102; PMCID: PMC5516328.

21. Hadler JL, Vugia D, Bennett NM, et al. Emerging Infections Program Efforts to Address Health Equity. Emerging Infectious Diseases. 2015;21(9):1589-1594. doi:10.3201/eid2109.150275.

22. Schwartz B, Schuchat A, Oxtoby MJ, et al. Invasive group B streptococcal disease in adults. A population-based study in metropolitan Atlanta. JAMA. 1991;266:1112–4. http://dx.doi.org/10.1001/jama.1991. 03470080082034

23. Centers for Disease Control and Prevention. Diminishing racial disparities in early-onset neonatal group B streptococcal disease. MMWR. 2004;53:502–5
24. Centers for Disease Control and Prevention. 2019. Active Bacterial Core Surveillance Report, Emerging Infections Program Network, Group B Streptococcus, 2019. www.cdc.gov/abcs/downloads/SPN_Surveillance_Report_2019.pdf
25. Forthman KL, Colaizzi JM, Yeh HW, Kuplicki R, Paulus MP. Latent Variables Quantifying Neighborhood Characteristics and Their Associations with Poor Mental Health. *Int J Environ Res Public Health*. 2021;18(3):1202. Published 2021 Jan 29. doi:10.3390/ijerph18031202
26. Centers for Disease Control and Prevention. 2022. BactFacts; invasive Group B *Streptococcus*. https://www.cdc.gov/abcs/bact-facts-interactive-dashboard.html
27. Navarro-Torné A, Curcio D, Moïsi JC, Jodar L.Burden of invasive group B Streptococcus disease in non-pregnant adults: A systematic review and meta-analysis. *PLOS ONE*. 2021:16(9), e0258030. https://doi.org/10.1371/journal.pone.0258030
28. Kramer, MR. A model of the Neighborhood Deprivation Index adapted for Emory University. 2022.

TABLES AND FIGURES

Table 1: Demographics and clinical characteristics of incident invasive Group B *Streptococcus* cases obtained from 2017-2019 EIP ABCs surveillance data, stratified by race

	White (n = 654) No./Mean (%/SD)	Black (n = 503) No./Mean (%/SD)	Asian (n = 17) No./Mean (%/SD)	American Indian (n = 5) No./Mean (%/SD)	Pacific Islander / Hawaiian (n = 4) No./Mean (%/SD)	Hispanic/Latino (n = 67) No./Mean (%/SD)
Patient Demographics						
Sex						
Female	292 (43.9)	238 (47.3)	10 (58.8)	1 (20.0)	2 (50.0)	33 (49.3)
Male	372 (56.1)	265 (52.7)	7 (41.2)	4 (80.0)	2 (50.0)	34 (50.7)
Age (years)	62.7 (15.8)	54.72 (16.9)	60.64 (16.5)	45.4 (16.1)	59.5 (22.0)	50.4 (17.6)
Residence Type						
Residential	649 (97.7)	471 (93.6)	17 (100.0)	5 (100.0)	4 (100.0)	66 (98.5.0)
LTCF	14 (2.1)	32 (6.4)	0	0	0	1 (1.5)
Other	1 (0.15)	0	0	0	0	0
County						
Barrow	15 (2.3)	3 (0.6)	0	0	0	0
Bartow	38(5.7)	3 (0.6)	0	0	0	0
Carroll	23 (3.5)	4 (0.8)	0	0	0	0
Cherokee	43 (6.5)	5 (1.0)	0	2 (40.0)	0	0
Clayton	19 (2.9)	41 (8.2)	1 (5.9)	1 (20.0)	0	5 (7.5)
Cobb	81 (12.2)	51 (10.1)	1 (5.9)	0	0	8 (11.9)
Coweta	38 (5.7)	4 (0.8)	1 (5.9)	1 (20.0)	0	2 (2.99)
Dekalb	48 (7.4)	120 (23.9)	1 (5.9)	0	0	7 (10.5)
Douglas	21 (3.2)	26 (5.2)	0	0	0	3 (4.5)
Fayette	14 (2.1)	7 (1.4)	0	0	0	0
Forsyth	24 (3.6)	1 (0.2)	0	0	1 (*25*.0)	2 (2.99)
Fulton	68 (10.2)	126 (25.1)	4 (23.5)	0	2 (50.0)	14 (20.9)
Gwinnett	84 (12.6)	46 (9.2)	9 (52.9)	0	0	23 (34.33)
Henry	29 (4.4)	20 (3.9)	0	1 (20.0)	1 (25.0)	2 (2.99)
Newton	16 (2.4)	19 (3.8)	0	0	0	1 (1.5)
Paulding	34 (5.1)	5 (1.0)	0	0	0	0
Pickens	14 (12.1)	0	0	0	0	0
Rockdale	10 (1.5)	11 (2.19)	0	0	0	0
Spalding	19 (2.9)	9 (1.8)	0	0	0	0

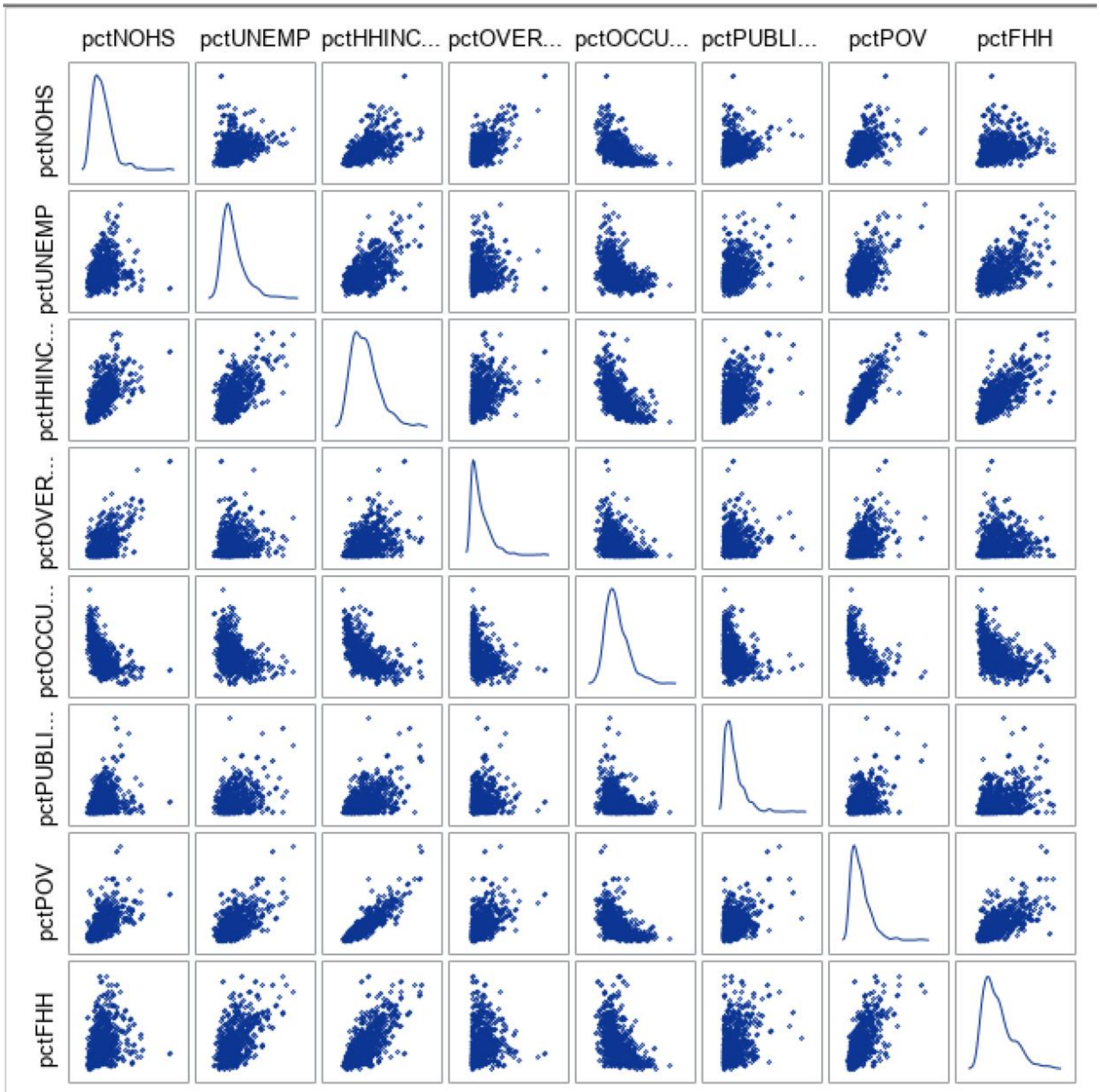
Walton	26 (3.9)	2 (0.4)	0	0	0	0
Insurance Type						
Private	373 (56.1)	195 (38.8)	6 (35.3)	3 (60.0)	2 (50.0)	14 (20.9)
Medicare	53 (7.9)	230 (45.7)	6 (35.3)	1 (20.0)	3 (75.0)	16 (23.9)
Medicaid	106 (15.9)	155 (30.8)	2 (11.8)	0	2 (50.0)	19 (28.4)
Military	21 (3.2)	24 (4.8)	1 (5.9)	1 (20.0)	1 (25.0)	3 (4.5)
Indian Health Service	0	0	0	0	0	0
Prison	0	0	0	0	0	0
Uninsured	53 (7.9)	50 (9.9)	6 (35.3)	1 (20.0)	0	22 (32.8)
Other Insurance	2 (0.3)	1 (0.2)	0	0	0	0
Clinical Characteristics						
Hospitalized	632 (95.2)	485 (96.4)	16 (94.1)	5 (100.0)	4 (100.0)	63 (94.0)
ICU admittance	179 (28.3)	161 (33.2)	5 (31.3)	3 (60.0)	0	15 (23.8)
Outcome - Death	31 (4.7)	28 (5.57)	0	1 (20.0)	0	6 (8.9)
Diabetes mellitus	346 (52.1)	295 (58.7)	9 (52.9)	4 (80.0)	3 (75.0)	40 (59.7)
Obesity	184 (27.7)	123 (24.5)	3 (17.6)	4 (80.0)	0	40 (59.7)

Table 2: Results of Full Poisson Model of Neighborhood Deprivation Index and Race Interaction[2]

Contrast Estimate Results										
Label	**Mean Estimate**	**Mean Confidence Limits**		**L'Beta Estimate**	**Standard Error**	**Alpha**	**L'Beta Confidence Limits**		**Chi-Square**	**Pr > ChiSq**
Increase in iGBS incidence rate among non-Hispanic White race	1.2187	1.187	1.2511	0.1977	0.0134	0.05	0.1715	0.224	217.31	<.0001
Increase in iGBS incidence rate among non-Hispanic Black race	0.9989	0.9716	1.0269	-0.0011	0.0141	0.05	-0.029	0.027	0.01	0.9359
Increase in iGBS incidence rate among Asian race	1.4512	1.1521	1.8279	0.3724	0.1177	0.05	0.1416	0.603	10	0.0016
Increase in iGBS incidence rate among American Indian race	0.8252	0.3436	1.9815	-0.1921	0.447	0.05	-1.068	0.684	0.18	0.6673
Increase in iGBS incidence rate Among Hawaiian/Pacific Islander race	1.2187	1.187	1.2511	0.1977	0.0134	0.05	0.1715	0.224	217.31	<.0001
Increase in iGBS incidence rate Among Hispanic/Latino race	0.9978	0.8849	1.1252	-0.0022	0.0613	0.05	-0.122	0.118	0	0.9716

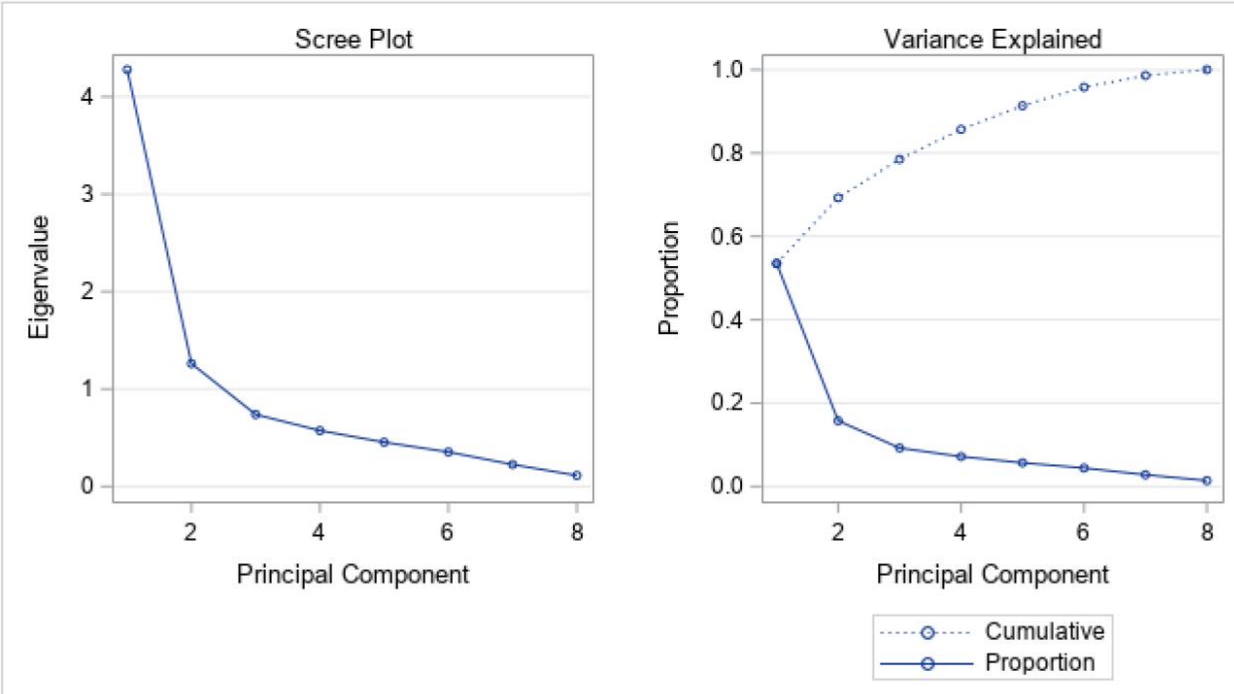
[2]Increase in incidence rate of invasive GBS as a result of a 1-unit change in the Neighborhood Deprivation Index, stratified by racial group to examine effects on the relationship of modification by race. Values over 1.00 indicate an increase in incidence rate associated with an increase in NDI, rates below 1.00 indicate a decrease.

Figure 1: Correlations of Census Tract Data[3]



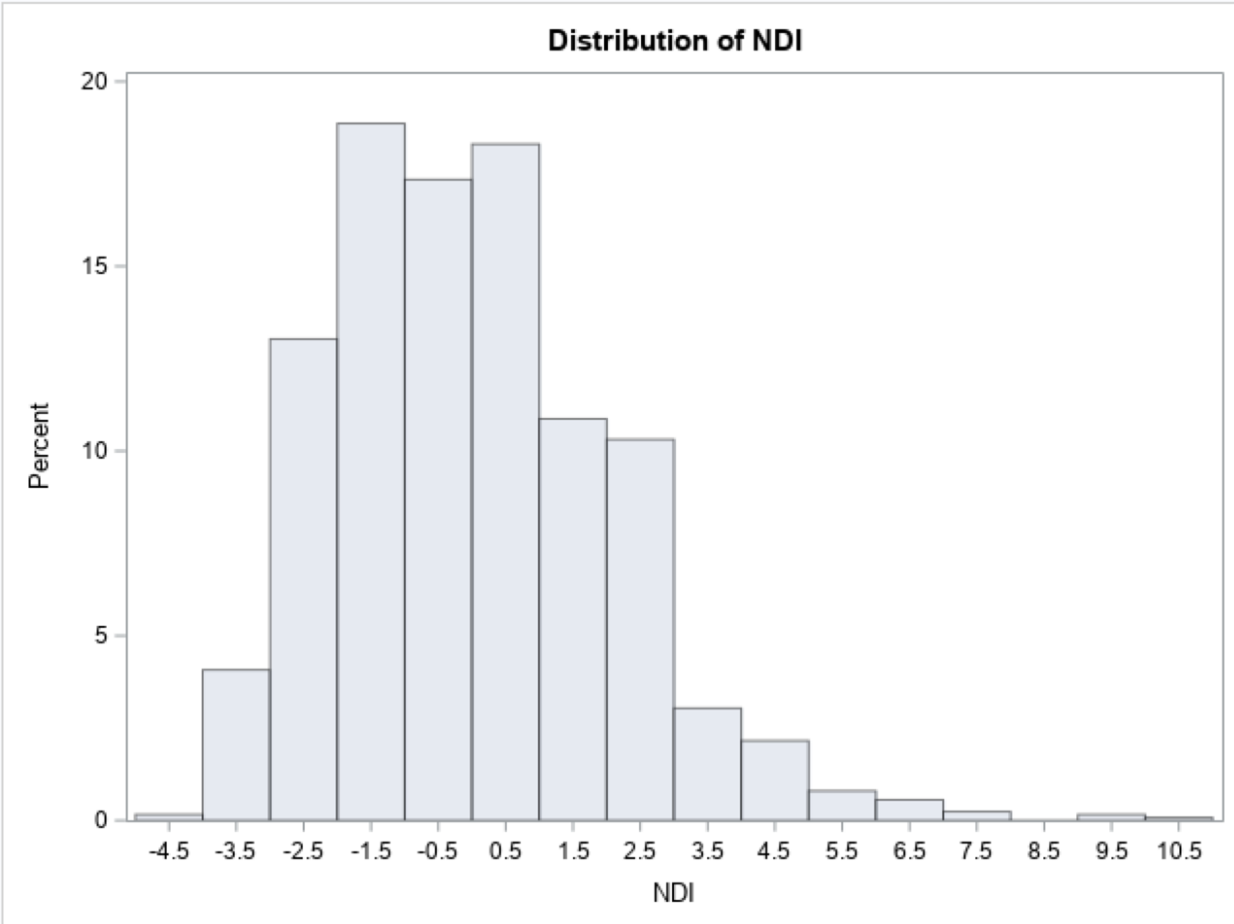
[3]Natural correlation between census variables that is avoided by using PCA.

Figure 2: Variance explained by PCA[4]



[4]Eigenvalue 1 explains and accounts for the most variance in the principal component analysis.

Figure 3: Distribution of Neighborhood Deprivation Index for 20-County MSA of Atlanta[5]



[5]Distribution of Neighborhood Deprivation Index over the census tracts of the 20-County MSA area. Negative NDI indicates less depravity, while positive NDI indicates greater depravity.

CHAPTER III: PUBLIC HEALTH IMPLICATIONS

This study utilized an analysis that addressed the correlation that is present among census-level variables used in socioeconomic research that have commonly been linked to adverse health outcomes by establishing a reproducible deprivation index at the neighborhood level. Because of this, deprivation can be defined in more of a concrete way, and a stronger relationship can be investigated regarding disparity and a health outcome of interest. Additionally, modifying relationships can be easily introduced into the analysis to further understand the impact of co-determining factors if the relationship between deprivation and an adverse health outcome exists.

For researchers looking to investigate the effects of deprivation who have access to patient information on a census tract level, this is a method that can be used to assist them in their investigation. As stated in the discussion, this methodology is reproducible at any level where the patient data is available.

Regarding the proposed relationship between deprivation and invasive Group B *Streptococcus* incidence, this may be a foundation to investigate further into the relationship that deprivation has with race and other risk factors that severely predispose non-pregnant adults to these infections. The base of social determinants of health-related research is to eventually understand what social or economic injustice or barriers may have led to the instance of said health disparity.