ABSTRACT

AN EPIDEMIOLOGICAL STUDY OF INVASIVE PNEUMOCOCCAL DISEASE IN A SOUTH TEXAS-MEXICO BORDER METROPOLITAN COMMUNITY

Steven Hinojosa, PhD
College of Health and Human Sciences
Northern Illinois University, 2018
Jinsook Kim, Director

This research study examines the epidemiology of invasive pneumococcal disease in Hidalgo County, Texas. Invasive pneumococcal disease is a bacterial infection causing severe types of clinical manifestations, including bacteremia and meningitis. Hidalgo County is the largest metropolitan statistical area along the Texas-Mexico border. Supporting a 2016 census population estimate of nearly 850,000, this community facilitates frequent bi-national travel, providing unique attributes towards disease transmission. This study aims to identify if population characteristics for infection aligned with current Advisory Committee for Immunization Practice recommendations for vaccination against pneumococcal disease. This includes identifying if the general healthy adult population has an increased risk for infection, now that other populations are eligible for vaccination. Furthermore, this study aims to identify other population characteristics at increased risk for infection, so that prevention and intervention methods can be targeted to such groups.

Keywords: pneumococcal disease, south Texas, epidemiology, pneumococcal vaccine, Hidalgo County, McAllen-Edinburg-Mission MSA, border health

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

DECEMBER 2018

AN EPIDEMIOLOGICAL STUDY OF INVASIVE PNEUMOCOCCAL DISEASE IN A SOUTH TEXAS-MEXICO BORDER METROPOLITAN COMMUNITY

BY

STEVEN HINOJOSA

A DISSERTATION PROPOSAL SUBMITTED TO THE COMMITTEE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

COLLEGE OF HEALTH AND HUMAN SCIENCES

Doctoral Director:
Jinsook Kim

ACKNOWLEDGEMENTS

I would like to acknowledge and thank my dissertation chair Dr. Jinsook Kim for all the information, guidance, and technical support that she has provided through these years of the program. I would also like to thank my dissertation committee members, Dr. Jeanne Isabel and Dr. James Wilson, for their educational support in this dissertation. I would like to acknowledge Northern Illinois University and the College of Health and Human Sciences for the education commitment that has been provided to me over the years.

In addition, I would like to acknowledge and thank Mr. Eddie Olivarez, chief administrative officer, and the Hidalgo County Health and Human Services Department for their leadership in public health and for the data information exchange support for this research study.

Finally, I would like to thank my family for all of their support over these years in my education. This include my sisters, Vanessa and Kristy, and my mom, Elsa Hinojosa. All three have been instrumental in me pushing onwards. I would also like to thank my best friend, Oberlin Cortez, for all of his support; he witnessed all that was involved and helped me push to excel in my goals. Finally, I would like to thank my dad, Daniel Hinojosa, for always having high expectations in my education and for always guiding me through the years in life.

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF APPENDICES

Page

CHAPTER ONE: INTRODUCTION

Pneumococcal disease is a bacterial infection causing varying types of infections, including otitis media, pneumonia, sepsis, and meningitis (Centers for Disease Contol and Prevention [CDC], 2015a, p. 279). Pneumococcal pneumonia is the most common form of pneumonia among adults, with an incubation period of one to three days (Wattal, Goel, & Byotra, 2017). Pneumococcal meningitis is also the most common form of bacterial meningitis, causing more than 50% of all types of bacterial meningitis (CDC, 2015a, p. 281). This infection is a vaccine-preventable disease. As a bacterial infection, antibiotic therapy can be used as a part of the treatment regimen (CDC, 2015b, p. 6). Although the introduction of vaccination has helped decrease disease incidence, a disease burden may still exist in certain populations.

Definitions of Terms

Invasive Pneumococcal Disease (IPD): an infection of *Streptococcus pneumoniae* from a normally sterile site, such as blood or cerebral spinal fluid (CSF; Gosciminski, Bandy, & Luther, 2017, p. 57)

PCV13: pneumococcal conjugate vaccine, 13 valent

PPSV23: pneumococcal polysaccharide vaccine, 23 valent

Bacteremia: the presence of bacteria within the blood

Meningitis: an inflammation of the meninges (i.e., brain and spinal cord membranes)

Otitis Media: middle ear infection

Trypanophobia: the fear of needles

Socioeconomic Status: a combined measure, typically incorporating economic, social, and work status (Adler & Stewart, 2010)

CDC: Centers for Disease Control and Prevention

ACIP: Advisory Committee on Immunization Practices

MSA: metropolitan statistical area

Case: an episode of infection

Significance of the Problem

Prior to the introduction of the PCV7 vaccine in 2000, pneumococcal disease was a leading cause of disease morbidity and mortality in the United States. Within the United States, approximately 400,000 hospitalizations occur each year due to pneumococcal pneumonia (CDC, 2015a, p. 281). This also includes an estimated 3,000 cases of meningitis and 50,000 cases of bacteremia each year (Gosciminski et al., 2017, p. 57). Additionally, it has been documented that certain populations, such as pediatric and geriatric age groups, have presented concerns for increased risk of infection (Nhantumbo et al., 2016).

In children, *Streptococcus pneumoniae* is the leading cause for otitis media, pneumonia, sepsis, and meningitis. Each year, an estimated 1.6 million children under the age of five die due to pneumococcal-related illnesses worldwide (Nhantumbo et al., 2016, p. 2). With other bacterial infections decreasing in incidence in the pediatric population, pneumococcal disease is now the most common type of infection in children under five in the United States. Additionally, an estimated five million cases of otitis media occur each year in children under five years old, with at least 60% of the population having at least one episode of this infection by the age of 12 months (Centers for Disease Control and Prevention, 2015a, p. 281).

Infection for adults can be just as serious. Each year, approximately 600,000 to 800,000 adults die due to pneumococcal pneumonia, sepsis, and meningitis worldwide (Muley, Ghadage, Yadav, & Bhore, 2017, p. 31). More specifically, 20% of adult cases of pneumococcal meningitis result in death, increasing to 50% in developing countries (Muley et al., 2017, p. 31). During the year 2004, pneumococcal disease was accountable for over 22,000 deaths in the United States, and over 4 million episodes of disease (Song et al., 2013, p. 1). Furthermore, when examining invasive pneumococcal disease (IPD), the Centers for Disease Control and Prevention (CDC) estimates 2,400 cases of mortality due to bacteremia and an additional 1,200 cases due to meningitis to occur annually within the United States (CDC, 2015a, pp. 281-282).

Antibiotic therapy regimens can be used to treat patients with pneumococcal disease, especially those with invasive infections. However, research has shown that some antibiotics are no longer effective due to resistance from the pathogen (Muley et al., 2017; Nhantumbo et al., 2016; Stacevičienė et al., 2016). Therefore, vaccination becomes more important as prevention treatment options may become more limited in the future.

Finally, pneumococcal disease not only presents a health impact but also an economic one as well. Invasive pneumococcal disease (IPD) presents an economic burden estimated at a cost of $3.5 billion in 2004 for direct medical costs in the United States (Song et al., 2013, p. 1). This creates an economic burden for the patient as well. Within a 970-patient, multicenter hospital-based study, the median cost for direct medical care was $7,452 per patient (Song et al., 2013, p. 3).

Research Topic

This research study will explored the epidemiology of invasive pneumococcal disease (IPD) in a south Texas metropolitan area using an ecological study design. It incorporated

statistical and geographical analyses to identify which populations are at most risk for infection. Variables include age, infection type, hospitalization, vaccination status, and other clinical and socioeconomic factors. Specifically, this study aims to identify if current ACIP pneumococcal vaccine recommendations are adequately protecting the general adult healthy population and whether further research is needed to amend current recommendations to protect a broader population.

The study's community is defined as Hidalgo County, Texas. This is also the McAllen-Edinburg-Mission metropolitan area, which represents all jurisdictions and populations located within the county lines of Hidalgo County, Texas (U.S. Census Bureau, 2017).

Hidalgo County Population

Hidalgo County is a Texas county located along the United States-Mexico border in the southern tip of Texas. The county as a whole also represents the McAllen-Edinburg-Mission metropolitan statistical area (MSA). The United States Census Bureau estimates the 2016 population size of this community as 849,843 (U.S. Census Bureau, 2017). The McAllen-Edinburg-Mission MSA is the largest MSA along the US-Mexico border and fifth largest in Texas. Hidalgo County is made up of 22 cities, 27 zip codes (26 complete and 1 partial), and 113 census tracts (U.S. Census Bureau, 2016). These 22 cities within the county, along with their populations are listed in Table 1.

The population of Hidalgo County is relatively young, with a median age of 28.8 years. The county's residents are predominant Hispanic or Latino (91.8%), followed by White alone/not Hispanic (6.6%), Asian alone (1.1%), and African American (0.8%). Between 2011 and 2015, the percentage of foreign-born persons was estimated at 28.8%. In terms of socioeconomic representation from 2012 to 2016, the median household income was $36,094. The per capita

income for the last 12 months of the period between 2011 and 2015 was $14,689. Over a third (34.2%) lived under the poverty line. Additionally, nearly a third (32.0%) of those under the age of 65 years did not have health insurance (U.S. Census Bureau, 2016). Nearly a quarter of the population (22.8%) had less than a 9th-grade education. The overall educational proportion of those graduating high school or higher was 63.3%, but increased to 77.1% in the 25-to-34 year-old population (U.S. Census Bureau, 2016). The median household size for owner-owned properties (non-rentals) was 3.70 in Hidalgo County compared to 2.70 for the United States (U.S. Census Bureau, 2016). Finally, the 2010 population per square mile was 493.2, and the total geographical land area of the county is 1570.87 square miles (U.S. Census Bureau, 2017).

Table 1: Hidalgo County Cities and their Populations

City	2016 Population
Alamo	19,679
Alton	17,278
Donna	16,638
Edcouch	3,365
Edinburg	90,280
Elsa	7,134
Granjeno	301
Hidalgo	13,931
La Joya	4,286
La Villa	2,740
McAllen	142,696
Mercedes	16,734
Mission	84,424
Palmhurst	2,738
Palmview	5,792
Penitas	4,886
Pharr	79,487
Progreso	5,938
Progreso Lakes	248
San Juan	36,981
Sullivan City	4,153
Weslaco	40,358

Purpose of Study

The primary objective of this research is to examine the disease burden and population characteristics for infection of invasive pneumococcal disease. This includes examining which age groups are most susceptible to infection and which populations are at increased risk for severe complications. Secondary objectives include examining and comparing subpopulations to determine if socioeconomic factors contribute to disease morbidity and other outcomes. Geographical techniques were employed to determine potential correlation of disease alongside income and population density. Additionally, current literature is limited in regards to the epidemiology of pneumococcal disease within United States-Mexico border communities. Since the literature has shown different characteristics of disease between the two countries in regards to serotype, infection type, morbidity and mortality (Sherchan et al., 2015; Song et al., 2013; van der Linden, Falkenhorst, Perniciaro, Fitzner, & Imöhl, 2016), this study will provide insight into disease outcomes where a population is influenced by two countries. Furthermore, this research will contribute information and outcomes that have not previously been established and assist in identifying if disease incidence and infection characteristics differ from national representation.

Study Justification

Research has shown that the introduction of the pneumococcal vaccine has decreased the incidence of the disease (Askim et al., 2016; Gosciminski et al., 2017). However, with current vaccination recommendations, those who do not qualify under these recommendations are still at risk for infection, resulting in disease morbidity and even the possibility of death. Therefore, the question arises whether there is adequate herd immunity protecting those ineligible for vaccination or if this is a susceptible population currently unprotected. Additionally, by

conducting research on those already infected, targeted populations can be identified for future public health interventions.

Pneumococcal vaccination has decreased incidence in pediatric and geriatric populations (Gosciminski et al., 2017). This study's aim focuses on the general adult healthy population, establishing risk for this disease within this population segment. It is currently unclear if this population has significant risk for disease. It is also unclear if this issue stems from lack of vaccination or a lack of eligibility for vaccination. Therefore, this study examined if those cases infected in the general adult population should have been vaccinated or if they did not qualify, as per Advisory Committee Immunization Practices (ACIP) recommendations.

Within the epidemiological literature, there has been little focus on border populations. Few studies have examined invasive pneumococcal disease in Latin and southern America, with even fewer studies in focusing on Mexico. With bi-national transit allowing various infectious diseases to transmit across borders, it is vital that increased research takes place in these border regions so that clearer transmission risks are identified. Furthermore, infection types can vary globally, and with a bi-national population it is vital to understand whether national vaccination recommendations are protecting such communities. Additionally, this population represents a much lower income representation as compared to communities in other studies. Therefore, studying this population provides contextual opportunities when exploring socioeconomic status and its role in disease outcomes for pneumococcal disease.

Currently, the literature presents a strong representation of pediatric populations globally, especially those groups under the age of five years (Hsiao et al., 2015; Sherchan et al., 2015). Additionally, the current literature has reviewed geriatric populations and those with comorbidities currently covered under ACIP recommendations (Marcus et al., 2016; Wattal et al.,

2017). However, studies examining the adult population are less abundant. This is especially true when studying the healthy adult population, where no comorbidities or conditions constitute vaccine recommendations. Therefore, this study will assist in contributing to a research focus of the adult population where little to no research currently exists.

Theoretical Frameworks and Models Overview

Theoretical frameworks and models for this study encompass concepts of disease transmission and population health. The epidemiologic triad represents a model of disease causation (Centers for Disease Control and Prevention [CDC], 2012). The triangle of human ecology depicts the health status of the individual and the influence of external factors (Meade & Emch, 2010, p. 30). Herd immunity defines the threshold for population immunity (Hendrix, Sturm, Zimet, & Meslin, 2016). The Health Disparities framework examines health inequity among social groups (Adler & Stewart, 2010). The socioecological model of health depicts the influences that can be placed on a person's health (Ferrer, Trotter, Hickman, & Audrey, 2014, pp. 3-4). Finally, the willingness-to-pay concept examines the patient and healthcare cost relationships (Prosser et al., 2004, pp. 283-284). The frameworks and models presented are illustrated in the appendices.

Research Questions

This research brings forth questions on infection of invasive pneumococcal disease and its relationship to a susceptible population.

- With vaccination recommendations established for pediatric, geriatric, and high-risk populations, is the healthy general adult population still at high (or increased) risk for infection?

- Within this study population, do high-risk patients experience a longer hospitalization duration and higher risk of mortality?
- Does socioeconomic status, such as household income, influence the rate of disease within the study population?

Limitations

This epidemiological study of invasive pneumococcal disease had some limitations. The study was a longitudinal and ecological study, spanning from 2010 to 2016. Although this study design does provide extended annual data, it creates cross points in patient data. This means that once a patient is discharged from a hospital, their ultimate recovery and outcome are unknown. It is also unknown if they returned to a hospital due to complications, unless the patient tests positive for pneumococcal disease again. Furthermore, with additional and cheaper healthcare access options just south of the border, those seeking medical care in Mexico will likely not be reported to the Hidalgo County Health and Human Services Department. It is Texas state law to report this disease, but this law is not applicable in Mexico. Therefore, cases may be missed for those identified in Mexico.

Since this is not a cohort study, little to no information will be known after hospital discharge. Information was limited to only during the initial medical visit or hospitalization. This type of case investigation does not routinely involve patient contact or follow-up. Therefore, long-term patient outcomes, such as post-discharge survival and person-time, were collected.

Some variables of interest were not complete or available, such as vaccination history. Texas has a state-wide immunization registry. However, this is an opt-in program, meaning that patients must choose to signup, rather than automatically being enrolled. Adult patients are less

likely to be in this registry as compared to pediatric patients. Additionally, patients who are poor historians of their vaccination history or lack a medical home may have missing information for vaccination status. Patients may also mistake the pneumococcal vaccine for other vaccines, such as the influenza or TDaP (tetanus, diphtheria, and acellular pertussis) vaccines. Furthermore, since it is suspected that most cases of disease will present to a hospital at time of diagnosis, many of these facilities may not have the vaccine history on record, as compared to primary care providers.

This research study focused on those who had a positive infection for pneumococcal disease. Therefore, as compared to a case-control study, there is no data reflecting individuals protected from disease. Examples include populations who were already vaccinated or not vaccinated but successfully were not infected at the time of study. These populations could have provided further information on population characteristics, positive and negative, for infection versus protection.

Summary of the Study

A vaccine has been developed for pneumococcal disease, creating an opportunity to prevent disease. This includes preventing more severe clinical manifestations of disease, such as sepsis and meningitis. However, even with the introduction of vaccine options, pneumococcal disease is still a risk for certain populations, potentially including those who do not qualify under national recommendations. Therefore, the theoretical frameworks described in this chapter identify the foundational platforms that depict and exemplify the significance of this problem. With the multifaceted presence of this health issue, this ecological study addressed the relationship of disease, impact of at-risk populations, and identified the interaction of explanatory and response variables. Additionally, with disease characteristics varying globally,

limited research has been conducted among communities in Latin American, Mexico, and along the United States-Mexico border. Furthermore, this epidemiological study assessed a south Texas border community and its characteristics for risk of infection from invasive pneumococcal disease. This study will identified those populations most susceptible for infection and those age or risk groups that may be susceptible to more severe disease outcomes. This study also examined whether it is indeed an issue of risk of infection among the general adult population or if those considered high risk are not receiving protection from disease due to undervaccination of eligible patients. In addition, socioeconomic factors, such as income were assessed to determine if they have any influence on disease risk. Geographical analysis assisted in identifying population characteristics associated with higher rates of infection. Furthermore, this study assessed the disease outcomes to determine areas of concern so that public health intervention and prevention methods can be identified.

CHAPTER TWO: LITERATURE REVIEW

Pneumococcal disease's cause is the pathogen *Streptococcus pneumoniae.* This pathogen is a lancet-shaped bacterium that normally forms as diplococci or less commonly can be single or in chains (Centers for Disease Control and Prevention, 2015a, p. 279). Pneumococcal disease can take place in different forms of infections, this including otitis media, pneumonia, sepsis, and meningitis (CDC, 2015a, p. 280).

Streptococcus pneumoniae was first discovered by Louis Pasteur in 1881 from the saliva of a patient with rabies. By 1940, more than 40 serotypes of the bacteria had been identified (CDC, 2015a, p. 279). The prevalence of the organism *Streptococcus pneumoniae* is estimated to be present in about 10% of the general population. This can then be transmitted to others, leading to infection. If the bacterium passes beyond the epithelial cells, it can then lead to invasive disease (Askim et al., 2016, p. 2). Currently, a total of over 92 serotypes have been identified, which are classified based on their type-specific antisera (CDC, 2015a, p. 279).

Pneumococcal pneumonia, a clinical syndrome of pneumococcal disease, has an estimated incubation period of one to three days. Symptoms can include fever, cough, chest pain, tachypnea, hypoxia, malaise, and weakness (CDC, 2015a). Pneumococcal pneumonia represents over a third (36%) of adult-acquired pneumonia. About 25-30% of those with pneumococcal pneumonia develop pneumococcal bacteremia (CDC, 2015a, p. 280). Pneumococcal bacteremia is defined as a bacterial presence within the blood. This type of infection can be without focus and cause general febrile illness (Azzari et al., 2015a, p. 2). Each year in the United States over 12,000 cases of pneumococcal bacteremia occur. Additionally, the case fatality rate is estimated

at 20% in the general population and as high as 60% within the geriatric population (CDC, 2015a, p. 278). Another type of infection from pneumococcal disease is pneumococcal meningitis. With 3,000 to 6,000 cases annually, this type of infection is the most common type of bacterial meningitis in the United States, accounting for over 50% of infections. Symptoms can include headache, lethargy, vomiting, irritability, fever, nuchal rigidity, cranial nerve signs, seizures and coma (CDC, 2015a, pp. 279-282).

Literature of Theoretical Frameworks and Models

The following are existing theoretical frameworks that describe components that are linked to this study. These include concepts of disease transmission and population health. These components also support a holistic model, the socioecologic pyramid of human disease, which describes the relationships and influences of this study. First, the epidemiologic triad represents a model of disease causation (Centers for Disease Control and Prevention [CDC], 2012). This model depicts a triangle, with each vertex representing a factor in disease transmission. This includes the host, the agent, and the environment. The host represents the patient or individual susceptible to infection. The agent is the pathogen causing disease, such as the *Streptococcus pneumoniae* bacterium leading to pneumococcal disease. The environment depicts extrinsic factors that influence disease transmission. This can include physical factors, such as geology and climate, or socioeconomic factors, like crowding and poor sanitation (Centers for Disease Control and Prevention, 2012). The epidemiological triad presents the overall relationship of communicable diseases. This framework depicts the foundation of how *Streptococcus pneumoniae* spreads person to person, potentially manifesting as pneumococcal disease. Furthermore, it provides a foundation for epidemiologic research for identifying risk factors important in public health prevention and intervention.

A similar framework is the triangle of human ecology. Central to this model is the health status of the individual or population under study and how it is influenced by the categories of habitat, population characteristics, and behavior (Meade & Emch, 2010, p. 30). These three categories form the vertices of the triangle. The first vertex, habitat, encompasses the home, neighborhoods, schools, healthcare facilities, workplace, or built environment, such as roads. Next, the population vertex includes age, genetics, immunity, and nutritional status. The third vertex is behavior. Behavior is defined as the observable acts in a cultural context. This includes economic constrains, social norms, education, political, technological, and other influences (Meade & Emch, 2010, pp. 30-31).

Another theoretical model is herd immunity. Herd immunity refers to a theory that once a threshold is met within the population, others who are not vaccinated may potentially have a reduced risk and ultimately protection against infection (Hendrix et al., 2016). These thresholds represent community protection of vaccine-preventable diseases, where a certain percentage of the population vaccinated can prevent future clusters and outbreaks. This theory can be formulated in vaccine-preventable diseases, including invasive pneumococcal disease (IPD). Along with herd immunity, the basic reproduction number (R_0) plays a role as well. The basic reproduction number refers to the average number of susceptible people who become infected (Hendrix et al., 2016). This number predicts the scale of a disease during an outbreak, usually set to be either above or below one (1). With increased vaccination, the susceptible population decreases and so does the R_0 (Hendrix et al., 2016). This model presents the importance of research in both the epidemiology and vaccination of pneumococcal disease. Specifically, within this study, vaccine and epidemiological research is supported using this model. This includes

examining disease incidence and determining whether ACIP recommendations for the pneumococcal vaccines are supporting herd immunity protection within the adult population.

The health disparities framework is defined as health inequity among social groups (Adler & Stewart, 2010). This framework examines how the health of certain populations can be poorer when compared to others due to their race/ethnicity, income, education, and other characteristics. When examining this border community, health inequity may be of concern due to overall social standing and access to care when compared to other communities. Hidalgo County has lower socioeconomic status as measured by median household income, health insurance coverage, and education completion rates when compared to Texas and the United States (U.S. Census Bureau, 2017).

The socioecological model of health, as presented in Appendix A, depicts the influences that can be placed on a person's health. This model shows the various layers of influences towards an individual's health, including individual, interpersonal, organizational, community, and public policy (Ferrer et al., 2014, pp. 3-4). The individual level represents how individual patients influence their own health. At the interpersonal level, health is impacted by the patient's family members, friends, and healthcare providers. The organizational level begins incorporating health insurance providers and health departments, offering access to care. The community level encompasses media, community-based organizations, and employers. Both the organizational and community levels represent an important influence for public health education on disease prevention. The public policy level represents the policies and laws, such as with federal vaccine recommendations (Ferrer et al., 2014, p. 3). As a whole, this model categorizes the important influences for determining data variables to be collected in this research study.

Economics and financial factors are also important to be examined when exploring frameworks for this research study. The concept of "willingness to pay" is defined as the theoretical amount a patient will pay for an intervention to prevent a negative health outcome (Prosser et al., 2004, pp. 283-284). Within this willingness-to-pay study for the pneumococcal vaccine, economic modeling was developed to identify such trends as vaccine costs while incorporating medical and non-medical event outcomes (Prosser et al., 2004). Examples include vaccination cost to prevent episodes of specific infection (i.e., otitis media or pneumonia) or even a timed series (i.e., three days of illness or quality-adjusted life years). In regards to this focus of research, the pneumococcal vaccine is an available prevention tool. However, with the lower median household income of this population, it is of interest whether infected patients qualified for vaccination and if cost could be a potential barrier.

Epidemiology of Invasive Pneumococcal Disease

Morbidity and Mortality

Prior to the year 2000, invasive pneumococcal disease was a leading cause of disease morbidity and mortality (Gosciminski et al., 2017, p. 57). However, even in the present day, IPD presents high disease morbidity and mortality globally (Hsiao et al., 2015, p. 1). In many countries, although a vaccine-preventable disease, pneumococcal disease is still a leading cause of childhood deaths (Sherchan et al., 2015, p. 47). Furthermore, in developed countries, overall mortality rate is estimated at 20% but increases to as high 50% in developing countries (Muley et al., 2017, p. 31).

A regional study in Sweden (Backhaus et al., 2016) examined the epidemiological characteristics of invasive pneumococcal infections. A total 2,977 cases were included from 1996 to 2008. Descriptive statistics were performed to identify population characteristics, including

age, gender, and comorbidity presence. The mean age was 60.6, and the median age was 65 years. Men made up 1,483 cases, and women made up 1,494 cases. Of all cases, 1,994 (67%) had at least one comorbidity, and only 67 (2.25%) cases had a documented history of vaccination, mostly from the PPSV23 vaccine (Backhaus et al., 2016, pp. 3-6).

In another study (Askim et al., 2016) the Charlson Comorbidity Index (CCI) was used to examine risk of infection. The CCI combines both age and comorbidities to represent a numerical risk value. Within this study, a total of 414 patients with pneumococcal sepsis were reviewed. Of these, 144 (34.8%) patients had a CCI score of zero, 190 (45.9%) has a score of 1-2, and 80 (19.3%) had a score 3 or greater (Askim et al., 2016, p. 4). Comorbidity infections were also examined individually, with 98 hypertension patients (23.7%) and 96 chronic obstructive pulmonary disease (COPD) patients (23.2%) as the most common conditions (Askim et al., 2016, pp. 4-5).

High-Risk Populations

The Centers for Disease Control and Prevention (CDC) and the Advisory Committee on Immunization Practices (ACIP) have identified certain population characteristics as high risk for infection of pneumococcal disease. For age groups two years and older, this includes chronic illness, anatomic or functional asplenia, immunocompromised conditions, HIV infection, cochlear implants, or cerebral spinal fluid (CSF) leak. Additionally, those 19 years and older who smoke cigarettes or have asthma are also considered high risk for infection (CDC, 2015a, pp. 289-292).

A healthcare system study in the United States (Marcus et al., 2016) compared risk of infection between HIV-infected (n=13,079) and uninfected (n=137,643) individuals. Incidence of disease was compared among these two groups from 305 cases per 100,000 person-years in 1996

to 88 cases per 100,000 person-years in 2011. HIV-infected individuals had a significantly higher risk for infection ($p<0.001$) during all time periods, with decreasing ratios as years progressed. In 1996-1999, HIV-infected individuals had an adjusted risk ratio of 18.5, as compared to uninfected individuals. More recently, in 2010 to 2011, HIV-infected individuals had an adjusted risk ratio of 6.6, as compared to uninfected individuals, presenting findings that pneumococcal vaccination among the HIV-infected population helped decrease the incidence rates of IPD (Marcus et al., 2016, pp. 465-467).

Diabetes is another risk factor for pneumococcal disease. In a United States medical claims study (Weycker et al., 2016) diabetes mellitus was the most common chronic condition among at-risk patients (18-64 years), representing 43% of the 13,368,935 claims. Additionally, patients with diabetes had an overall incidence rate ratio (IRR) of 4.3 (Weycker et al., 2016, p. 4). Another study (Fay, Hoppe, Schulkin, & Eckert, 2016) surveyed obstetrics and gynecology residents to assess their knowledge on pneumococcal vaccination indications for pregnant patients. A total of 238 respondents were included within the survey. In regards to pneumococcal vaccine indications, only 35.8% of respondents successfully identified diabetes as an indication for vaccination (Fay et al., 2016, p. 3).

Population Characteristics

Pneumococcal disease epidemiology presents population characteristics, such as age and other risk factors. A Norwegian observational study (Askim et al., 2016) examined the epidemiological outcomes of adults with septic pneumococcal disease. Within the 1993-2011 dataset, the mean age was 67 years old, with 48.3% aged 69 or younger (Askim et al., 2016, pp. 3-4). A study in Sweden (Backhaus et al., 2016) examined the epidemiology of invasive pneumococcal infections. Between 1996 and 2008, a total of 1,083 cases were identified, with

485 cases aged 18-64 years and 530 cases aged 65 or greater (Backhaus et al., 2016, p. 7). In a South Korean retrospective cohort study, data from ten university hospitals were collected on invasive pneumococcal disease (IPD). A total of 970 patients were identified with IPD, and of these, the 50-64 age group was most common (33.0%), with a median age of 60.9 years (Song et al., 2013, pp. 6-7).

Border Health and Latin American Populations

Current literature presents limited epidemiological information on those populations in Latin America, Mexico, or along the United States-Mexico border. These regions may include populations that have different characteristics when compared to other populations across the nation. Furthermore, with these transient Latin American populations, disease risks and attributes may meld with United States-Mexico border communities.

A study in Mexico City (Gómez-Barreto, Espinosa-Monteros, López-Enríquez, Jiménez-Rojas, & Rodríguez-Suárez, 2010) assessed the epidemiology and mortality risks factors of invasive pneumococcal disease in a pediatric hospital. A total of 156 patients were included in this study, with hospitalizations from 1997 to 2004. Of these patients, 69.3% had at least one underlying condition (Gómez-Barreto et al., 2010, p. 393). The overall mortality rate of these patients was 27.5%, and of those cases, 81.4% had at least one underlying condition. Bacteremia and meningitis have an odds ratio of 2.3 for mortality risk, as compared to other types of infections (Gómez-Barreto et al., 2010, p. 395).

A study in Guatemala (Contreras et al., 2015) examined the 2008-2012 incidence of pneumococcal pneumonia in 188 hospitalized adults. In this study, that most common age group affected was those aged 65 and older (38%), with an incidence rate of 31.3 per 100,000 (Contreras et al., 2015, p. 10). The most common symptoms presented included cough (92%),

difficulty breathing (84%), and reported fever (72%). Of all patients, 65% had consolidation or a large effusion identified on chest x-rays (Contreras et al., 2015, p. 4). With data representing adult cases 2008-2012, the study presented a baseline prior to the 2013 vaccine introduction in children (Contreras et al., 2015).

An August 2014 Morbidity and Mortality Weekly Report (MMWR; Nyangoma et al., 2014) assessed respiratory infections of unaccompanied minors traveling to the United States from Central America. In the 2014 report, clusters of pneumococcal disease were identified in both California and Texas. A total of six cases, aged 14-17 years old, were identified. All six cases were identified as serotype 5 (Nyangoma et al., 2014, p. 698).

The research articles presented above describe the limited literature currently available on Latin American and United States-Mexico border communities. Therefore, this presents the need for additional research examining these communities to identify if similar disease characteristics occur or if instead novel public health interventions will need to be developed to better combat invasive pneumococcal disease in these areas.

Spatial Epidemiology

Limited research is currently available on the spatial epidemiology of pneumococcal disease. However, the available literature does present research on spatial analysis methods in regards to other diseases. In one spatial epidemiology study (Stopka et al., 2017) hotspots, an area of elevated disease burden, of hepatitis C in Massachusetts was characterized. In that study, spatial analysis techniques included counts and incidence rates per census tract, kernel density, and hotspot maps (Stopka et al., 2017, p. 5). The study also presented methods for data cleaning, such as correcting mismatched zip codes, and geocoding to obtain longitude and latitude coordinates (Stopka et al., 2017, p. 3). Another study (Jeffery, Ozonoff, & Pagano, 2014) focused

on spatial aggregation to map risk of disease. In that study, leukemia data in New York were analyzed using a non-parametric distance-based mapping (DBM) approach (Jeffery et al., 2014, p. 2). The method used disease point data, placed onto grids, to create high-risk areas.

A study in southern Ethiopia (Tadesse, Enqueselassie, & Hagos, 2018) examined the spatial epidemiology of tuberculosis. In that study, global spatial autocorrelation was conducted to examine clustering of cases. Moran's index was compared on an annual basis to identify patterns, resulting in significant clustering each year (Tadesse et al., 2018, pp. 6-10). A study on avian influenza in Zhejiang Province, China (Wu et al., 2017) also incorporated spatial autocorrelation to examine five waves of Avian Influenza epidemics. In that technique, Moran's index was also used for local analysis, where high-low and low-high autocorrelations were examined (Wu et al., 2017, p. 3).

In a Netherlands dietary study (Dekker, Rijnks, Strijker, & Navis, 2017) the Getis Gi hot and cold spot analysis tool was utilized to study snack patterns. In consideration of pneumococcal disease, these spatial analysis techniques support methods for identifying additional risk factors and disease trends. Examples of this include geocoding point data into polygon symbology to compare disease incidence rates and through conduction spatial statistical analysis to identify if any significant risk factors are present regarding toward infection.

Vaccination

Research has shown that pneumococcal disease can be prevented through vaccination (Chih-Cheng, Sheng-Hsiang, Chun-Hsing, Wang-Huei, & Po-Ren, 2014; Gosciminski et al., 2017). There are two main types of vaccines overall to protect against diseases: live-attenuated and inactivated. A live attenuated vaccine refers to when the vaccine contains a disease-producing wild-type pathogen (i.e., bacteria or virus). This pathogen can then replicate within the

body to produce immunity without causing illness. Some examples of live-attenuated vaccines include the mumps-measles-rubella-varicella (MMRV) vaccine and the nasal influenza vaccine (CDC, 2015b, pp. 4-6). Inactivated vaccines refer to those that contain pathogens that have been killed. This is accomplished through techniques using heat and chemicals. These vaccines do not cause disease from an active infection, creating the need for a multidose series as immunity wanes over time (CDC, 2015b, pp. 5-6).

Pneumococcal Vaccination

In the National Health and Wellness Survey (NHWS), from 2007 to 2011, adult vaccination was assessed. For pneumococcal disease, a total of 43.8 million adults had received at least one vaccine, representing 19.2% of adults (Annunziata, Rak, Del Buono, DiBonaventura, & Krishnarajah, 2012, p. 5). Pneumococcal vaccination rates were also assessed in another United States surveillance study (Williams et al., 2016). In the 2014 study, pneumococcal vaccination coverage in high-risk adults, aged 19-64 years, was 20.3%. In adults 65 years and older, vaccination rates were 61.3%. In the 19-64 age group, Caucasians had the highest rate of vaccination at 21.1%, second was African Americans (20.2%), third was Hispanics (16.4%), and fourth was Asians (14.6%; Williams et al., 2016, pp. 4-5). In a 2011-2015 pediatric study in the United States (Hill, Elam-Evans, Yankey, Singleton, & Dietz, 2016), pneumococcal vaccination rates among infants 19-35 months old were examined. For a three-dose series coverage, vaccination rates ranged from 92.3% (2012) to 93.6% (2011). For a four-dose series coverage, this rate ranged from 81.9% (2012) to 84.4% (2011; Hill et al., 2016, p. 1066).

The pneumococcal vaccine belongs to the category of polysaccharide vaccines. These are inactivated vaccines, composed of long sugar molecule chains with a surface capsule of the pathogen. The two types of polysaccharide vaccines for pneumococcal disease include pure and

conjugate (CDC, 2015b, pp. 6-7). Once inoculated, antibody development for the protection of disease usually occurs within two to three weeks (CDC, 2015a, p. 286).

The pure version is called the pneumococcal polysaccharide vaccine (PPSV). The first PPSV was licensed in 1977 as a 14-valent vaccine, protecting against 14 serotypes of bacteria. In 1983, the PPSV23, protecting against 23 serotypes, replaced the previous 14-valent version, which can still be used currently. This version of the vaccine contains 25 mcg of each antigen per dose and 0.25% phenol as a preservative (CDC, 2015a, p. 284). Some areas of concern are present with pure polysaccharide vaccines. First, pure polysaccharide vaccines are not consistently immunogenic in children under two years of age. This may be because of immaturity in their immune systems. Second, a booster vaccine does not evoke an increased response for immunity, inhibiting long-term protection against disease (CDC, 2015b, pp. 6-7).

The pneumococcal conjugate vaccine (PCV) is a vaccine that undergoes conjugation. This process chemically combines the polysaccharide to the antigen, resulting in immunogenic response in infants and through antibody boosters (CDC, 2015b, p. 7). The first pneumococcal conjugate vaccine was licensed as a 7-valent vaccine in the year 2000. In 2010, the 13-valent vaccine was introduced in the United States. In 2012, ACIP recommendations for the PCV13 were for immunocompromised adults aged 19 years and older and in 2014 for all adults aged 65 years and older. One 0.5mL does of PCV13 contains 2.2µg of polysaccharide of 12 of the serotypes and 4.4µg of polysaccharide for serotype 6B. The 13 serotypes within this conjugate vaccine include 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, and 23F (CDC, 2015a, pp. 284-285).

Current ACIP vaccine recommendations for the general pediatric population include a four-dose series at 2, 4, 6, and 12-15 months of age of the pneumococcal conjugate vaccine (PCV) (CDC, 2015a, p. 287). In unvaccinated adolescents aged 6 to 18 years, a single PCV dose

is recommended if a high-risk condition is present (i.e., anatomic asplenia, immunocompromising conditions, cochlear implant, or cerebrospinal fluid leak). For patients 65 years and older, a single dose of PCV is recommended, followed by a single dose of PPSV 6-12 months after the initial dose. If the PPSV was the initial dose, an additional single dose of PCV is recommended no sooner than 12 months after the initial PPSV dose. For high-risk adults aged 19-64 years old, an initial dose of PCV is recommended, followed by a dose PPSV eight weeks later (CDC, 2015a, pp. 288-289).

Serotype Studies

Serotype studies can be beneficial in identifying specific infections that are circulating within a population (or community). This is also helpful in targeting certain geographic areas, based on their serotype presence among infections. Globally, serotypes can vary based on the region where infection is occurring. With this information, preventive methods such as vaccination and vaccine research can be implemented. This includes identifying if the most prevalent serotypes are matched to currently available pneumococcal vaccines and identifying if certain serotypes cause more severe infections. In a study within a Taiwanese hospital (Hsiao et al., 2015), the epidemiology of infection by serotype was conducted to determine clinical features and outcomes. Among meningitis patients, the following serotypes were identified: 19F (3 cases), 6B (1 case), 15 (1 case), 19A (1 case), 23F (1 case), and unknown (3 cases). Pneumonia patients had the following serotypes: 19A (19 cases), 14 (5 cases), 6B (2 cases), 6A (1 case), 23F (1 case), non-typeable (1 case), and unknown (8 cases; Hsiao et al., 2015, p. 88). By categorizing infection type by serotype, the study was able to determine if certain serotypes are prone to causing more severe infections.

Specifically, regarding acute bacterial meningitis, a study in Mozambique (Nhantumbo et al., 2016) examined the serotype distribution of children with invasive pneumococcal disease. In the study, a total of 352 samples were collected. Of these samples, 119 were isolated with *S. pneumoniae*. Of these samples, 50 met criteria to be serotyped. Of the 50 samples, 17 were not typeable through the study's detection methods. The following serotypes were identified: 9V (12 cases), 1 (6 cases), 5 (5 cases), 14 (4 cases), 4 (3 cases), 6A (3 cases), 23F (3 cases), and 6B (2 cases). Serotypes 12, 15B, and 3 each had one case (Nhantumbo et al., 2016, pp. 4-5). By focusing on this population within a specific geographical area, the researchers were able to identify the most prevalent serotypes to cause acute bacterial meningitis within the study's parameters. In addition, both of these studies present an example of how serotypes can differ based on geographic region, population, and infection type.

Vaccination Barriers, Hesitancy, and Acceptance

Even with a vaccine currently available, barriers still exist against vaccination. Without adequate protection, patients, especially those considered high risk, are susceptible to infection as well as complications from the disease. Trypanophobia refers to the fear of needles, and this fear can lead to health consequences and implications, such as vaccine refusal (Taddio et al., 2012). This type of fear has been documented to impede vaccination efforts among both children and adults (Taddio et al., 2012, p. 4807). In a Canadian survey (Taddio et al., 2012), over 1,900 parents and children were asked about vaccination at a local museum. Of these respondents, 24% of parents and 63% of children reported at least some form of trypanophobia. Additionally, among those surveyed, 7% of adult and 8% of children immunizations resulted in non-compliance (Taddio et al., 2012, p. 4807). Studies like this can be used to focus on areas of concern and where interventions can be developed. One example is in the age groups presenting

trypanophobia within the Canadian study. The 4-8 age group presented the highest rate of fear, encompassing 56% presenting with fear within that particular age group (Taddio et al., 2012, p. 4809).

Vaccination hesitancy and acceptance have been described in the literature to be attributed to multiple factors (Ferrer et al., 2014; How How, Phua See Chun, Shafi, & Jakes, 2016; Prosser et al., 2004). Within a research article on vaccine hesitancy (MacDonald, 2015), the 3 C's model is presented. This model focuses on the concept of vaccine acceptance and hesitancy. This vaccine hesitancy model encompasses behaviors contributing to the decision-making process of vaccines. Vaccination hesitancy refers to either a delay in acceptance or refusal of vaccination and can be the result of different influences. In correspondence, vaccine acceptance is a behavior that stems from a complex decision-making process that encompasses many factors (MacDonald, 2015, p. 4162). These determinants can be represented through models and matrices to depict the influences on the vaccine decision-making process.

As presented in Appendix B, the 3 C's model represents three core areas: complacency, confidence, and convenience. Complacency refers to not getting vaccinated due to the perceived risk of getting infected. Influential factors of complacency can include life and health responsibilities, where other responsibilities are regarded as more important at that point in time of decision. Another component of complacency is self-efficacy, which refers to the self-perception of ability to actually get vaccinated (MacDonald, 2015, p. 4163). Confidence refers to the trust in a vaccine. More specifically, this is divided into three key segments. First, trust in effectiveness and safety of the vaccine. Second, trust in system delivery (i.e., health services and health professionals). Third, trust in the policymakers who passed regulations on the need of the vaccine. Vaccine convenience refers to the accessibility of the vaccine. This includes physical

availability, affordability and willingness to pay, geographical accessibility, and health literacy of uptake for the vaccine (MacDonald, 2015, pp. 4162-4163). The 3 C's model represents the potential barriers to vaccination, such as with the pneumococcal vaccine. In addition, it supports this work of research within pneumococcal vaccination and disease epidemiology because it presents the factors associated with vaccination and the factors within the decision-making process. As this study explores vaccination history, explanatory variables, such as census tract income, will help identify whether socioeconomic influences play a role in vaccination coverage within the community.

Treatment and Antibiotic Resistance

Pneumococcal disease, like other bacterial infections, can be treatable with antibiotic therapy (CDC, 2015a). This is utilized through the incorporation of an antibiotic treatment regimen in patients. However, antibiotic resistance can greatly limit the treatment options for pneumococcal disease. Studies within the literature have been conducted to determine if any antibiotic resistance is present that could compromise treatment plans and regimens for patients.

Antibiotic resistance is determined through antimicrobial susceptibility panels. These profiles are developed through culturing and isolating the bacteria and then testing the bacteria against a series of antibiotics. When testing bacterial resistance to antibiotics, two main categories are developed: susceptible and non-susceptible. Susceptible refers to bacteria that are sensitive, displaying no resistance to a particular antibiotic. Non-susceptible refers to bacteria either being intermediate or fully resistant to an antibiotic (Stacevičienė et al., 2016). Within the antimicrobial susceptibility panel, each antibiotic is given a numerical value, known as the minimum inhibitory concentration (MIC). MIC refers to the lowest concentration of an

antimicrobial needed to limit bacterial growth (Muley et al., 2017, p. 32). The scores are then categorized into susceptible, intermediate, or resistant.

Geographically, antimicrobial susceptibility profiles can range greatly. The literature has shown that non-susceptibility to Penicillin for *S. pneumoniae* can range from as low as 1.7% in Norway to as high as 83% in Romania. Additionally, non-susceptibility to Erythromycin can range from 1.2% in the Czech Republic to up to 65.5% in Italy (Stacevičienė et al., 2016, p. 2). The discrepancies in antimicrobial profiles depict how disease management may need to be different for each population affected.

An observational study in a tertiary care hospital in India (Muley, Ghadage, Yadav, & Bhore, 2017) examined patients with invasive infections admitted to one facility. A total of 120 patients with suspected invasive infections were included within the study. Of these patients, 55 patients had *S. pneumoniae* isolated. The following antimicrobial susceptibility patterns were identified: 22/55 isolates were resistant to Cotrimoxazole, 7/55 resistant to Chloramphenicol, 4/55 resistant to Ciprofloxacin, 2/55 resistant to Erythromycin, and 0/55 resistant to Cefotaxime. No more than two antibiotics were resistant within a single isolate (Muley et al., 2017, pp. 32-34).

An antimicrobial study in Nepal (Sherchan et al., 2015) examined the susceptibility patterns of *S. pneumoniae* isolates from children with clinical meningitis. A total of 24 isolates were tested for antimicrobial susceptibility. Penicillin, Cefotaxime, Chloramphenicol, and Erythromycin were sensitive to almost all isolates (23 out of 24). More specifically, both Penicillin and Cefotaxime each had 23 samples sensitive, 1 sample intermediate, and 0 samples resistant. Cotrimoxazole had 15 samples sensitive, 4 samples intermediate, and 5 samples resistant. Both Chloramphenicol and Erythromycin each had 23 samples sensitive, 0 samples

intermediate, and 1 sample resistant (Sherchan et al., 2015, pp. 48-59). Cotrimoxazole shows an increased occurrence of resistance, as compared to other antibiotics, creating potential concerns in usage of this antibiotic over others as a part of the treatment regimen.

In Lithuania (Stacevičienė et al., 2016), a similar study was conducted examining *S. pneumoniae* isolates and their resistance patterns. A total of 367 isolates were included within the study. The majority of isolates (56.7%) were sensitive to all antibiotics tested. Overall, of all samples, 15.8% were non-susceptible to Penicillin, 21.3% to Erythromycin, 16.9% to Clindamycin, and 27.3% to Trimethoprim–Sulphamethoxazole. All isolates were susceptible to Norfloxacin and Vancomycin. In addition, 67.9-82.4% of non-susceptible isolates were those found in pneumococcal conjugate vaccines (Stacevičienė et al., 2016, pp. 4-6). This finding depicts how vaccination can help reduce the prevalence of pneumococcal serotypes exhibiting antimicrobial resistance.

A Mozambique study (Nhantumbo et al., 2016) presented higher rates of resistance as compared to the previous studies. A total of 17 isolates were profiled with susceptibility panels. Of these, 15 were resistant to Penicillin, 11 resistant to Tetracycline, 6 resistant to Chloramphenicol, 4 resistant to Erythromycin, and 2 resistant to Vancomycin. None of the isolates were resistant to Ceftriaxone. However, all of the 17 isolates were resistant to Trimethoprim-Sulfamethoxazole (SXT; Nhantumbo et al., 2016, pp. 5-7). With increased resistance, vaccination is an important preventative tool, since treatment options are becoming more limited within this population.

Policy and Economics

The Advisory Committee for Immunization Practice (ACIP) sets forth the national recommendations for vaccination (Black et al., 2017). This includes utilizing current research to

determine which populations should receive which vaccines, at what age, and how many doses (Annunziata, Rak, Del Buono, DiBonaventura, & Krishnarajah, 2012, p. 1). These decisions reflect not only who is recommended to receive vaccination but also which claims will be reimbursed when submitted through insurance, such as with Medicaid and Medicare (Black et al., 2017, p. 728).

A cost analysis study (Weycker et al., 2016) examined the rates and financial outcomes of pneumococcal disease from three healthcare claims repositories. Within the study, invasive pneumococcal disease (IPD) and pneumonia medical costs were compared among healthy, at-risk, and high-risk patients. In the study, healthy was defined as the absence of comorbidities or risk factors. At risk was defined as immuno-competent with one or more chronic medical conditions. High risk was defined as those who were immune-compromised, immune-suppressed, or had a cochlear implant (Weycker et al., 2016, p. 2). Overall, patients considered at risk or high risk resulted in 3 to 43 times higher healthcare costs per 100,000 person-years. In the 18-64 age group, the difference in the annual cost of pneumonia increased from the healthy group of $1.8 million per 100,000 persons to $13.9 million per 100,000 persons for the at-risk group and to $50.4 million per 100,000 persons for the high-risk group (Weycker et al., 2016). This observed increase in costs can be considered to be multifactorial, due to increased disease incidence, in-patient care costs, and hospitalization duration.

In a multicenter hospital study in South Korea, the mean direct medical costs per case was $7,542 ,which was not significantly different between age groups (Song et al., 2013, pp. 3-4). Hospitalization duration was greatest (22.0 days +/-16.5) in the 18-49 year high-risk group and lowest (15.3 days +/-15.7) in the 18-49 years moderate-risk group (Song et al., 2013, pp. 5-

6). Overall, clinical disease burden increased with age, but medical costs associated with infection had no significant variation among the age-risk groups.

In a Colombian economic study (Ordóñez & Orozco, 2015), the preventive costs for the pneumococcal vaccination was assessed. In that study, a comparison was made between the costs of different types of pneumococcal vaccines and their projected medical costs averted. Vaccinating 90% of the pediatric population with the PCV10 would result in a cost of $28,920,564 but would result in an excess savings of $13,703,271 in medical costs. Additionally, vaccinating 90% of the pediatric population with the PCV13 would cost more, at $31,904,432, but would result in an even greater cost avoidance of $19,479,395 (Ordóñez & Orozco, 2015, p. 6).

A Chinese study (Caldwell, Roberts, An, Chieh, & Bruce, 2015) focused on the economic impact of pneumococcal vaccination during an annual influenza epidemic as well as an influenza pandemic. Within that study, modeling was predicted in terms of routine pneumococcal vaccination of infants. If routine vaccination were to occur in the population, over 75,000 deaths could be prevented from secondary pneumococcal pneumonia during an annual year of epidemic influenza. This number increases to over 700,000 deaths during an influenza pandemic, resulting in an estimated $6.3 billion in avoided cost (Caldwell et al., 2015, pp. 8-9). The study recommended the integration of the pneumococcal vaccine for routine childhood vaccination in China to help reduce healthcare costs during influenza seasons and pandemics.

Willingness to Pay

In a study examining willingness to pay (WTP) for the pneumococcal conjugate vaccine (Prosser et al., 2004), adults' values in disease prevention associated with pneumococcal infection were measured. A series of questions were asked through an algorithm-styled interview,

where vaccine cost and disease outcome questions were based on previous answers. Overall, adults responded with a median WTP of $100 to prevent one episode of otitis media, $400 to prevent one severe episode of pneumonia, and upwards of a median of $500 to reduce the risk of meningitis. The overall median WTP for parent participants was $250 and $300 for community participants (Prosser et al., 2004, pp. 284-288).

Another concern of vaccination coverage is accessibility through insurance coverage and income. In an adult vaccination study in the United States (Annunziata et al., 2012), the National Health and Wellness Survey (NHWS) was used to identify vaccination coverage. The odds ratio for receiving the pneumococcal vaccine was 1.783 for those who possess health insurance. Additionally, income was compared within the study. Income was categorized into three levels: <$25k, $25-50k, and >$50k. Overall, a slight general trend showed decreased vaccination rates among those in lower income levels. For pneumococcal disease, the vaccination odds ratio per income level were <$25k (OR=0.992), $25-50k (OR=1.14), and >$50k (OR=1.144; Annunziata et al., 2012, p. 7).

Conceptual Framework Diagram

The pre-existing frameworks and literature presented here describe a foundational platform for this research study. Furthermore, they help transcribe the development of a model reflective of these themes. Appendix C depicts the socioecologic pyramid of disease. This model depicts an evolutionary theoretical framework, incorporating the previous foundational models mentioned and expanding them to describe the research study of interest. The human (or patient) is at the top, with internal factors making up the middle tier and external factors forming the base tier. This structure presents how external factors can affect middle-tier factors resulting in overall risk of disease for the individual. At the base level, larger scaled factors are presented. These

include population, physical environment, and healthcare policy and economics. These factors present external factors that can affect disease of the individual. Socioeconomic status and access to care are just above the base level, where they have both community and individual level influence. Next is behavior towards vaccination, where this factor is determined by the individual patient. This includes making health-related decisions that affect the overall health of the individual. Finally, towards the top of the pyramid, just below the human, is pathogenesis. This refers to both the organism and act of disease transmission. All these factors are interrelated and the inputs of each factor affect the outcome of disease within the individual.

Finally, there is the conceptual framework diagram, Appendix D, which presents the relationship of variables. This encompasses variables both included and excluded from this research study. This diagram is divided into five areas: patient risk factors, lower socioeconomic status, increased human disease transmission, pathogenies, and disease outcomes. The first four core areas defines the input variables while the fifth core area (disease outcomes) define the output variables. Those variables shaded in green present the variables included within this study. The variables shaded in yellow represent those variables identified within the literature but are not a part of this study due to this information is not routinely collected within epidemiological disease investigations.

The first core area is patient risk factors. In this core area, age and comorbidities are variables of interest within this study. This then includes vaccine history, which can be affected by the patient's eligibility based on age or comorbidities. Not included within this study is trypanophobia, the fear of needles or vaccination. This variable has an impact on vaccine history as well. Finally, in this core area gender and occupation are included as well. Occupation information was not available for this study and therefore was not included in the study.

The second core area is socioeconomic status. Since individual socioeconomic data are not available for case patients, population-level data are represented at the census tract level. The two variables for this core area include median household income and private health insurance coverage. These two variables will be proxies for socioeconomic status and its relationship to disease incidence and risk.

The third core area is human disease transmission. This core area encompasses the variables that can influence the human-to-human transmission model. More specifically, this core area looks at median household size and population density at the census tract level to determine whether this is an increased risk for disease.

The fourth core area is pathogenesis. In this core area, the factors related to the pathogenicity of the disease are examined. This includes infection type, serotype, and antibiotic resistance. These three variables affect the patient prevention and intervention methods for infection. Only infection type was included within this study because serotype and antibiotic resistance information was not be available from the abstracted dataset.

Finally, the last core area is disease outcomes. This core area is different than the first four core areas because it is dependent (as a variable or variables) on the risk factors or independent variables. Variables in this core area include hospitalization duration and mortality, which indicate the morbidity and severity of infection from the disease.

Conclusion

The literature review explored the different areas of pneumococcal disease. This included examining the epidemiology of disease, where morbidity and mortality, high-risk populations, and other population characteristics were reviewed. Additionally, border health and Latin American studies were identified and presented. However, this also identified that there is

limited literature that focuses on these specific populations. Literature in spatial epidemiology was reviewed, depicting the different analysis techniques for geographical disease analysis. The area of prevention and treatment was also incorporated, encompassing vaccination and antimicrobial susceptibility. Policy and economic literature presented barriers and areas of consideration for disease prevention and transmission. Finally, a theoretical framework and a conceptual framework diagram were developed, incorporating these common themes, to describe the research study of interest.

CHAPTER THREE: METHODS AND DESIGN

To complete this research, a retrospective cross-sectional study was developed to assess the epidemiology of invasive pneumococcal disease in Hidalgo County, Texas. As a quantitative cross-sectional study, the same data variables were collected and analyzed for the duration of the identified time frame. Data were primarily analyzed statistically as single-grouped data as a long-term cross-sectional dataset. However, annual case trends were compared with one another in a longitudinal format. Data were analyzed from the years 2010 to 2016. This range in years was selected based on potential external influences, such as vaccine introduction. Therefore, this range limits potential population infection shifts, as compared to including prior years of data. This seven-year dataset was selected to provide information on infection post-2010 PCV13 vaccine introduction, as well as current case counts and details. These data reflected only infected cases, as compared to a case-control study that would examine both populations. Furthermore, with a vaccine now available as prevention, this study examined future populations that need to be targeted for public health interventions.

Research Questions

The literature review presented findings of disease epidemiology and outcomes leading to several research questions. First, with vaccination recommendations established for pediatric, geriatric, and high-risk populations, is the healthy general adult population still at high risk for infection? This is of novel interest since the literature has shown decreased incidence within the pediatric, geriatric, and high-risk groups based on their respective vaccination recommendations.

This leads to the question of whether there is a different population at increased risk that does not currently fall under ACIP vaccine recommendations.

Another question is whether infection type and comorbidities affect final outcomes. Final outcomes can include hospitalization duration and deceased indicator. Therefore, within this population of study, do high-risk patients experience a longer hospitalization duration and higher risk of mortality? With concerns such as diabetes and obesity in this population it is a question whether this can contribute to more severe outcomes from infection.

A third question focuses on population characteristics and infection. This includes examining population-based socioeconomic attributes and comparing them to disease rates. This includes determining whether certain socioeconomic status contributes to risk for infection within this population. Therefore, does socioeconomic status, such as income, influence the rate of disease in a population? Identifying these socioeconomic risk factors can help determine if additional specific interventions need to be addressed. Furthermore, identifying these risk groups for infection can help determine if health equity is present within the community.

Hypotheses

The first research question examined age groups and incidence of infection. Therefore, this epidemiological study assisted in identifying disease trends within each group. This included defining the adult population as *vaccine ineligible* (*vi*) and the pediatric, geriatric, and high-risk populations as *vaccine eligible* (*ve*). In addition, this question then presented the following null and alternative hypotheses. H_0: The vaccine-ineligible adult population (18-64 years) has no significant difference in the probability of pneumococcal disease, as compared to vaccine-eligible populations (H_0: $p_{vi} < p_{ve}$). H_1: The vaccine-ineligible adult population (18-64 years) has

a significant difference in the probability of pneumococcal disease, as compared to the vaccine-eligible populations (H_1: $p_{vi} \geq p_{ve}$).

The second research question compared risk factors to hospitalization duration and recovery status. The literature has presented the notion that the higher risk and older populations correlate to longer hospital stays and increased risk of mortality. Therefore, the following hypotheses are stated. H_0: There is no significant difference in hospitalization duration for pneumococcal disease between the age-risk groups identified (H_0: $\mu_{Pedi} = \mu_{HA} = \mu_{HRA} = \mu_{Geri}$). H_1: There is a significant difference in hospitalization duration for pneumococcal disease between the age-risk groups identified (H_0: $\mu_{Pedi} \neq \mu_{HA} \neq \mu_{HRA} \neq \mu_{Geri}$).

The third research question explored socioeconomic status and the risk factors for infection. The literature has shown lower income populations to have lower vaccination rates and higher infections. Therefore, the following hypothesis tested if this is true for the population of study. H_0: Within the population of study, lower income populations have no increased risk in the rate of infection for invasive pneumococcal disease (H_0: $B_1 = 0$). H_1: Within the population of study, lower income populations have an increase risk in the rate of infection for invasive pneumococcal disease (H_1: $B_1 \neq 0$).

Pre-Data Collection

This research study followed and complied with appropriate protocols and procedures necessary to conduct such research. This included establishing a memorandum of understanding for data exchange and submitting for institutional review board (IRB) approval.

The Office of Research Compliance, Integrity and Safety at Northern Illinois University was contacted prior to data collection and analysis. A screening application was submitted for

IRB review and approval. On May 7, 2018, the Office of Research Compliance, Integrity and Safety determined that this research study does not meet the definition for human subject research and was accepted to continue with this study (Protocol # HS18-0133). To ensure that this research study abides by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), no identifiable data was collected, and no individuals were contacted. Instead, only de-identified medical information was collected, reviewed, and analyzed for the purposes of this research project.

Data and Information Exchange

Data reflected previously collected information as a part of routine epidemiological investigation of public health. Data and information, for the development of this research, were exchanged through collaboration with the Hidalgo County Health and Human Services Department, located in Edinburg, Texas. A memorandum of understanding (MOU) was presented and approved at Hidalgo County Commissioners Court on Tuesday, September 12, 2017. This memorandum allows for the exchange of public health information and data with Northern Illinois University for the purposes of research and prevention of notifiable conditions. Both parties, Northern Illinois University and Hidalgo County, Texas, have agreed upon its terms and information exchange, and this MOU is active until December 2020.

Data Collection

Data collection took place through the exchange of information with Hidalgo County, Texas. Hidalgo County Health and Human Services Department (HCHHSD) provided epidemiological case information on pneumococcal disease and associated laboratory result data for analysis. In addition, supportive information and data were referenced from the United States Census Bureau for population and geographic analysis.

The information from Hidalgo County, Texas, reflected data extracted from the National Electronic Disease Surveillance System (NEDSS). This is a public health information system for data entry and electronic information exchange of notifiable conditions (Center for Surveillance, 2017). This is a national system used by all fifty states where public health workers investigate cases of notifiable diseases, such as invasive pneumococcal disease, and enter this information into the system (Center for Surveillance, 2017).

The data from Hidalgo County was exported using the surveillance system, NEDSS. This occurred by running the Bacterial Meningitis Invasive/Respiratory Disease Report (BMIRD) in NEDSS while incorporating certain parameters. The time parameter was defined as January 1, 2010, to December 31, 2016, defined by collection date. Within the query, the following parameters were set: Jurisdiction equaled Hidalgo County Health and Human Services Department, Morbidity and Mortality Weekly Report (MMWR) year was *equal to or greater than* 2010 and *equal to or less than* 2016. Case status was *not equal to* not a case, so that all types of cases were included (i.e., confirmed and probable). Next, variable selection included demographical and clinical information. No patient-identifying variables were included within the report. Finally, the report was exported as a comma space value (.csv) file.

The American Fact Finder, from the U.S. Census Bureau, was used to access the American Community Survey (ACS) data. A total of five datasets were extracted from this data source. To retrieve economic data, search queries were enacted within the American Fact Finder for median income for Hidalgo County, Texas. Once identified, the table builder was specified to include census tracts, and a .csv file was created and downloaded. Population and demographic data were also retrieved. From this file, data was cleaned up by pulling only necessary variables, such as the median and mean household income, tied to their appropriate census tract. The first

dataset extracted was the *Median Household Income in the Past 12 Months (2016)* ACS report. The second dataset extracted was the *Health Insurance Coverage by Type (2012-2016)* ACS report. The third dataset extracted was the *Educational Attainment (2012-2016)* ACS report. The fourth dataset extracted was the *Hispanic or Latino Origin (2012-2016)* ACS report. The final dataset extracted was the *Place of Birth for the Foreign-Born Population in the United States (2012-2016)* ACS report. A final query was built for population size, so that an incidence rate could later on be calculated. A raw .csv file was downloaded, and data cleanup was also needed for this dataset. The census tract identification codes were transposed into actual census tract numbers through multiple steps of deleting and adjusting prefixes and suffixes. Once completed, all data components were compiled into a single Excel spreadsheet for future analysis.

Population

Data was collected from the Hidalgo County Health and Human Services Department, under a memorandum of understanding for data exchange. This included epidemiological data of invasive pneumococcal disease from 2010 to 2016, to reflect a seven-year timespan. This data was reflected as cases, defined as patients or clients who were reported to the local health department with an identified laboratory culture of *Streptococcus pneumoniae* from a normally sterile site (i.e., blood or cerebrospinal fluid).

Residency was defined as living within Hidalgo County, Texas, at time of specimen collection date. Citizenship status did not affect residency. Therefore, all patients were included within the data analysis as long as the identified case had a Hidalgo County physical or PO box address represented within the medical record data. This included special populations, such as winter Texans and undocumented individuals, who have defined residency within Hidalgo County. The term “winter Texans” refers to individuals from the Northeast, Midwest, amd

Pacific Northwest of the United States or Canada who migrate to Texas during the winter (Rio Grande Valley TX, 2018). However, patients who were under United States Customs and Border Patrol custody at time of specimen collection were considered out of area and excluded as Hidalgo County residents for this study. This is because their identified exposure likely occurred prior to arriving to the United States and would incorrectly contribute data for analysis of this local population and community.

Case Definition

All cases were determined and defined by the Texas Department of State Health Services' *Epidemiology Case Criteria Guide* in regards to case status of invasive pneumococcal disease. This included clinical compatibility and the presence of an identified culture for *S. pneumoniae* from a normally sterile site. Appendix F provides an example of the 2017 Texas DSHS case definition for *Streptococcus pneumoniae* invasive disease (IPD). The case determinations are made through an epidemiological investigation to collection data and information of the infection. Appendix G presents the Texas DSHS case investigation form for streptococcal disease, including invasive pneumococcal disease. Appendix G shows which type of information are collected during a case investigation for this notifiable condition. Patients who have had more than one infection were included once, for their first infection meeting case definition.

Methods in Data Cleanup

The data was first received from Hidalgo County Health and Human Services as a raw .csv file, containing the whole requested dataset. Therefore, this required the necessary steps involved with data cleanup to make this data applicable to this research study.

Data Quality

For the first step in data cleanup, the .csv file of the epidemiology dataset was first converted and saved to an Excel spreadsheet to allow for edits and changes to be saved. Next, each variable was reviewed for completeness and accuracy. Variables that did not contain any entries were deleted and excluded for the purposes of this study. This included information that is not routinely collected within the standard epidemiological investigation. Examples of these variables included daycare history, other vaccine histories (i.e., meningococcal and *Haemophilus influenza* type b), and medical insurance. Next, variables needed for data analysis were reviewed for completeness. Any variables with a data entry completion rate greater than 98% were allowed to stand as is. Variables with less than 98% were reviewed to determine subsequent steps. The primary variables of interest where this threshold was not met was hospital discharge date and its related variable, hospitalization duration. To assist in rectifying this missing data issue, patterns of missing data were assessed to determine if any trends existed in the missing data. In the comment section of many of the cases with missing discharge dates, it was noted that the patient was still admitted into the hospital at the time of the investigation being completed. Therefore, those patients who had prolonged hospitalization durations were at risk for being mispresented in the data. A retrospective review of the data was requested from Hidalgo County to include these missing dates.

All data variables and their content were reviewed to ensure adequate data quality. Data quality began with examining variable completion. This included examining demographical and clinical variables of interest for this study, such as age, gender, zip code, city, ethnicity, and hospitalization status. Additionally, duplicate entries were reviewed and subsequent results were deleted. This was done to ensure that if a patient was infected and reported multiple times, their

data would not construe any outcomes. Data was verified with Hidalgo County Health and Human Services Department to ensure accuracy and completeness. When information was identified to be missing, a request was sent to Hidalgo County, as allowed within the memorandum of understanding. For data with minimal missing fields, data analysis was run normally, while excluding those rows with missing fields. Variables with notable missing values were reassessed to determine the actions of such data analysis.

Another technique in data cleanup was cross-referencing of case investigations and laboratory results. Lab accession IDs were cross-referenced to verify specimen site and infection type. This included verifying that all identified clinical meningitis cases derived from cerebrospinal fluid (CSF) specimen sources.

Data Variables

Data variables were used for numerical and categorical classification. Identification keys were created with no linkage or connection to original patient identification numbers. All patient identification and case numbers were excluded. Variables of interest included both explanatory and response variables. Explanatory variables of interest included*, AdultGeri, Age Category, AgeCatName, Age Reported, Concentrated Race, CurrentSexBinary, EthnicityBinary, Hospital ID, Infection Type, Underlying Conditions*, *Vaccination History, Vaccine Eligible,* and *Risk Group.* For the spatial analysis among the census tracts, variables of interest included *Median Household Income, Private Health Insurance Coverage, Household Size, Educational Attainment, Hispanic Representation, Foreign-Born Representation and Population Density*. Response variables included *Hospitalization Status, Hospitalization Duration, and Deceased Indicator*. *Hospitalization Duration* was a continuous variable while *Deceased Indicator* was a dichotomous variable. A case fatality rate (CFR) was calculated, stratified by age groups.

Comorbidities and risk factors were assessed for those who may be eligible for pneumococcal vaccination but were identified to have no current vaccinated status.

Demographic Data

The following demographic data was extracted for data analysis: age, gender, census tract, city, ethnicity, and race. Gender was a dichotomous variable for both male and female. Age was represented in its numerical value and was also translated into a categorical variable, *AgeCatName*, to compare pediatrics, adults, and geriatrics. The age category variable, *AgCatName*, supported representing vaccination recommendation age groups, such as pediatric and geriatric, compared to non-recommendation age groups. Demographic variables for this study include:

- Age
- Age Category
- Sex
- Census Tract
- City
- Ethnicity
- Race
- Population Density
- Household Size
- Household Income
- Educational Attainment
- Foreign-Born

Ethnic and race data was collected and reviewed as well. Overall data was likely expected to produce minimal outcomes, due to the high single-ethnicity (Hispanic) representation in the county. However, ethnicity and race data was compared among age groups to determine if a higher rate of Caucasian cases occur in the geriatric group. This is because a large winter Texan population is present within the county and helped identify if this special population presented an overall higher incidence of rate in proportion to other resident populations.

Census tract data was used to identify pockets of incidence and disease burden. This data was also paired with US Census data to integrate socialeconomic analysis. All geographical data that was collected and analyzed abided by the Health Insurance Portability and Accountability Act (HIPAA).

Clinical Data

Clinical data variables included Facility/Hospital Location, Hospitalized vs Outpatient, Hospitalization Duration, Infection Type, Deceased Indicator, Vaccine History, Comorbidities, and Symptom Onset. Hospitals were renamed and assigned an ID, such as *Hospital 1, Hospital 2,* etc, to protect their identifiable information within this study. Infection types were divided to include meningitis, bacteremia, and other as their categories. Comorbidities were included individually and also grouped as equal to or greater than one for the creation of a high-risk variable. This high-risk variable was based on risk factors defined by ACIP vaccine recommendations. The variables of study include:

- Facility Location
- Hospitalization
- Hospitalization Duration
- Infection Type
- Deceased Indicator
- Vaccine History
- Comorbidities
- Symptom Onset
- Risk Group

Descriptive statistics were conducted to determine the in-patient versus out-patient frequency for case infections. In-patient settings were defined as including any hospitalization stay with an admission. This definition included both same-day discharge and extended stays. Visits to the emergency department were defined as out-patient visits. Hospitalization duration was also assessed. Hospitalization duration was examined in correlation to explanatory variables of interest. This included hospital facility and age-risk group (i.e., pediatric, health adult, high-

risk adult, and geriatric). Age-risk groups included identifying and representing comorbidity presence as a high-risk adult category (i.e., immunocompromised-conditions, diabetes, and chronic conditions). Hospitalization duration was also be compared among hospital facilities to determine if any hospital has a significant difference is duration, as compared to others.

Vaccination history was assessed to determine the following objectives. First, to identify if adequate data is available to identify vaccine history in patients. This was completed, since it was suspected that many patients were hospitalized and that their primary care provider information was unavailable. Therefore, if the patient was not in the state immunizations registry, his or her vaccination status for pneumococcal disease would likely be unknown. Second, for those who were eligible for vaccination, their history can be examined to determine if they received the recommended vaccinations.

Age-risk groups were categorized as well, based on comorbidities and risk factors. This included healthy and high-risk. As mentioned earlier, the four groups included pediatric, healthy adult, high-risk adult, and geriatric. Healthy adult was defined as patients (18-64 years) who do not currently fall under any recommendations for pneumococcal vaccination. High-risk adult was defined as patients (18-64 years) with at least one factor or condition identified as a risk factor indication for pneumococcal vaccination, excluding age. A high risk variable, *HighRisk*, was established based on the presence of one or more high-risk comorbidities in a patient, as defined by ACIP vaccine indications. This was created as a dichotomous variable, where 1 represented high-risk and 0 represented healthy (no risks identified). A separate variable, *VaccineEligible*, was also created as a dichotomous variable (*0=no, 1= yes*) to represent all factors, both comorbidities and age, of vaccine eligibility.

Mortality was assessed based on associated explanatory variables. This included analyzed comorbidities to determine their effect on deceased indicator. More specifically, the abundance and type of comorbidities were examined to identify areas of higher risk for mortality. Mortality was compared between hospitals to identify if significant proportions or risk ratios are present, based on hospital facility.

Reporting Data

Reporting measures were also examined to determine if private provider and public health investigation are taking place in a timely manner. As per Texas Department of State Health Services and Texas Administrative Code, notifiable diseases must be reported to the local health department within a designated time frame. This is to ensure adequate response and intervention time by public health staff. Invasive pneumococcal disease must be reported to a local health department within seven days of an identified case by a private provider (Texas Department of State Health Services, 2017). Once reported, the local health department has thirty days to complete the investigation and submit it to the state health department. Both these numerical timeline measures were evaluated to determine if timeliness is taking place for proper public health surveillance and intervention.

Data Analysis and Software

Data analysis took place with Stata (v14.2 StataCorp), Excel, and ArcGIS (v10.6). These three forms of software supported statistical analysis, data visualization, and geographical analysis. This data analysis also tested the three hypotheses of this research study, as well as other supportive analyses. Although QGIS software provided a strong platform for case and incidence mapping, once statistical and spatial analysis was needed, this software did not provide full capability. This included conducting ordinary least squares (OLS) and geographic weighted

regression (GWR). Therefore, a student license of ArcGIS was procured to more efficiently handle these spatial analysis techniques.

Data was first graphically examined to determine if normal distribution is present. Visual data included box plots that were created to identify the range, quartiles, and median. This also helped visualize the spread of the data. Specifically, this was completed for variables such as age and hospitalization duration. Other data visualization methods included performing quantile-quantile plots to determine if the data is normal. Histograms were created as well. Histograms were created to visually compare gender incidence, age, and overall annual disease incidence. Longitudinal comparisons were made, based on annual case trends and subgroups within these annual trend analyses.

Statistical Analysis

Descriptive statistics were performed to present the population of study. The annual case counts were depicted in a histogram to show case trends. Then a table was created to depict disease case representation by sex, age, ethnicity, underlying conditions, hospitalization duration, deceased status, and infection. To further examine case representation, box plots were created for annual age distribution. Cases were also examined as seasonal trends to identify which months had a greater case occurrence over the seven-year period. Different statistical analysis techniques were used to identify the population of study and to test the hypotheses previously addressed. In this study, three hypotheses were of focus and tested to examine vaccination history, hospitalization, and population characteristics. These hypotheses were also supported through additional analysis techniques, such as post hoc and spatial analysis.

Data analysis for this research study also focused on the three presented hypotheses. This included calculating coefficients and odds ratios (ORs) to analyze which characteristics

presented the greatest risks for infection. This was completed to compare age categories, risk groups, hospital facility, and vaccine eligibility. Additionally, stratified analysis took place to identify if any effect modification or confounding bias exists. This included conducting stratification among demographical characteristics and comparing these variables among the overall coefficients and odds ratios. Furthermore, census population data was incorporated as well to assess and identify population-level risk factors for infection. This included utilizing demographical and socioeconomic data to assess disease incidence.

1st Hypothesis Methods: Vaccine Eligibility in Cases

Initial statistical analysis focused on the first hypothesis of this study. This hypothesis is concerned with the case occurrence of vaccine eligibility among the age-risk groups of the population within this study. To begin, descriptive statistics were created, focusing on the demographics identified. A table was created to summarize the overall case representation of infections. This included presenting case counts by age group (i.e., pediatric, adult, and geriatric) and then depicting case breakdowns by sex, ethnicity, and infection type.

To support this hypothesis, the variable *Age* was assessed as both numerical and categorical variables. To further assess the numerical age values, an age categorical variable was created to allow for further analysis. The pediatric group included cases under the age of 18, the adult group included ages 18 to 64, and the geriatric group included 65 years and older. The variable *AgeCatName* represented a string name for the three categories, which was then transcribed into the following numerical categories: pediatric (1), adult (2), and geriatric (3) for the variable *AgeCategory*.

All 329 cases were assigned a vaccine eligibility status. This was accomplished by reviewing medical risk factors in the case data entry, the age, and other social risk factors that

determined ACIP and CDC vaccine indications. To determine vaccine eligibility and indication, descriptive statistics were reviewed for the proportion of cases that met vaccine eligibly based on the age or risk factors. A variable for vaccine eligibility was created as a dichotomous variable. This included a value of 1 for "yes" and a value of 0 for "no" in regards to a vaccine indication for the pneumococcal vaccine. Eight of the cases were identified as having disease exposure prior to two months of age, making them ineligible for vaccination. Therefore, these eight cases were excluded from this portion of the analysis to not misrepresent disease risk in the vaccine-eligible pediatric population. A 2x2 chi-square cross-tabulation association table was created to compare the frequency of vaccine history and vaccine eligibility. This cross-tabulation was created to help identify if these primary cases of disease could have been preventable through vaccination, if a vaccinated population was still infected, or if the population did meet vaccination indications. Next, binary logistic models were run to assess hospitalization and deceased status. Patients were categorized into a dichotomous variable on whether or not they had at least one comorbidity that was considered high risk for infection. Then binary logistic regression were run for both hospitalization status and deceased indicator to determine if there was any increased risk for either based on comorbidity presence.

To test the first hypothesis, a binomial probability test was performed upon certain parameters. First, the binomial test examined the adult versus geriatric age groups to identify if there was any significant difference between the two groups. This analysis looked specifically at age and did not compare comorbidities or other health risk factors. Next, a binomial probability test was also performed to determine if there was no significant difference in case probability between the vaccine-indicated population and the vaccine-ineligible population. Finally, a post

hoc binomial test was examined specifically for the 18-64 age group. The case occurrence was compared between the vaccine-eligible and vaccine-ineligible groups.

2nd Hypothesis Methods: Hospitalization Duration in Risk Groups

The second hypothesis is concerned with disease outcomes by age and risk of infection. Before testing the 2nd hypothesis, descriptive statistics were conducted to identify the attributes of hospitalization and mortality. Whisker-box plots were created to compare the groups of study interest (i.e., pediatric, adult, high-risk adult, and geriatric) in terms of hospitalization duration (in days). Other data visualizations included box plots of overall hospitalization duration by hospital facility. As with the 1st hypothesis, the numerical age and hospitalization duration continuous variables were assessed to determine if these variables were normally distributed. In this hypothesis, hospitalization status, hospitalization duration, and mortality were the three outcome variables of interest.

First, hospitalization status was assessed. This was defined as being admitted into a hospital facility. Emergency departments without admission were not considered as hospitalized. Descriptive statistics were preformed to define the age-risk groups that were hospitalized. Then, a chi-square test of association was performed to compare hospitalization rates of these groups. Next, binary logistic regression was then performed to assess hospitalization status among high-risk versus generally healthy groups. A binary logistic regression model was also conducted to assess age and sex.

After hospitalization status was analyzed, hospitalization duration was examined next. The count, range, and average was calculated of overall hospitalization durations, as defined in days. As mentioned earlier, a box plot was created to identify the duration of each hospital facility. To assess normality of the data, a Shapiro-Wilk test was run. Linear regression was

conducted to compare the numerical age and lengths of hospitalization duration. Additional analysis included conducting a one-way ANOVA for age categories and hospitalization duration. This one-way ANOVA was used to compare the means of hospitalization duration between age groups to determine if disease severity is similar or significantly different across age groups. The same linear and ANOVA analysis methods were conducted to compare the four identified age-risk groups of this study. To account for non-normal data, a non-parametric method was utilized. The Kruskal-Wallis test was used to compare hospitalization duration among the eight acute-care hospitals located in Hidalgo County. This test method was also used for hospitalization duration and age-risk group.

Finally, for this 2nd hypothesis, case fatality was also analyzed to determine if any specific population groups or other factors contribute to final patient outcome. Descriptive statistics examined overall mortality and also specifically by infection type and age group. A binary logistic model was run to assess mortality with age and sex. This logistic regression was conducted to statistically identify if certain age-risk groups were more at risk for a morality outcome.

3rd Hypothesis Methods: Spatial Analysis of IPD Incidence

The third hypothesis examined the influence of socioeconomic status and disease incidence. Socioeconomic status was defined by median household income. This data variable is defined at the population level of census tracts. Hidalgo County has 113 census tracts, and United States Census Bureau data was incorporated to assess if income influences invasive pneumococcal disease risk and incidence.

Census tract information was collected from the United States Census Bureau. This information provided demographic data for the population of interest. The mapping software was

first layered with a base layer of the county. This included layering a .shp file to place the county and its census tract boundaries onto the map. The .kml file type is a Keyhole Markup Language file for geographic annotation and visualization. Shapefiles (.shp) were extracted from the United States Census Bureau to create a mapping vector layer for the census tracts within Hidalgo County. Other boundary shapefiles and .kml files were received from the Hidalgo County Health and Human Services Department and Hidalgo County Planning Department.

Population-level information was collected from the American Community Survey (ACS) from the United States Census Bureau. This included the median household income, average household size, private health insurance coverage (%), and land area. Land area (square miles) was collected to calculate population density. The total land area for each census tract was divided by the population estimate of the ACS. A new datasheet was created to combine the population-level data from the ACS. The incidence rate was also included. This was calculated by creating a case count per census tract. The count was then divided by the 2016 population estimate of the census tract. This created the initial incidence variable. Next, the incidence was multiplied by 10,000 and created as an additional variable, *IncidenceX10,000*. This was completed to provide a more workable numerical figure and to determine case incidence per 10,000 individuals. The range of incidence was calculated from both the cities and census tracts. The incidence of cities was also compared with one another.

To begin, descriptive statistics were analyzed to present data on the census tracts and their characteristics. Techniques included examining range, mean, and median of the associated census data. The population variables studied included median household income, private insurance coverage, household size, and population density. The mean and population range, minimum and maximum values were calculated from the census tracts. The mean and range were

also calculated for the median household income, private health insurance coverage, and household size.

Economic and spatial analysis. Economic and spatial analysis was conducted to support the third hypothesis. Spatial analysis was performed to create data visualization of the statistical outcomes. Data was collected from an information exchange with Hidalgo County, which included geographical variables (i.e., census tract and city). Since all geographical data derived from one county, no identifiable information was depicted. Techniques were used to ensure to stay in compliance with the Health Insurance Portability and Accountability Act (HIPAA). Geocoding was used so that the extracted data from Hidalgo County Health and Human Services Department (HCHHSD) did not contain any identifiable information connecting a patient to disease. This included using broader measurements, such as census tract and city, instead of patient address.

As mentioned earlier, ArcGIS version 10.6 was used as the software platform for the spatial analysis component of this data. To begin, new variables were created to intersect case count and population density. Census tracts were used as a marker for economic impact for disease. Although individual patient income was not be collected, census tract of residence was. Therefore, data was analyzed by comparing census tract median household incomes with the occurrence of disease. Population density was analyzed along with incidence, rather than case counts, so that a larger population did not create a bias in the statistical outcomes. This was accomplished by creating a variable, *CTIncidence*, so that there is an incidence calculated for each census tract studied. This incidence was calculated based on case and population data. Performing this analysis helped determine if income levels have an influence on disease risk.

The shapefile of Hidalgo County census tracts was attempted to be joined to the Excel data of incidence and socioeconomic data. However, the data would not join. Therefore, the attributes and properties of the shapefile were reviewed and identified to be string data. Therefore, the census tract labels and names were converted from numerical text values to text so that it could be recognized in the join step of ArcGIS. A unique ID integer variable was required to perform OLS. This was created and established in Excel.

A choropleth was mapped on case incidence for the census tracts in Hidalgo County. As with the table earlier, case incidence for this choropleth was calculated per 10,000 individuals. Then a series of choropleths were created to depict six of the explanatory variables of interest. Hot and cold spot analysis was performed, based on incidence within each census tract. A median and a mean center were also mapped to describe the center of overall disease representation. Next, spatial autocorrelation of variables was performed, utilizing Moran's index. This was utilized to identify the presence of case clustering among census tracts for the following variables: median household income, average household size, private health insurance coverage, population density, high school completion, bachelor's degree completion, Hispanic ethnicity, and foreign-born representation. Case incidence was also examined in the spatial autocorrelation analysis.

Next, ordinary least squares (OLS) was conducted to examine the four explanatory variables of interest and their overall relationship to case incidence in each of the census tracts. These seven explanatory variables included median household income, private health insurance, household size, educational attainment, ethnicity, foreign-born representation, and population density. An OLS model was created, incorporating all seven explanatory variables to explore if any had significant outcomes. Once the model was run, the following was identified: median

household income and private health insurance coverage were of interest to identify if any relationships existed with socioeconomic status and health insurance coverage in regards to disease incidence.

Geographic weighted regression was performed as an additional method for analysis. In this model, the same four variables (median household income, private health insurance, household size, and population density) were included within the analysis. Finally, a linear regression model was conducted in Stata to compare to ArcGIS and to test the null and alternative hypotheses. This linear regression model examined the incidence of the 112 census tracts with median household income. This included comparing the two continuous variables of income and incidence rate.

Summary

In summary, this research was a cross-sectional epidemiological study. The population assessed was represented through patients of invasive pneumococcal disease in Hidalgo County, Texas, a United States-Mexico border community. Through the establishment of a memorandum of understanding, data was exchanged and collected from Hidalgo County Health and Human Services Department. This dataset was utilized for secondary analysis to identify risk population for invasive pneumococcal disease. Primarily, this research study focused on age demographics, risk factors, and socioeconomic influences to determine presence of increased incidences of disease. This research study was conducted to identify which populations are at most risk for infections and whether these groups align with current vaccine recommendations.

CHAPTER FOUR: RESULTS

The primary objective of this research study was to examine the disease burden and population characteristics for infection of invasive pneumococcal disease. More specifically, this included identifying if certain age groups were most susceptible to infection and which groups were at higher risk of severe complications of infection. Furthermore, other research objectives included comparing subpopulations to identify if socioeconomic factors contribute to disease incidence. This included utilizing United States census data along with spatial analytics to determine population-based risk factors for disease. Population density, household size, median household income, and private health insurance coverage were data variables analyzed from the United States Census Bureau.

The first step before data abstraction was completing the Institutional Review Board (IRB) submission process. A screening application for IRB was submitted to Northern Illinois University's Research Compliance Office. From there, it was identified that this research project did not encompass human subject research, as described in their determination notice. This then allowed for official data abstraction from Hidalgo County.

Data Collection

Data was abstracted from Hidalgo County Health and Human Services, located in Edinburg, Texas. This data information exchange process included a dataset transfer, where all identifiable information was completely omitted from the report. Identifiable information that was excluded from this exchange included, but was not limited to, *patient name, patient address,* and *patient phone number*. This data information exchange took place through a memorandum of

understanding (MOU) with Hidalgo County to exchange de-identified data for the purposes of this research project. A total of 333 records were obtained in a comma space value file. This dataset included the variables from the public health case investigations conducted by Hidalgo County Health and Human Services for invasive pneumococcal disease case reports from 2010 to 2016. Case investigations conducted by the epidemiology and surveillance team at HCHHS included gathering and collecting medical history, risk factors, vaccination status, and disease outcomes. The cases were reported from medical providers who had a positive culture of *Streptococcus pneumoniae* in a patient residing in Hidalgo County.

Data was also collected from the United States Census Bureau. This included demographical data of Hidalgo County and its census tracts. The American Fact Finder was used from the U.S. Census Bureau to access the American Community Survey (ACS) data. A total of five datasets were extracted from this data source. The variables extracted from these datasets included population estimate, median household income, average household size, private health insurance coverage, population density, educational attainment, Hispanic representation, and foreign-born representation. Once completed, all data components were compiled into a single Excel spreadsheet for additional analysis.

Data Cleanup

As an initial step of data cleanup, the comma space value file of the epidemiology dataset was first converted and saved to an Excel spreadsheet to allow for edits and changes to be saved. In the pneumococcal disease datasheet, there were a total of 140 variables (columns) and 333 case patients (rows). Of the 140 columns, 59 were background geospatial data. This included examples like geocoding source and census that were not directly actively involved with the

analysis of data. In addition, many variables were duplicated to account for variations in data analysis, such as conversion into categorical, binary, or string variables.

Some variables required data cleanup, such as hospital discharge dates. A total of 241 cases were identified as hospitalized. Of these cases, 75 had a missing discharge date. An analysis of pattern for missing data was completed to identify if any trends existed in discharge dates missing. A subsection of late 2016 cases were identified with this pattern. Furthermore, the comment variable of the dataset was reviewed and cross-referenced to identify potential common reasons for missing discharge dates. Comments included the patient still being hospitalized at the time of the case being completed and closed, as well as notation of expiration. A total of six cases had notation of the patient expiring. Of these six cases, four were admitted into a hospital, and these discharge dates were updated to reflect this information. In addition, a retrospective review was initiated with Hidalgo County to limit the missing values.

Data cleanup also included verifying all other variables that may have had missing information. For example, if there was mortality information, such as deceased date and time, the deceased indicator was updated from blank to yes. Another example included if the comment section included the specific comorbidities and risk factors, this was then transferred and recorded into the appropriate variables.

Demographic data was then reviewed for completeness. Age, sex, city, census tract, and zip code were all complete, with zero missing values. City, zip code, and census tract information was also cross-verified to ensure that all values correlated to correct values that could be located within Hidalgo County, Texas. This encompassed comparing these variables' data entry fields with the United States Census Bureau's defined jurisdictions within Hidalgo County.

Another section of data cleanup encompassed string variable conversion. In instances where string variables were provided, an additional variable was created as a numerical value. One example of this included assigning a hospital ID. Each hospital name was assigned an identification number. The variables *Hospitalized* was numerically assigned zero for no, and one for yes. *Gender* was assigned a zero for females and a one for males. For the variable *AdultGeri*, adults were assigned a zero and geriatrics a one. Zero and one were selected as the binary values to allow for binomial and other numerical testing. Variable categories were also named and assigned to align with the hypotheses of this research study. The numerical age value was assigned an age category. This included an additional age category that was created for all three age groups, *AgeCategory*, representing pediatric (1) for under 18 years old, adult (2) for 18-64 years old, and geriatric (3) for 65 years and over.

Within the Excel database containing the United States Census Bureau census tract information, variables were reviewed and cleaned. A foundational dataset was established by creating a variable of all census tracts in Hidalgo County. A total of 18 variables were created, encompassing both Hidalgo County case data and American Community Survey (ACS) data. This included, but was not limited to, population estimate, median household income, average household size, private health insurance coverage, land area, population density, educational attainment, ethnicity representation, foreign-born representation, case count, and incidence.

A total of 333 observations were identified in Stata. A pivot table was performed in Excel to identify any duplication of case identification numbers. A total of four patients had two infections during this study's timeframe. Therefore, the second infection was removed for the purposes of this study's data analysis. This was completed to prevent bias in the occurrence of infection representation.

Data was then imported into Stata to begin analysis. The Stata/IC version 14.2 software package was utilized for the data analysis of this study. This was accomplished by importing the dataset as an Excel spreadsheet, with the first row marked as variable headers. The dataset was also imported into ArcGIS for further spatial analysis.

Overview of Data

A descriptive statistical analysis was performed to identify overall findings of the data. A total of 329 unique cases were identified in Stata. Four cases were identified to have two infections each. Therefore, the latter infection was removed from all further analyses to prevent bias or data inflation. For this dataset, the range of annual case counts was as low as 35 cases in 2010 to as high as 59 in 2011 (see Figure 1). Other annual disease characteristics, such as sex, ethnicity, underlying conditions, hospitalization duration, deceased indicator, and infection type, were reviewed (see Table 2).

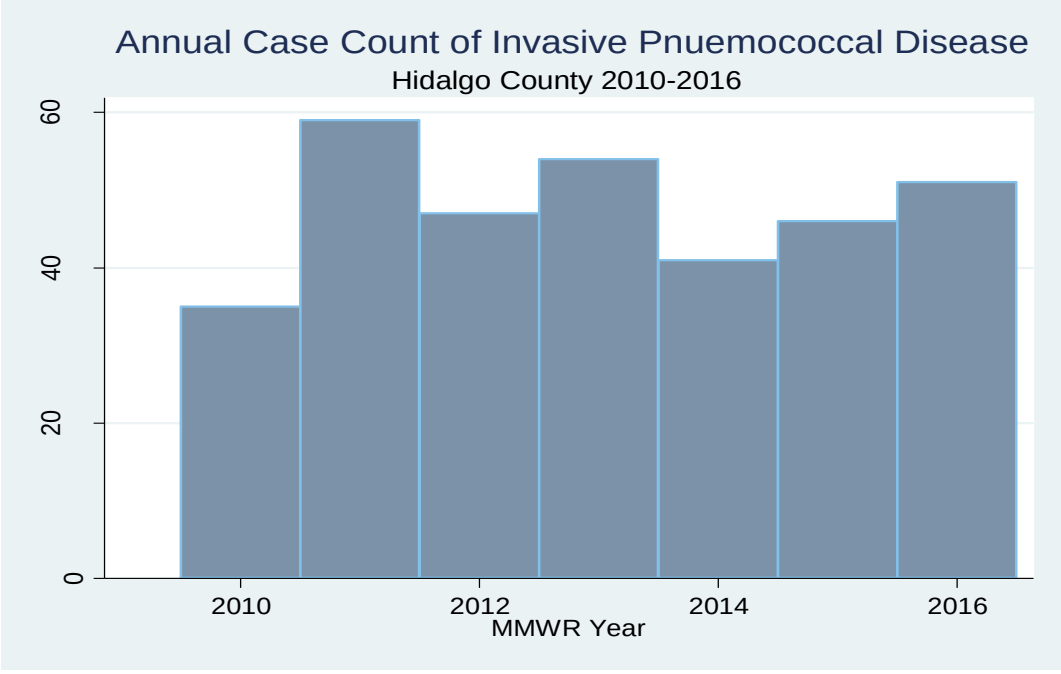


Figure 1: Annual Case Count of Invasive Pneumococcal Disease. Cases defined by Texas DSHS Epidemiology Case Criteria Guide.

Table 2: Population Demographics and Characteristics

	2010	2011	2012	2013	2014	2015	2016
Cases	35	59	47	54	41	46	51
Sex (n)							
• Male	20	28	24	27	22	23	29
• Female	15	31	23	27	19	23	22
Median Age	52.0	50.0	61.0	58.0	60.0	57.5	59.0
Hispanic/ Latino Ethnicity	31 88.57%	50 84.74%	39 82.98%	40 74.07%	38 92.68%	37 80.43%	43 84.31%
Underlying Conditions Present	7 20.00%	38 64.41%	32 68.09%	39 72.22%	33 80.49%	33 71.74%	28 54.90%
Avg. Hospitalization duration (Days)	1.667	4.810	4.625	5.963	12.294	10.441	6.324
Deceased (n / %)	3 8.57%	6 10.17%	2 4.65%	3 5.56%	2 4.88%	3 6.52%	3 5.88%
Infection Type (n)							
• Bacteremia	33	52	42	53	40	45	51
• Meningitis	2	1	1	1	1	1	0
• Other	0	6	0	0	0	0	0

The overall median age of the seven-year data set was 57.00 years old, with a mean of 48.81 years. Age of cases ranged from as young as 3 days old to up to 100 years old. The 25th percentile age was 26, and the 75th percentile was 73 years old. The youngest median age occurred in the year 2011 at 50 years old, and the oldest median age in 2012 at 61 years old. These trends were also graphed annually to show the annual distribution of age through box plots (see Figure 2). This figure also depicts where in 2011 and 2016 a larger proportion of pediatric

cases under the age of five occurred. The primary race and ethnicity was Hispanic/Latino with 90.27% of cases (n=297), and 9.73% as not Hispanic/Latino (n=32).

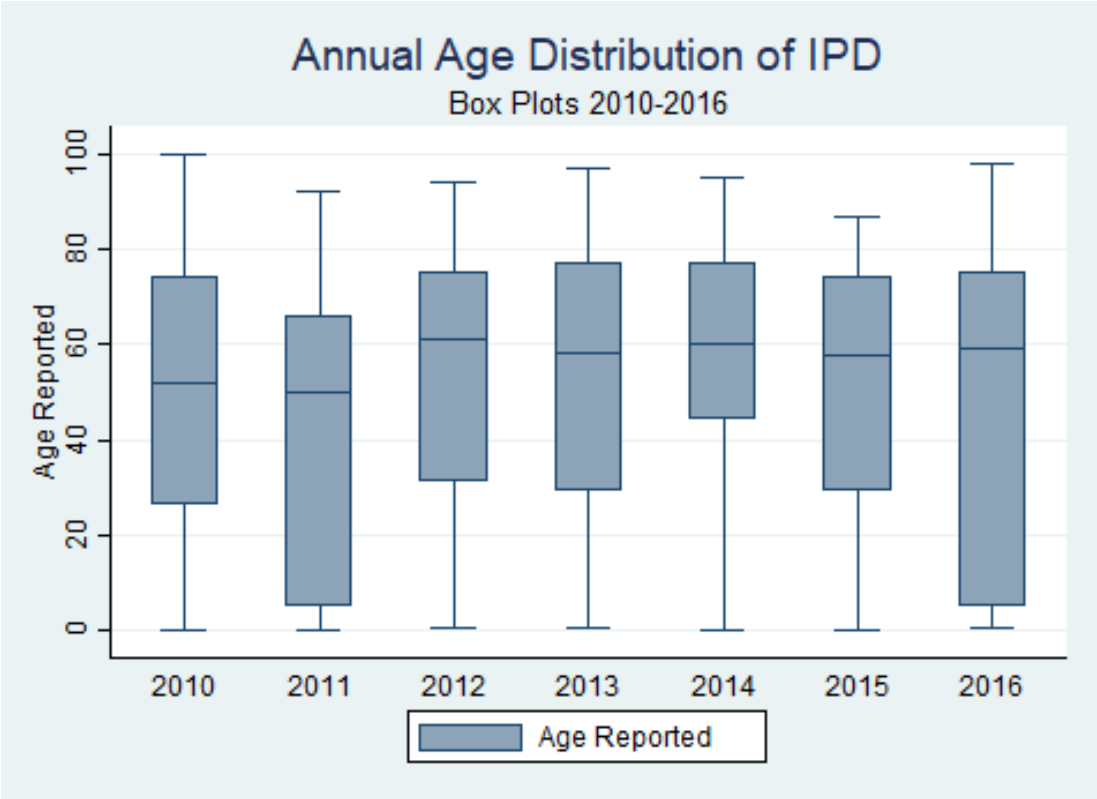


Figure 2: Annual Age Distribution of IPD Cases

For the seven-year period, 157 cases (47.72%) were female and 172 cases (52.28%) were male. The year 2011 was the only year when the female case count (n=31) was higher than the male case count (n=28). Of the 329 cases of invasive infection, 96.05% (n=316) were bacteremia infections, 2.13% (n=7) were meningitis infections, and 1.82% (n=6) were either classified as other or unknown infections. Cases occurred primarily in the fall and winter months. December had the highest monthly average (7.43 cases), and June had the lowest monthly average (1.86 cases). December 2014 had the highest case count frequency, occurring in a single month. Cases

were four times (4.0000) more likely to occur in December versus June. The months of May to August presented the lowest overall case counts (see Figure 3).

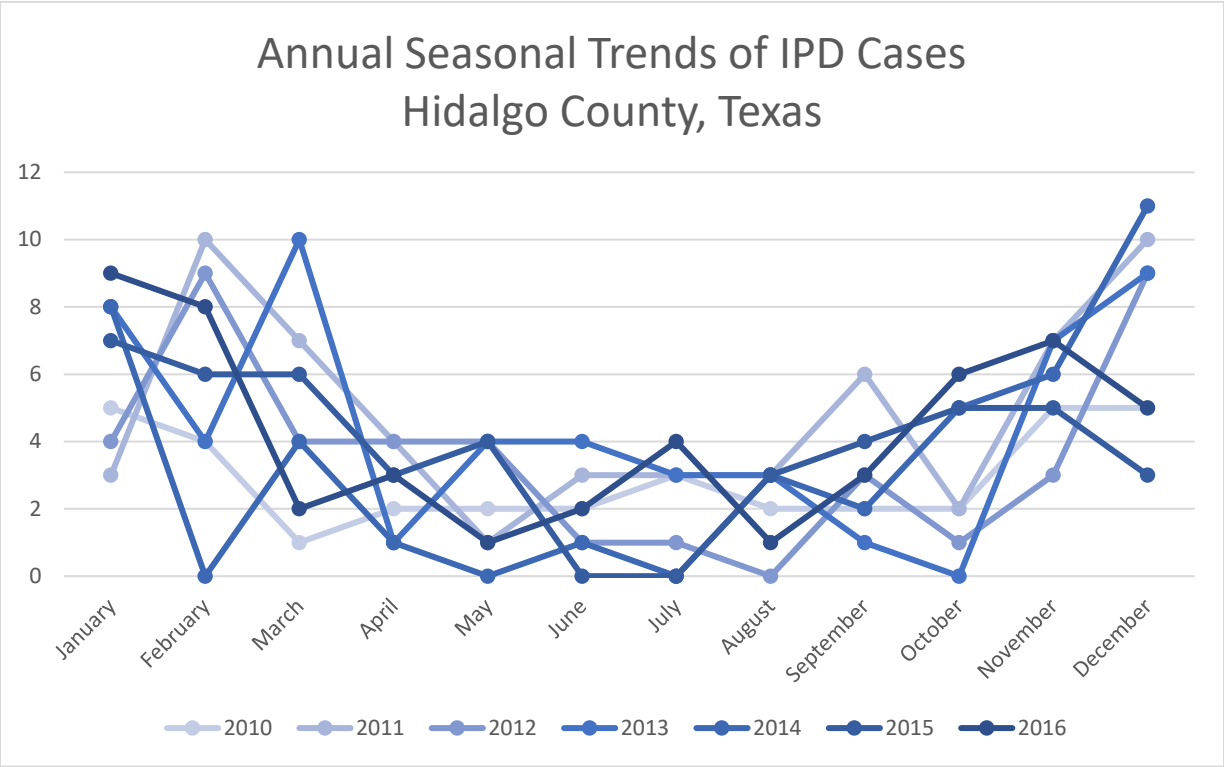


Figure 3: Annual Seasonal Trends of IPD Cases. Dates defined by specimen collection date.

Hypothesis 1 Results: Vaccine Eligibility in Cases

As identified in the methods section, the following is the first null and alternative hypotheses. H_0: The vaccine-ineligible adult population (18-64 years) has no significant difference in the probability of pneumococcal disease, as compared to vaccine-eligible populations (H_0: $p_{vi} < p_{ve}$). H_1: The vaccine-ineligible adult population (18-64 years) has a significant difference in the probability of pneumococcal disease, as compared to the vaccine-eligible populations (H_1: $p_{vi} \geq p_{ve}$). These hypotheses focus on vaccine eligibility and disease probability within these subgroups of the population.

Age and Comorbidities

A total of 329 cases had a numerical age that was transcribed into either pediatric, adult, or geriatric. As mentioned earlier, *AdultGeri* was another variable established for binary analysis of the adult and geriatric populations. Over the seven-year timespan, the adult age group was most common, with 131 cases, followed by geriatric (n=121) and pediatric (n=77). The three age categories (pediatric, adult, and geriatric) were described by sex, race, and infection type (see Table 3). The 18-64 age group was the most common for males, females, and Hispanic ethnicity. Non-Hispanic ethnicity was the most common, by both count and rate, among the 65 years and older age group.

Table 3: Age Category Characteristics and Distribution

Parameter (Age Category)		0-17	18-64	65+
• Sex				
	Male	45 / 58.4%	64 / 48.9%	63 / 52.1%
	Female	32 / 41.6%	67/ 51.1%	58 / 47.9%
• Ethnicity				
	Hispanic	74 / 96.1%	120 / 91.6%	103 / 85.1%
	Non-Hispanic	3 / 3.9%	11 / 8.4%	18 / 14.9%
• Infection Type				
	Bacteremia	75 / 97.4%	124 / 94.7%	117 / 96.7%
	Meningitis	1 / 1.3%	5 / 3.8%	1 / 0.8%
	Other	1 / 2.3%	2 / 1.5%	3 / 2.5%

The presence of underlying conditions was assessed. Of the overall population, 62.61% (n=206) had at least one underlying condition documented. Documented underlying conditions included, but were not limited to, diabetes mellitus, renal failure, liver cirrhosis, heart disease, heart failure, chronic lung disease, and cancer. Binary logistic models were used with these

underlying conditions as independent variables and with each hospitalization status and mortality as the dependent variable. To meet the criteria of having an underlying condition in this study, the case patient needed one more of the conditions noted above. Underlying conditions did not significantly increase the risk of mortality, with an odds ratio of 1.048 ($p = 0.918$). However, there was a significant increase in the odds of hospitalization for those with underlying conditions, with an odds ratio of 2.19 ($p = 0.02$).

$$logistic(p_{hospitalized}) = \log\left(\frac{p_{hospitalized}}{1 - p_{hospitalized}}\right) = \beta_0 + \beta_1 X_{comorbidity}$$

$$logistic(p_{deceased}) = \log\left(\frac{p_{deceased}}{1 - p_{deceased}}\right) = \beta_0 + \beta_1 X_{comorbidity}$$

Vaccine Eligibility

A variable for vaccine eligibility was created as a dichotomous variable. This included a value of 1 for "yes" and a value of 0 for "no" in regards to a vaccine indication for the pneumococcal vaccine. This eligibility was determined by current 2010 to 2016 vaccine recommendations and patient risk factors, such as age and comorbidities. All 329 cases were assigned a vaccine eligibility status. This was accomplished by reviewing medical risk factors in the case data entry, the age, and other social risk factors that determined ACIP and CDC vaccine indications. Of the 329 cases, 321 cases were included in hypothesis testing. Eight of the cases were identified as having disease exposure prior to two months of age, making them ineligible for vaccination, and therefore were excluded from this portion of the analysis to not misrepresent disease risk in the vaccine-eligible pediatric population.

Vaccine eligibility was analyzed for a total of 321 case infections. A total of 37 (11.53%) patients had an identified vaccination history from the case investigation outcomes. A total of

284 (88.47%) patients had no vaccine history identified within the case findings. However, after review of age, comorbidities, and risk factors, an additional 195 patients were eligible for vaccination. Of these 195 patients eligible but not vaccinated, 119 were 65 years or older. In addition, 38 patients were 18 to 64 years old with at least one indicated risk factor. Finally, an additional 38 were pediatric patients with no previous vaccine history identified.

A chi-square cross tabulation table was created to compare the frequency of vaccine history and vaccine eligibility. The Pearson chi-square test equaled 10.8380, with 1 degree of freedom. This resulted in a significant p-value from both Fisher's exact test ($p < 0.0005$) and the chi-square test ($p = 0.001$), presenting dependence among vaccination eligibility and history (see Table 4).

Table 4: 2X2 Table of Vaccination History and Vaccine Eligibility

Vaccination History		Vaccine Eligibility		
		Vaccine Ineligible	Vaccine Eligible	Total
Not Vaccinated	Count	89	195	**284**
	%	31.34%	68.66%	**88.47%**
Vaccinated	Count	2	35	**37**
	%	5.41%	94.59%	**11.53%**
Total	**Count**	**91**	**230**	**321**
	%	**100%**	**100%**	**100%**

The two cases that were vaccinated without indication included two cases (aged 60 and 62) who did not have any documented risk factors. An additional case was identified from the eight infant cases, where one infant just under two months was also vaccinated slightly before the vaccine-indicated time frame age of two months.

To test the first hypothesis, the binomial probability test was selected as the inferential statistic function to compare the expected probability of subgroups of invasive pneumococcal disease cases with one another. First, the binomial test examined the adult versus geriatric age groups to identify if there was any significant difference between the two groups. This analysis looked specifically at age and did not compare comorbidities or other health risk factors. A total of 252 observations occurred, with an expected *k* of 126 cases. A total of 121 cases were observed as geriatric cases and 131 as adult cases. Therefore, the observed probability of geriatric case occurrence was 0.4802, compared to the assumed probability of 0.5000. Furthermore, there was no significant difference ($p = 0.2854$) in case occurrence between adult and geriatric patients when no other contributing risk factors are considered.

$$P(X) = \frac{n!}{(n-X)!\,X!} \times (p)^X \times (q)^{n-X}$$

A binomial probability test was also performed to determine if there was no significant difference in case probability between the vaccine-indicated population and the vaccine-ineligible population. A total of 321 observations were included in this analysis. Again, eight cases were removed from this analysis since they did not meet the pediatric age criteria for vaccination. All eight cases were less than three months old. The probability of success for this binomial test was set at 0.5 to represent an equal probability between both groups. In this analysis the observed *k* for vaccine eligible was 230, with an expected *k* of 160.5, translating to

an observed probability of 0.7170. This resulted in a significantly greater case occurrence within the vaccine eligible subgroup, compared to the vaccine-ineligible subgroup ($p < 0.00005$).

Finally, a post hoc binomial test was examined specifically for the 18-64 age group. The case occurrence was compared between the vaccine-eligible and vaccine-ineligible groups. Between both groups, a total of 131 observations occurred, which resulted in an expected *k* of 65.5, representing a 0.5 standard equal probability. The case occurrence of the vaccine-eligible group in this age group was 30.53% (n=40), while the vaccine-ineligible case occurrence was 69.47% (n=91). This resulted in a significantly lesser case count within the vaccine-eligible subgroup, as compared to the vaccine-ineligible group, among the 18-64 age range ($p < 0.00005$).

Hypothesis 2 Results: Hospitalization Duration in Risk Groups

In the next section of study, it is of interest where hospitalization and hospitalization duration is significantly different between the different risk groups, as noted in the following null and alternative hypotheses. H_0: There is no significant difference in hospitalization duration for pneumococcal disease between the age-risk groups identified (H_0: $\mu_{Pedi} = \mu_{HA} = \mu_{HRA} = \mu_{Geri}$). H_1: There is a significant difference in hospitalization duration for pneumococcal disease between the age-risk groups identified (H_0: $\mu_{Pedi} \neq \mu_{HA} \neq \mu_{HRA} \neq \mu_{Geri}$). These hypotheses represent the following four risk groups: pediatric, healthy adult, high-risk adult, and geriatric. Healthy adult represents adults aged 18-64 years old with no risk factors indicating pneumococcal vaccination, while the high-risk adult group represents the same age group with vaccine-indicated comorbidities and risk factors.

Hospitalization Status

Hospitalization status was also assessed. This was defined as being admitted into a hospital facility. Emergency departments without admission were not considered as hospitalized. All 329 records identified with a hospitalization status. Of these cases, 241 (73.25%) patients were hospitalized, and 88 (26.75%) were not. Of the pediatric group, 47 patients (61.04%) were hospitalized, and 30 patients (38.96%) were not hospitalized. Of the healthy adult group, 71 patients (78.02%) were hospitalized, and 20 patients (21.98%) were not hospitalized. Of the high-risk adult group, 30 patients (75.00%) were hospitalized, and 10 patients (25.00%) were not hospitalized. Of the geriatric group, 93 patients (76.86%) were hospitalized, and 28 patients (23.14%) were not hospitalized.

A chi-square test of association was performed to compare hospitalization rates of these groups. With a Pearson's chi-square of 7.7846 and a p-value of 0.051, this association is not statistically significant, but it does present a strong trend for the relationship of hospitalization status along with the age-risk group.

Binary logistic regression was then performed to assess hospitalization status among high-risk versus generally healthy groups. This regression model resulted in an odds ratio of 2.1870, with a standard error of 0.2493 and a z-score of 4.62. Therefore, there is a significant increase in the high-risk group *($p < 0.005$)* to be hospitalized over the generally healthy group.

$$logistic(p_{hospitalized}) = \log\left(\frac{p_{hospitalized}}{1 - p_{hospitalized}}\right) = \beta_0 + \beta_1 X_{Risk}$$

A binary logistic regression model was also fitted to assess the association of age and sex with hospitalization status. In this model, the risk of hospitalization was not different between

males and females, with an odds ratio of 1.413 (p = 0.172). However, age showed a significant association with the risk of hospitalization, with an odds ratio of 1.011 (p = 0.008).

$$logistic(p_{hospitalized}) = \log\left(\frac{p_{hospitalized}}{1 - p_{hospitalized}}\right) = \beta_0 + \beta_1 X_{Age} + \beta_2 X_{Sex}$$

Hospitalization Duration

For hospitalization duration, 161 patients had both an admission and discharge date. Hospitalization duration ranged from 1 day to 84 days, with a mean of 8.025 days. Of those hospitalized, a total of 12 hospitals were identified. Hospitals were categorized into nine categories, including the eight local hospitals within Hidalgo County and a ninth category to represent all other facilities outside of the county. Figure 4 depicts a histogram of hospitalization duration of these nine categories.

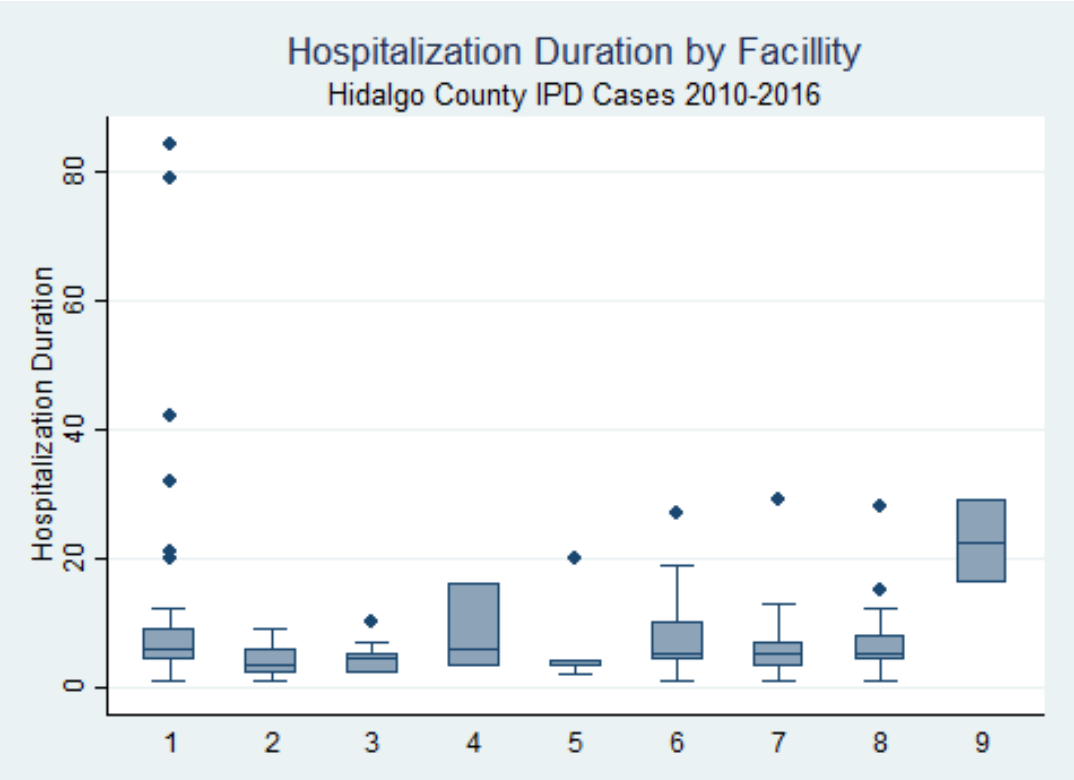


Figure 4: Hospitalization Duration by Facility. Ninth category represents other hospitals outside of Hidalgo County, Texas.

When examining hospitalization duration, the shortest average duration was Hospital 2 (n=8), with a mean of 4.125 days. The longest average duration was the Other category, Hospital 9 (n=2), with a mean of 22.5 days. The second longest average duration was Hospital 1 (n=52), with a mean of 10.5 days. Prior to ANOVA, a Shapiro-Wilk test was run in Stata to assess normality in the variable *HospitalDuration*. In this analysis, the data was identified to not have a normal distribution, with a z-score of 9.254 ($p < 0.00005$). Therefore, caution was taken when analyzing with one-way ANOVA and linear regression. A one-way ANOVA was performed to compare hospitalization duration lengths between each hospital. In this model, the ninth hospital category (Other) was removed, since this included all other hospitals outside of Hidalgo County, Texas. This ANOVA model resulted in an F statistic of 0.92, indicating no significant difference in means in the duration between any pair of hospitals ($p = 0.4958$). The Kruskal-Wallis rank test, a non-parametric test, was run to compare hospital stay lengths among the eight local hospitals in Hidalgo County. The chi-squared result was 7.687 with 7 degrees of freedom and a p-value of 0.3610. Therefore, the hypothesis that these two variables are independent failed to be rejected.

Now focusing on the patient characteristics, a linear regression model was developed to examine the association of age with hospitalization duration. Linear regression was fitted using hospitalization duration as the dependent variable and numerical age as the independent variable. With an F statistic of 0.99, this model indicated no significant association ($p = 0.3224$) between age and hospitalization duration. To account for a non-normal distribution, Spearman's rank correlation was analyzed for *AgeReported* and *HospitalizationDuration*. In this analysis, Spearman's rho equaled 0.0340, with p-value of 0.6681. Therefore, the null hypothesis that both variables are independent failed to be rejected.

$$Y_{HospitalizationDuration} = \beta_0 + \beta_1 X_{AgeReported} + \varepsilon$$

Data analysis was then executed to account for age-risk factors. As stated earlier, the population of study was divided into four groups. This included pediatric, healthy adult, high-risk adult, and geriatric. These four groups were divided to identify which groups, if any, had higher rates of infection and longer hospitalization duration compared to others. The pediatric group equaled 1, the healthy adult equaled 2, the high-risk adult equaled 3, and geriatric equaled 4. Group 1, pediatric, had 31 cases with a hospitalization duration. This ranged from 1 to 79 days, with a mean of 9.06 days. Group 2, healthy adult, had 43 cases with a hospitalization duration. This ranged from 1 to 27 days, with a mean of 7.14 days. Group 3, high-risk adult, had 22 cases with a hospitalization duration. This ranged from 1 to 84 days, with a mean of 11.68 days. Group 4, geriatric, had 65 cases with a hospitalization duration. This ranged from 1 to 29 days, with a mean of 6.88 days (see Figure 5).

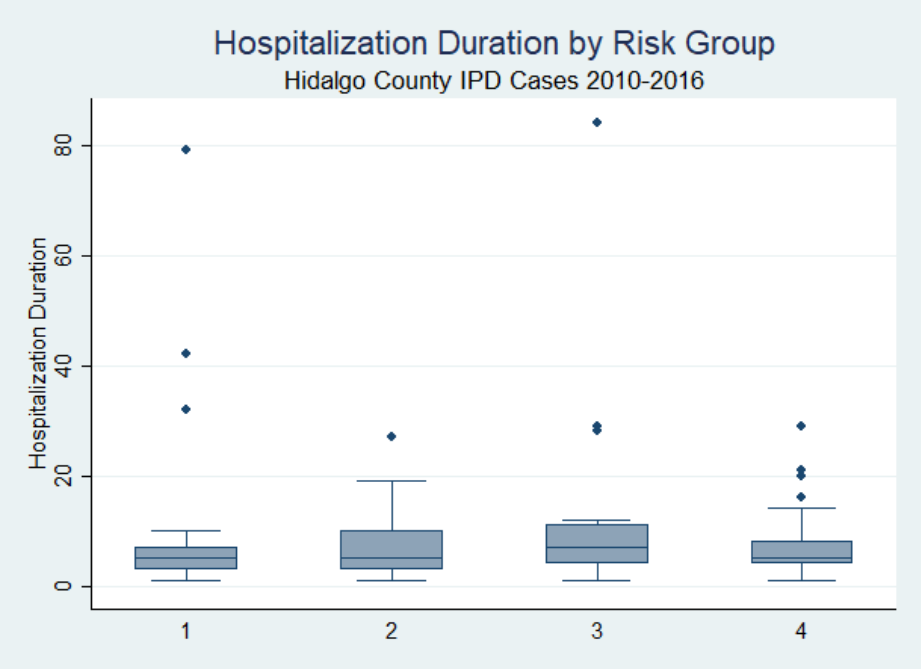


Figure 5: Hospitalization Duration by Age-Risk Group. Hospitalization in terms of days.

Next, a one-way ANOVA was completed to compare age-risk groups regarding hospitalization duration. With an F statistic of 1.01, the means of hospitalization duration between the four age-risk groups was not significant (p-value = 0.3894; see Table 5). In addition, a Kruskal-Wallis rank test was run, resulting in a chi-squared result of 2.667 with 3 degrees of freedom and a p-value of 0.4459. To further assess hospitalization duration, a one-way ANOVA model was run incorporating other explanatory variables (i.e., age category, sex, deceased outcome, Hispanic ethnicity, hospital, vaccination, vaccine eligibility, and underlying conditions). This was run multiple times for each variable of interest and compiled together. None of these variables presented significance within their prospective models (see Table 6).

Table 5: Analysis of Variance of Hospitalization Duration (Days) by Age-Risk Group

Source	Mean	Std. Dev.	Freq.
Group 1 (Pediatric)	9.0645161	15.513296	31
Group 2 (Healthy Adult)	6.9285714	6.2053983	42
Group 3 (High-Risk Adult)	10.857143	17.675649	21
Group 4 (Geriatric)	6.8769231	5.5297708	65
Total	7.8427673	10.475127	159

Source	SS	df	MS	F	Prob>F
Between groups	332.825687	3	110.941896	1.01	0.3894
Within groups	17004.2435	155	109.704797		
Total	17337.0692	158	109.728286		

Table 6: Analysis of Variance of Hospitalization Duration by Explanatory Variable

Source	Partial SS	df	MS	F	Prob>F
Age Category	146.798884	2	73.3994421	0.66	0.5207
Current Sex	27.6018976	1	27.6018976	0.6204	0.6204
Deceased	343.916138	1	343.916138	3012	0.0790
Hispanic Ethnicity	72.3576987	1	72.3576987	0.65	0.4223
Hospital ID	1130.57497	8	141.321872	1.29	0.2552
Vaccinated	50.9934727	1	50.9934727	0.46	0.5007
Vaccine Eligible	251.239752	1	251.239752	2.27	0.1338
Underlying Conditions	1240495027	1	124.495027	1012	0.2922

Case Fatality

The case fatality rate of each group was also of interest for review. A total of 22 (6.69%) cases of the 329 cases resulted in fatality. The other 307 cases (93.31%) had no documentation of mortality. Of the fatal outcomes, all 22 cases (100%) derived from bacteremia infections; no meningitis cases resulted in fatality. The case fatality rate for the pediatric group was 7.25%, for the adult group was 6.87%, and for the geriatric group was 6.61%. The mean age of cases resulting in mortality was 51.02 years old. The median age was 54.00 years old. The range from date of positive culture to date of death ranged from 0 days to 27 days. The mean of this time period was 3.45 days and the median was 1 day. A binary logistic model was run to assess mortality with age and sex, resulting in no significant association. Age presented an odds ratio of 1.003 *(p = 0.727)*, and sex presented an odds ratio of 0.912 *(p = 0.834)*.

$$logistic(p_{deceased}) = \log\left(\frac{p_{deceased}}{1 - p_{deceased}}\right) = \beta_0 + \beta_1 X_{Age} + \beta_2 X_{Sex}$$

Hypothesis 3 Results: Spatial Analysis of IPD Incidence

The third hypothesis in this study examined population-level data to compare other potential risk factors. This is more specifically defined in the following null and alternative hypotheses. H_0: Within the population of study, lower income populations have no increased risk in the rate of infection for invasive pneumococcal disease (H_0: $B_1 = 0$). H_1: Within the population of study, lower income populations have an increase risk in the rate of infection for invasive pneumococcal disease (H_1: $B_1 \neq 0$).

A total of 113 census tracts were identified in Hidalgo County, Texas, from the United States Census Bureau. Census tract names ranged from 201.01 to 246.00, with the exception of census tract 9800, which represents the McAllen Miller International Airport. No cases were reporting with this census tract. In addition, 27 zip codes and 22 cities were identified in Hidalgo County, Texas.

Population Data Results

United States Census data was collected using the American Community Survey (ACS) for the census tracts in Hidalgo County. The known information from the dataset included the case count per census tract. The total population for each census tract was extracted from the ACS in the American FactFinder of the U.S. Census Bureau. The average population of all census tracts in Hidalgo County was 7,396 people. Population of the 112 census tracts, excluding census tract 9800 with a population of zero, ranged from as low as 1,002 people in census tract 243.02 to as high as 17,165 people in census tract 240.00.

The average median household income for all census tracts was $36,182. The median household income of all census tracts in Hidalgo County was $32,819. This ranged from as low as $18,274 in tract 241.12 to as high as $76,971 in tract 235.10. Private health insurance

coverage averaged out to 33.65% for all census tracts, with a median of 29.50%. This ranged from as low as 10.40% in tract 241.12 to as high as 75.4% in tract 209.01.

Average household size and population density were of interest to determine if any relationships existed between an increase in person-to-person contact and disease incidence. The average household size of all census tracts was 3.59 individuals per household. Per census tract, this figure ranged from as low as 2.29 in tract 241.11 to as high as 5.03 in tract 205.1.
Next, the incidence was calculated for each census tract and multiplied by 10,000 to represent an incidence per 10,000 people. The incidence of cases ranged from as low as zero in multiple census tracts to as high as 17.57 per 10,000 people in census tract 204.03. Incidence was also measured for each of the 22 cities located in Hidalgo County, Texas. Of the 329 cases, three were removed due to not being located within a city. Of the 326 cases reviewed in their respective cities, the lowest incidence was zero cases. This included Granjeno, La Villa, Palmhurst, Progreso, and Progreso Lakes, Texas. The lowest calculative incidence was 1.74 cases per 10,000 people in Alton, Texas. The second highest incidence was 8.88 cases per 10,000 people in Mission, Texas, and the highest incidence was 17.83 per 10,000 people in Edcouch, Texas. This city also had an incidence 3.934 times greater than the average incidence for all cities. Each city's incidence was also calculated and compared (see Table 7).

Table 7: IPD Case Incidence Rates per City

City	Incidence per 10,000
Alamo	7.11418263
Alton	1.73631207
Donna	7.81343912
Edcouch	17.83060921
Edinburg	6.09215773
Elsa	4.20521447
Granjeno	0
Hidalgo	2.87129424
La Joya	4.66635558
La Villa	0
McAllen	4.5551382
Mercedes	1.79275726
Mission	8.88372975
Palmhurst	0
Palmview	5.17955801
Penitas	8.18665575
Pharr	5.03226943
Progreso	0
Progreso Lakes	0
San Juan	6.21940997
Sullivan City	4.81579581
Weslaco	2.72560583

Spatial Analysis Results

To perform spatial analysis for this study, ArcGIS (ArcMap) version 10.6 was used. First, a general choropleth was created to present the distribution of incidence across Hidalgo County. Next, a spatial autocorrelation (Moran's index) was completed to determine significance between dispersed, random, or clustering of incidence outcomes. Once run, with a z-score of 1.7943, this choropleth presented strong associations towards spatial clustering ($p = 0.0727$). Subsequent choropleths depict population-level explanatory variables, such as population density (per square

mile), median household income, private health insurance coverage, education, and ethnicity representation (see Figure 6). Next, high-low clustering (Getis-Ord general G) analysis was run to assess cluster patterns among case incidence. This analysis resulted in an observed G score of 0.000083. Therefore, a significant pattern was observed of a high-clustered pattern in regards to case incidence per census tract ($p = 0.0045$).

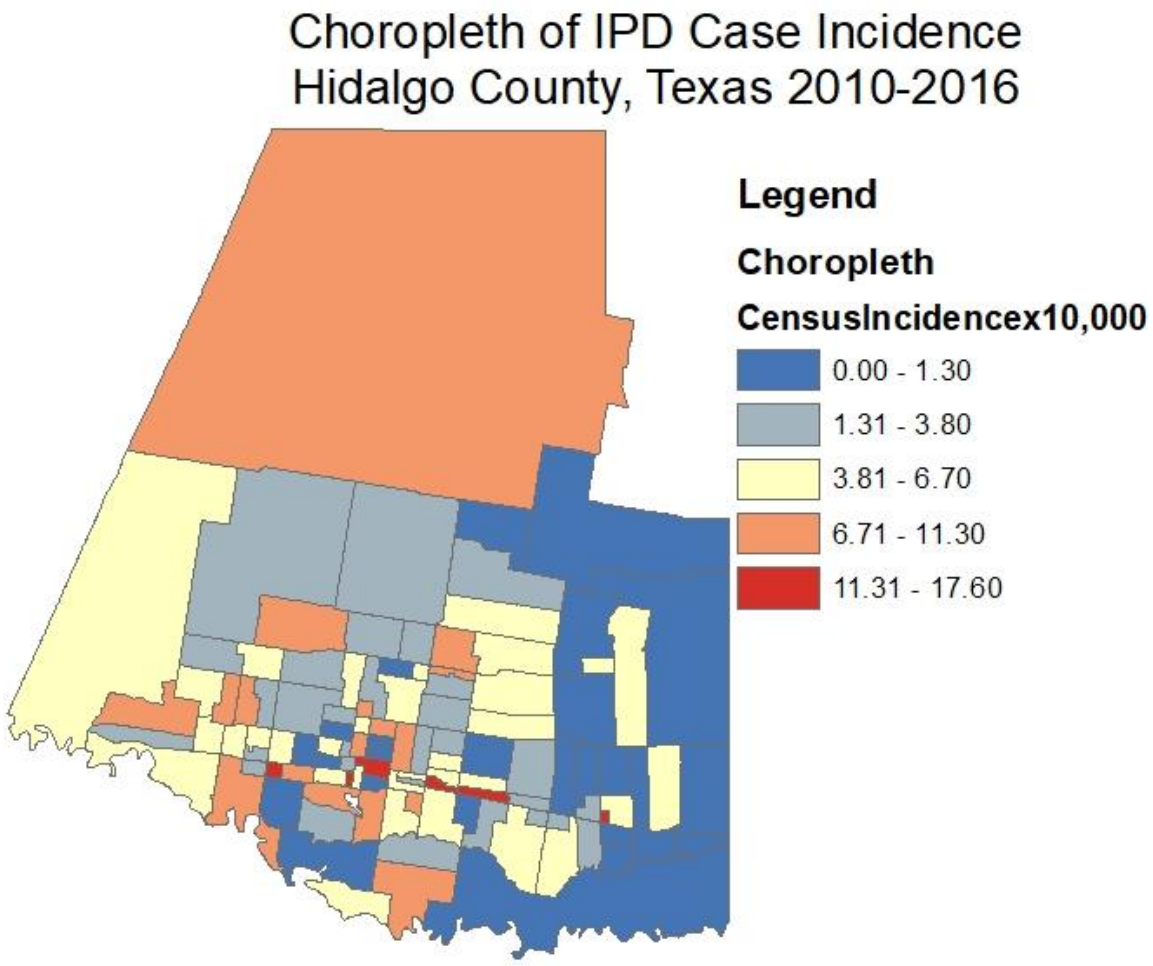


Figure 6: Choropleths of IPD Case Incidence Rates and Population Variable. Case incidence rates per 10,000 people. Continued on following page.

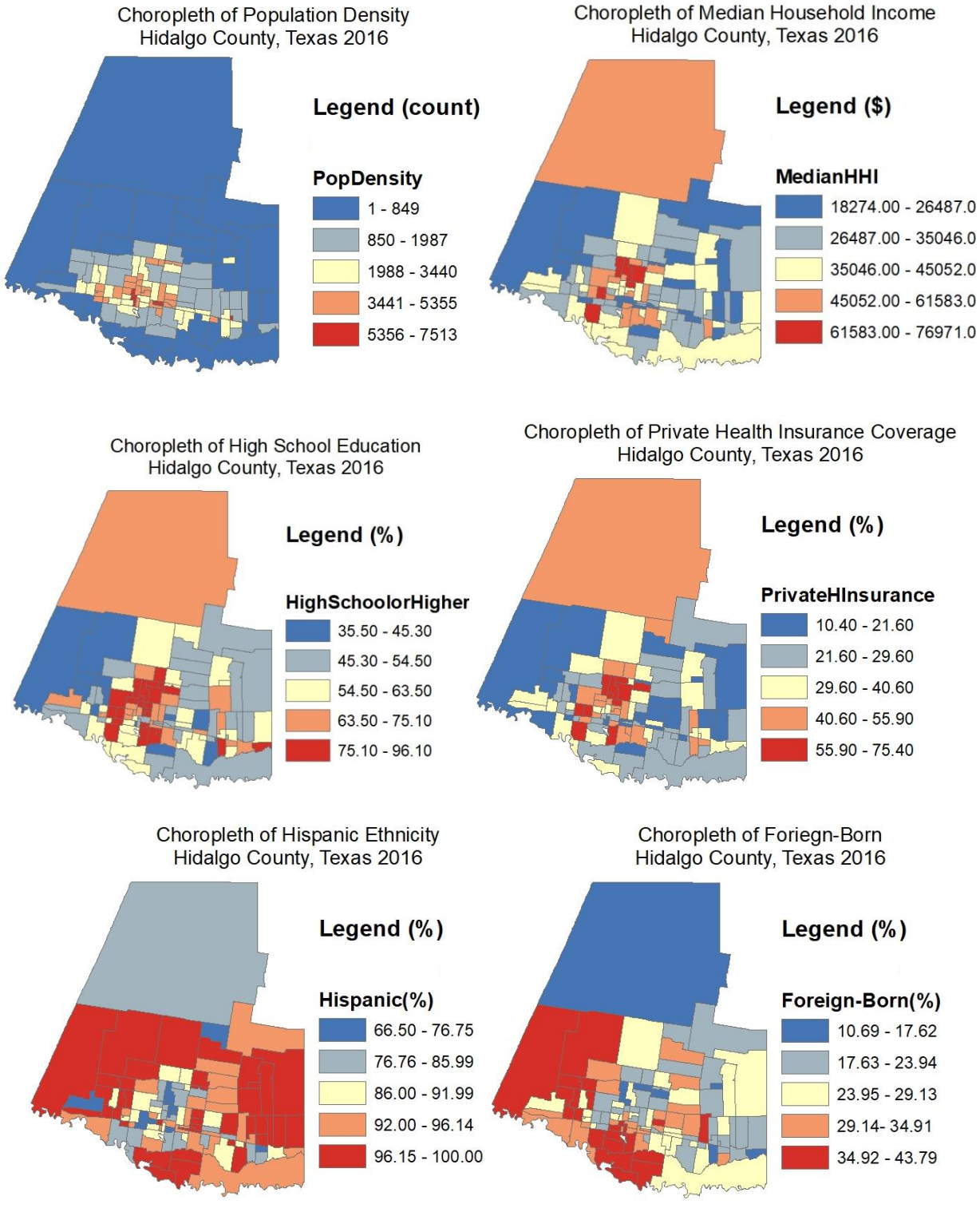


Figure 6 continued.

Next, a hot and cold spot analysis was performed, based on incidence rates within each census tract. Census tracts were statistically color coded to identify both hot and cold spots for incidence. The darker red areas describe hotspots and the darker blue areas represent coldspots. The yellow circle represents the median center for incidence. The green circle represents the mean center for incidence. Those shaded in dark red represent census tracts with a p-value less than 0.01 for high incidence coverage. The census tract shaded dark blue also represents a p-value less than 0.01 for low incidence coverage (see Figure 7).

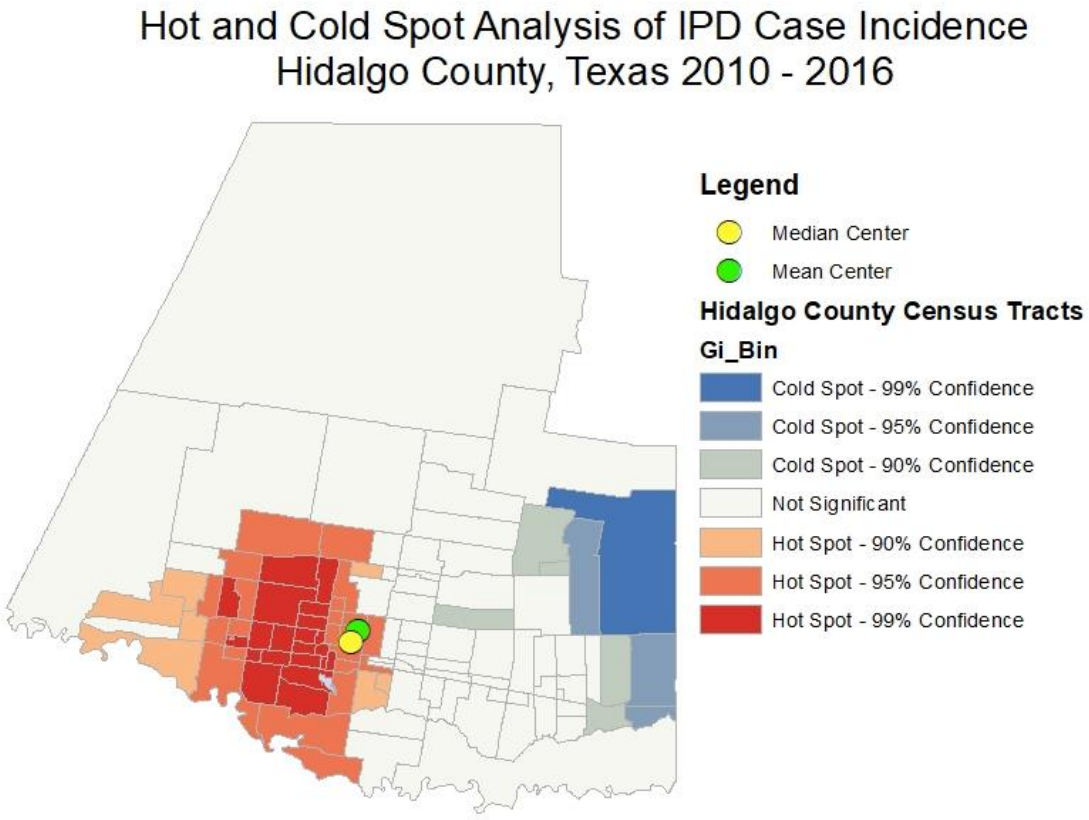


Figure 7: Hot and Cold Spot Analysis by Incidence Rates.

Spatial autocorrelation of variables (Moran's index) was also performed on the population-level variables and on case incidence. All explanatory variables presented with

significant results ($p < 0.00005$), presenting clustering based on these population attributes (see Table 8). Therefore, it can be interpreted that the independent variables of interest for this analysis present significant spatial association in regards to the clustering among the census tracts. Case incidence did not present significant clustering among census tracts ($p = 0.7276$).

Table 8: Spatial Autocorrelation of Variables

Variable	Moran's Index	Z-score	P-Value
MedianHHI	0.087268	6.594138	0.000000
AvgHHSize	0.055591	4.396873	0.000011
PrivateHInsurance	0.099732	7.414162	0.000000
PopDensity	0.260318	18.367830	0.000000
CensusIncidencex10,000	0.017161	1.794334	0.072760
HighSchoolorHigher(%)	0.141476	10.237786	0.000000
BachorHigher(%)	0.169649	12.222716	0.000000
Hispanic(%)	0.078939	6.006147	0.000000
Foreign-Born(%)	0.128167	9.330242	0.000000

Next, ordinary least squares (OLS) was conducted, using ArcGIS, to examine the four explanatory variables of interest and their overall relationship to case incidence in each of the census tracts. These four explanatory variables included median household income, private health insurance, household size, and population density. An OLS model, using the Getis-Ord (Gi*) statistic, was created incorporating all four explanatory variables to identify pattern detection. Once the model was run, the following was identified. Private health insurance had a

coefficient of 0.00329, resulting in no significant relationship (p = 0.4873) with incidence. Average household size had a coefficient of 0.042242, resulting in no significant relationship (p = 0.4756) with incidence. Median household income had a coefficient of -0.000010, depicting a strong trend relationship (p = 0.0651) with incidence. Finally, population density had a coefficient of 0.000056, resulting in a significant relationship (p = 0.0057) with incidence. Since population density depicted a significant association, an additional OLS spatial analysis was completed specifically for this variable (see Figure 8).

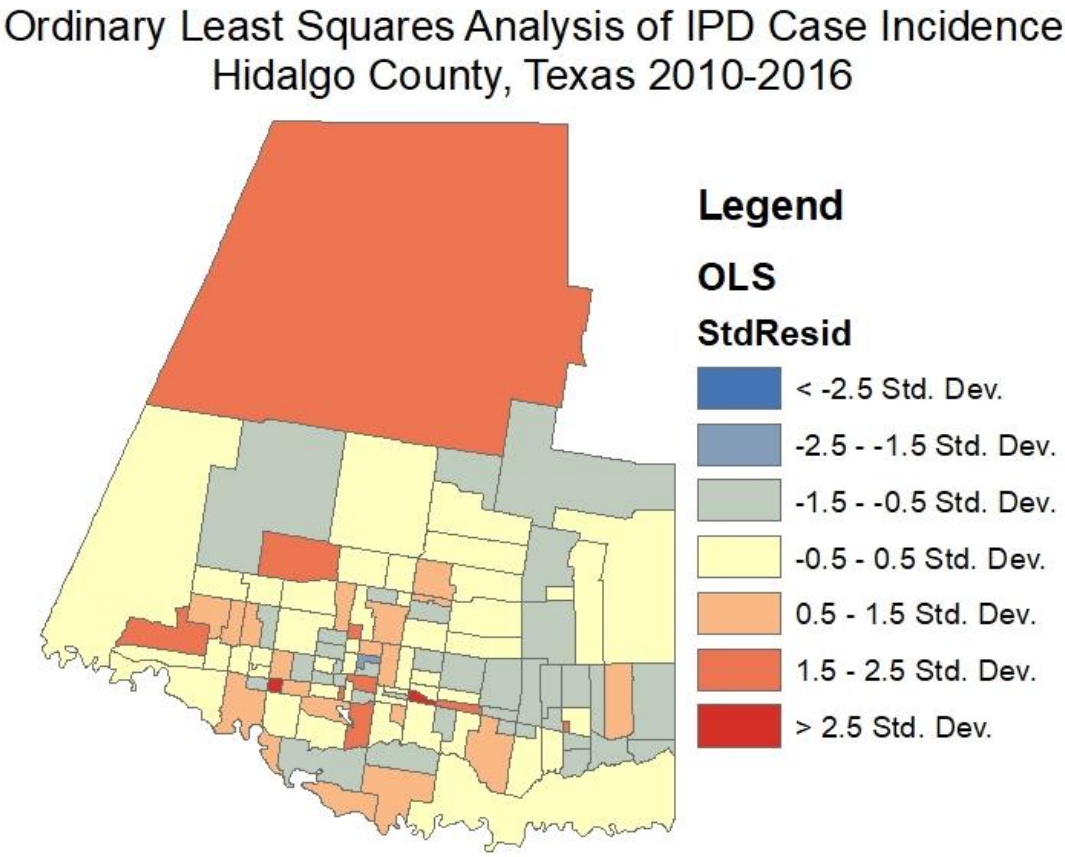


Figure 8: Ordinary Least Squares of Population Variables.

Spatial autocorrelation was conducted on the ordinary least squares model residuals above. In this technique, Morin's index was used, resulting in a value of -0.011547. With a p-value of 0.862269, no significant clustering was identified. Therefore, it can be concluded that

this was a result of random outcomes. Furthermore, this ordinary least squares model presented two independent variables of significance. Median household income ($p = 0.0398$) and population density ($p = 0.0129$) both noted a significant association with the invasive pneumococcal disease case incidence (see Table 9).

Table 9: Ordinary Least Squares Model Results

Variable	Coef.	Std. Error	t-Statistic	Prob.	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	8.3372	8.7940	0.9481	0.3453	9.4515	0.8821	0.3798	--------
Median HHI	-0.0001	0.0001	-1.9802	0.0503	0.0001	-2.0822	0.0398*	4.6685
AvgHH Size	0.4992	0.6204	0.8046	0.4229	0.6204	0.8046	0.4229	1.1475
PrivateH Insurance	0.0596	0.0702	0.8482	0.3983	0.0706	0.8430	0.4012	8.8431
Pop Density	0.0006	0.0002	2.9761	0.0036*	0.0002	2.5298	0.0129*	1.1686
Census Incidence x10,000	-0.0930	0.0739	-1.2593	0.2108	0.0732	-1.2704	0.2068	9.5097
High Schoolor Higher(%)	0.0866	0.0799	1.0835	0.2811	0.0843	1.0272	0.3067	7.6587
Bachor Higher(%)	-0.0075	0.0644	-0.1170	0.9071	0.0727	-0.1036	0.9177	2.3726
Hispanic (%)	-0.0050	0.0636	-0.0792	0.9370	0.0698	-0.0722	0.9426	1.8677

Geographically weighted regression was performed as an additional method for analysis. This form of analysis was selected to explore local relationships of the variables of interest. In

addition, this method was selected to account for the significant spatial clustering of the explanatory variables of interest for this analysis. In this model, the same four variables (median household income, private health insurance, household size, and population density) were included within the analysis. Figure 9 shows the categories of standard residuals for this model analysis. Furthermore, the local R^2 values were assessed for this geographically weighted regression model. Figure 9 was developed to depict the GWR model's fit, where areas with values closer to zero represent poorer performance of the local model.

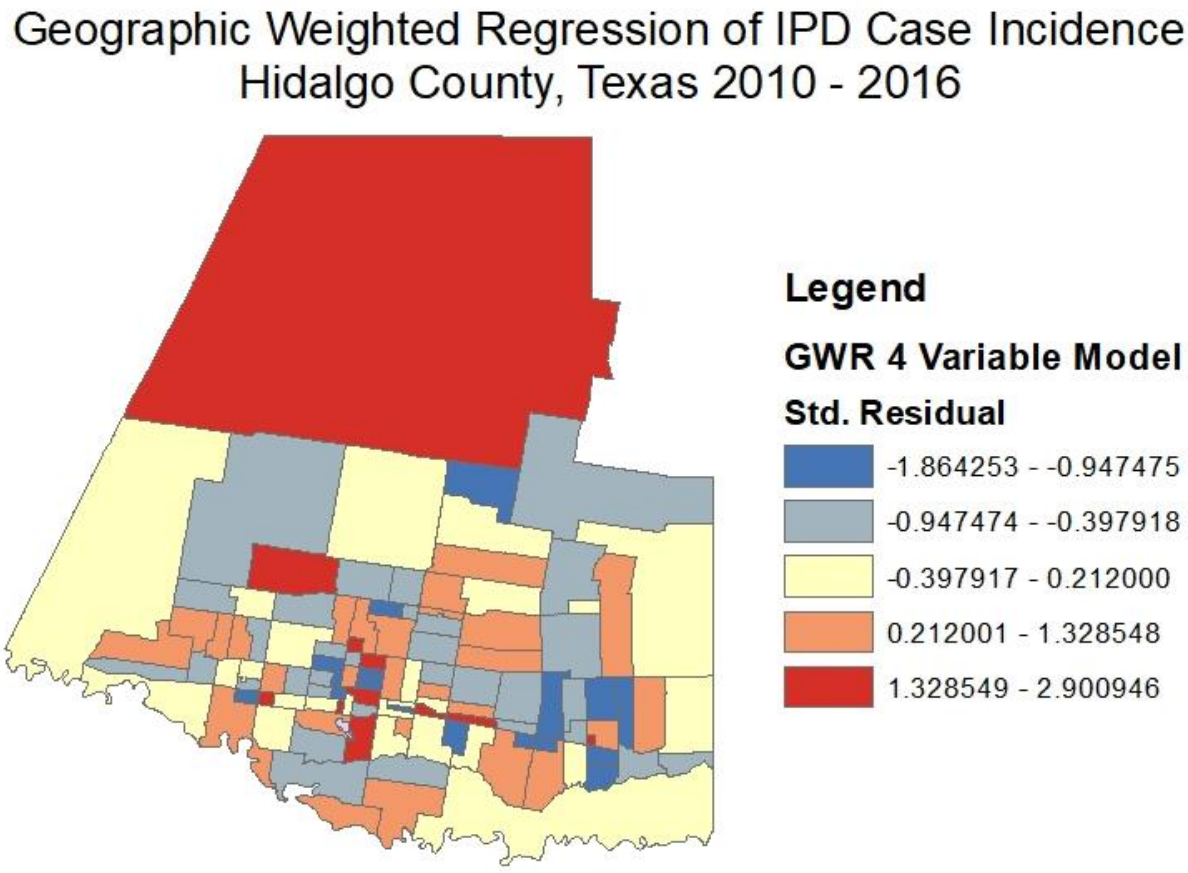


Figure 9: Geographic Weight Regression of Incidence Rates. Four GWR variables include median household income, private health insurance, household size and population density.

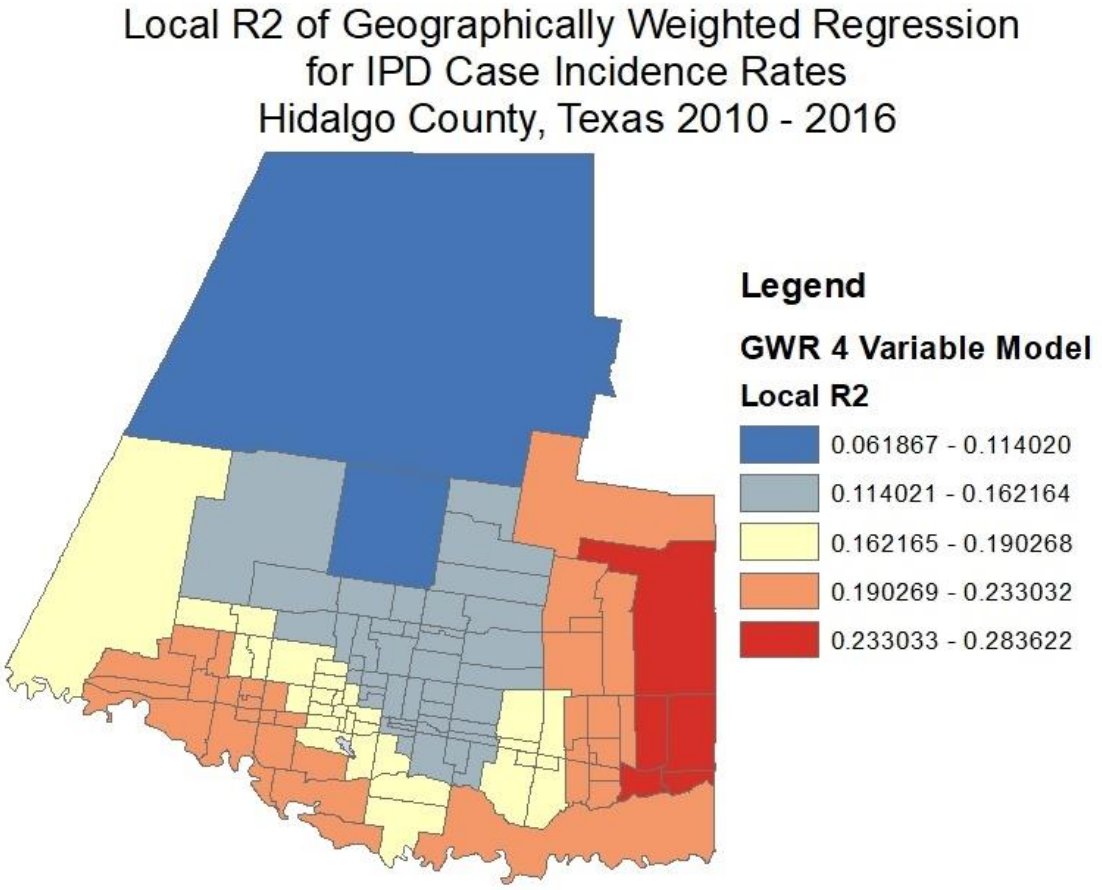


Figure 10: Local R^2 of Geographically Weight Regression of Incidence Rates

Finally, a linear regression model was conducted in Stata to compare to ArcGIS and to test the following null and alternative hypotheses: H_0: $B_1 = 0$ and H_1: $B_1 \neq 0$. This linear regression model examined the incidence of the 112 census tracts with median household income. As a four-explanatory-variable model, this resulted in a coefficient of 0.000056 and standard error of 0.0000199 for population density; thus reporting a significant relationship, with a p-value of 0.006. In addition, median household income showed a slight trend association (p = 0.065) as well, with a coefficient $-9.56e^{-6}$. Once this model was amended to a single-variable model, it resulted in a coefficient of $-6.26e^{-6}$ and standard error of $2.70e^{-6}$ for median household

income; thus reporting a significant p-value of 0.022 (see Table 10). The equations below represent this linear regression model of median household income as both a four-variable and single-variable model.

$$Y_{Incidence} = \beta_0 + \beta_1 X_{MedianHHI} + \beta_2 X_{AvgHHSize} + \beta_3 X_{PrivHealthIns} + \beta_4 X_{PopDens} + \varepsilon$$

$$Y_{Incidence} = \beta_1 X_{MedianHHI} + \varepsilon$$

Table 10: Linear Regression of Incidence by Income

Model	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]
MedianHHI	-0.0000626	0.000027	-2.32	0.022	[-.000116, -9.16e-06]
_cons	6.633215	1.037496	6.39	0.000	[4.577142, 8.689288]

Results Summary

In summary, a total of 333 case records were provided by Hidalgo County Health and Human Services. Of these, four cases represented a second infection later within the seven-year period. These four cases were removed from analysis. Therefore, a total of 329 observations (cases) were identified in the dataset. The center patient representation of all cases would be a 57.00 year old (median age) male (more common gender). Of these cases, 241 (73.25%) patients were hospitalized. Of the 329 cases of invasive infection, 96.05% (n=316) were bacteremia infections, 2.13% (n=7) were meningitis infections, and 1.82% (n=6) were either classified as other or unknown infections. Additionally, 22 cases (6.69%) resulted in mortality.

The first hypothesis examined case representation based on vaccine eligibility. First, data analysis assessed the adult and geriatric population to compare case differences. Although adult

cases were higher, there was no significant increase ($p = 0.2854$). Next, adults 18-64 years old were assessed based on vaccine eligibility. When comparing these two groups, there was a significant result ($p < 0.00005$) in greater case occurrence of the vaccine-eligible subgroup, compared to the vaccine-ineligible subgroup. The second hypothesis assessed hospitalization duration by comparing and analyzing each of the four age-risk groups. In this test, there was no statistically significant difference ($p = 0.3894$) in the means of hospitalization duration between the four age-risk groups. Finally, the third hypothesis was about population-level characteristics in invasive pneumococcal disease case incidence. This included variables representing socioeconomic status and demographics. Both median household income ($p = 0.022$) and population density ($p = 0.013$) presented significant associations with case incidence.

CHAPTER FIVE: DISCUSSION

To complete this research, a retrospective cross-sectional study was developed to assess the epidemiology of invasive pneumococcal disease in Hidalgo County, Texas. As a quantitative cross-sectional study, data was reviewed and analyzed from 2010 to 2016 of invasive pneumococcal disease (IPD). The aim of this research study was to identify the population characteristics at an increased risk for infection so that preventative health measures could be targeted and identified. This research was completed to answer the following questions:

- With vaccination recommendations established for pediatric, geriatric, and high-risk populations, is the healthy general adult population still at high risk for infection?
- Within this population of study, do high-risk patients experience a longer hospitalization duration and higher risk of mortality?
- Does socioeconomic status, such as household income, influence the rate of disease within this population?

To answer these questions, this research compared and identified characteristics and groups at higher risk for infection. In addition, with vaccine recommendations specified towards certain groups, this research examined the case occurrence and disease outcomes of the different age-risk groups identified. Finally, this research examined population-level data to conduct spatial analysis to identify potential trends and patterns of increased incidence.

Findings and Interpretations

The general descriptive statistics within this study identified the salient characteristics of disease cases. With this descriptive information, it was identified that male infections were more

common and that the median age was usually in the middle to late fifties each year. Although case representation can vary from each patient, these two demographics provide a center point of an invasive pneumococcal disease patient in Hidalgo County, Texas. With the exception of MMWR Year 2010, annual case counts stayed consistently in the 40s to 50s. However, the three most recent years have shown an increasing trend line. This increasing case trend line is of interest since the introduction of a vaccine should be decreasing case rates. In addition, this slight increase conflicts with the literature, where many areas have shown a decrease in annual cases. A Rhode Island study presented with a case trend line that consistently dropped annually from 2011 (94 cases) to 2015 (62 cases); (Gosciminski et al., 2017, p. 58). In a study at a Taiwan university hospital, the incidence of IPD was 3.8 per 10,000 in 2010, decreased to an incidence of 3.1 in 2011, and then to 2.1 in 2012 (Chih-Cheng et al., 2014, p. 6). However, a hospital study in Guatemala presented with a stable, non-decreasing trend line, similar to Hidalgo County. With both the hospital's Santa Rosa and Quetzaltenango departments combined, the case count for pneumococcal infections changed from 32 cases in 2009, to 27 cases in 2010, to 32 cases in 2011, to 22 cases in 2012 (Contreras et al., 2015, p. 9).

Mortality is another notable feature presented in these data. The literature review identified that the case fatality rate is estimated to be around 20%, and up to 60% in the elderly, at the United States national level (CDC, 2015a, p. 280). In the university hospital study in Taiwan, the overall case fatality rate (CFR) was 30.9%, and the 65 years and older CFR was 40.7% (Song et al., 2013, p. 4). However, the range of mortality was 4.65% to 10.17% during this study's time period. Furthermore, the overall case fatality rate for the 65 years and older population was 6.61% in this study. Therefore, it is of interest to explore what marks these

notable differences and if long-term tracking is needed to better identify final outcomes in the local population of south Texas.

Hypothesis 1 Discussion: Vaccine Eligibility in Cases

The age of onset for infection was a variable of interest in this research study. One reason for this is because the 18-64 age group is the only age group where there is no vaccine indication on the sole basis of age. In addition, the median age for Hidalgo County (28.8 years) is nearly nine years younger than the national median age (37.7 years; U.S. Census Bureau, 2017). The first hypothesis identified that there was no significant difference between the binomial probability of infection between the adult and geriatric populations ($p = 0.2854$). With these findings, it became notable to further explore the subgroups of these two populations. The adult age group (18-64 years old) represented the largest age group (n=131) of cases of the three age groups. This age group representation accounted for 39.8% of all cases.

As compared to the literature, a study in Sweden contrasts this finding, where the geriatric age group (65 and older) represented that largest age group with 51.1% (n=1,521; Backhaus et al., 2016, pp. 5-7). However, other studies displayed similar descriptive findings to those within this research study. In a Guatemalan adult pneumococcal study, the majority of cases, 62.2% (n=117), represented adults 18-64 years, as compared to geriatric patients (Contreras et al., 2015, p. 6). In addition, similar results were identified in a South Korean study, where adults comprised 55.2% (n=535), as compared to geriatrics, 44.8% (n=435; Song et al., 2013, p. 4). These initial findings in this south Texas research study showed concern for this age group, since this is the only group where age alone is not a vaccine indication. Once other vaccine indicators, such as comorbidities and risk factors, were accounted for, 38 cases had at least one vaccine indication, while 89 cases had no vaccine indication. Once vaccine eligibility

was accounted for, there was a significant difference ($p < 0.00005$) between the healthy and high-risk adult groups. Therefore, it was identified that the vaccine-ineligible subgroup had a significant case representation over the vaccine-eligible subgroup. However, when conducting binomial testing between the overall vaccine-eligible and -ineligible groups, the vaccine-eligible group had a statistically significant ($p<0.00005$) higher proportion of cases (n=230) over the vaccine-ineligible group (n=91). This resulted in failing to reject the null hypothesis that infections would be similar among both groups. However, this binomial testing does support the notion that within the adult population, infections were more commonly occurring in those with no documented risk factors, as compared to the high-risk group. This presented two notable outcomes. First, that the generally healthy adult group has an increased risk for infection. Second, it also identified that lack of vaccination in this high risk-group is a public health concern in regards to disease prevention.

The vaccination coverage among cases identified that a significant majority were unvaccinated, as identified through the chi-square association test. This helps identify introductory risk assessment of vaccine efficacy in this border region. With serotypes varying globally, concern was presented that those vaccinated were still at great risk for infection. The highest proportion of unvaccinated persons occurred in the geriatric age group. In addition, as expected, the findings found vaccinated individuals have lower overall infection case counts.

In a vaccination coverage study in the United States, data from the National Health and Wellness Study was analyzed. Among those vaccinated for pneumococcal disease, more were female (60.7%), as compared to male (39.3%; Annunziata et al., 2012, p. 5). Of the 38 vaccinated cases of infection in this research study, more females (n=21) were vaccinated as compared to males (n=17).

In addition to age, comorbidity presence was another variable of interest for this study. The presence of an underlying condition did not significantly increase the risk of case fatality, resulting in an odds ratio of 1.048 ($p = 0.918$). However, there was a significant increase in hospitalization for having underlying condition(s), resulting in an odds ratio of 2.19 ($p = 0.02$).

Hypothesis 2 Discussion: Hospitalization Duration in Risk Groups

Utilizing hospital data provided varied opportunities for data analysis of the cases, their characteristics, and their outcomes. This included incorporating explanatory variables such as demographics and clinical background to dependent variables like hospitalization status and hospitalization duration. Explanatory variables, such as age, sex, and underlying conditions, assessed typical characteristics that could depict increased risk factors for infection.

The second hypothesis examined hospitalization status and hospitalization duration. Of the 329 patients, 241 (73.25%) had record of hospital admission. A chi-square was performed for hospitalization status and age-risk groups. A strong trend in the relationship of these two variables occurred but was not statistically significant ($p = 0.051$). Therefore, the higher risk and older populations were more likely to be hospitalized. Next, ANOVA was conducted to compare the association of hospital facility and hospitalization duration. However, the data was identified to not be normally distributed through a Shapiro-Wilk test. Therefore, caution was taken in the interpretation of the ANOVA results, and other non-parametric analysis techniques were utilized to account for this. However, there was no significant relationship when hospital facility and hospitalization duration were analyzed through ANOVA ($p = 0.4958$). The Kruskal-Wallis rank test depicted no significance as well ($p = 0.3610$). Therefore, the hypothesis that hospitalization duration and hospital facility are independent failed to be rejected.

The current literature is limited on identifying hospitalization status, since the majority focus on a specific hospital or medical system. However, overall, this study's hospitalization rate is lower than those presented in the literature. In a Rhode Island study, 93.04% (n=361) patients were hospitalized with invasive pneumococcal disease (Gosciminski et al., 2017, p. 58). In a Sweden study, of the 2,977 episodes of disease, only 25 cases (0.84%) presented as outpatient (Backhaus et al., 2016, p. 5).

Hypothesis 3 Discussion: Spatial Analysis of IPD Incidence

Spatial analysis of the epidemiology of invasive pneumococcal disease provided the opportunity to explore and analyze population-based data. This supported the hypothesis examining socioeconomic status within the population. This included incorporating census data and utilizing census tracts to better understand population characteristics.

The third hypothesis identified population level-risk factors. To begin, case incidence was examined at both the city and census tract levels. At the city level, Edcouch, Texas, had the highest incidence rate at 17.83 cases per 10,000 people. However, when examining this geographical area at the census tract level, the case incidence rate did not rank among the highest. Therefore, examining both population levels assisted in giving additional insight on case occurrence. Furthermore, looking at US Census population level data assisted with incorporating socioeconomic factors for analysis. When studying human transmission, individuals per household and population density were examined. As these variables were analyzed, it became apparent that median household income did have a significant association with incidence. With a p-value of 0.022 in the linear regression model of case incidence and median household income, the correlation of these two variables was significant, thus identifying that income has a strong association with risk of infection within this specific population.

Patient-specific information was not available on income, household size, or other instances. Therefore, census tracts were used as a representation for these cases. Therefore, exact information and analysis directly related to the patient were unable to be performed. This means that a patient may possibly not align with his or her census tract characteristics for these variables. However, these data variables do bring insight on notable characteristics that create a foundation for future research in regards to these relationships. When examining these explanatory variables through spatial autocorrelation (Moran's index), a foundation was established that variables of interest had significant spatial clustering. This included median household income, average household size, private health insurance coverage, population density, high school education, bachelor's degree education, Hispanic ethnicity, and foreign-born representation ($p < 0.00005$). Therefore, with each of these variables significantly clustered, additional analysis could then examine the effect of case incidence.

The hot and cold spot analysis identified the areas with the most concentrated incidence. This also matched population density. However, since incidence, rather than case count, was analyzed, the result was not due to the increased population. Once identified, ordinary least squares (OLS) and geographically weighted regression (GWR) were performed to test the explanatory variables. Within the ordinary least squares analysis, only population density ($p = 0.0129$) and median household income ($p = 0.0398$) had significant findings. Therefore, these two variables presented significant association with IPD case incidence, resulting in rejecting the null hypothesis, $H_0: B_1 = 0$.

In the spatial analysis of this outcome in OLS, the specific census tracts and their residuals were visualized through mapping. Therefore, areas of highest residuals could be identified for future research and analysis. Next, GWR was run to assess a four-variable model

(i.e., median household income, average household size, private health insurance coverage, and population density). Performing this GWR model was based on the significant findings from the OLS analysis so that GWR could attempt to correlate geographical and spatial influences into the regression analysis. In this analysis under GWR, similar results occurred where population density presented significant findings ($p = 0.006$) and median household income presented a slight trend, not significant, for association ($p = 0.065$).

Finally, although there is current literature available on the epidemiology of IPD, it is limited in regards to spatial analysis. Therefore, techniques and methods were of focus for general infectious disease analysis. In a Massachusetts study, hotspot analysis was performed for hepatitis C. In this study, Hispanic ethnicity and food stamps (income level) showed significant spatial hotspots, and a high school education or greater showed a negative association for infection (Stopka et al., 2017, pp. 4-6).

Outliers in the Data

Throughout the research study, outliers were identified within the data analysis. Examples of this included the hospitalization duration length in days, where some patients stayed admitted in the hospital greater than 50 days. The average hospitalization duration for 2010 was much lower than other years. The sample size for hospitalized was also much smaller, bringing to question whether data completeness was reliable for this year. Another example included assessing the cities' disease incidence, where Edcouch, Texas, has a much greater incidence as compared to all other cities in Hidalgo County. One of the census tracts did not appear to align with the others. This was census tract 9800, as compared to all others in the 200's. There was no data available from the United States Census Bureau. After researching this census tract, it was

identified that this census tract is the McAllen International Airport. This census tract was not included within the spatial analytic methods of this study.

Hospitalization status was recorded for each case within the data set. As identified earlier, the overall hospitalization rate in this study (73.25%) was much lower than those identified in the literature. However, 81 of the 88 (92.05%) cases identified as not hospitalized were collected or reported from a hospital. Therefore, further review by the local health department could help identify if these patients were ever hospitalized later during their medical visit (e.g., admitted into the hospital from an initial emergency room visit).

Limitations

Although there was a large representation of hospitalization duration data on cases, many cases were still missing hospital discharge data. With discharge data missing, a full analysis of the cases and their outcomes could not be completed. This is especially true since it was documented in multiple cases that the discharge date was missing due to the patient still being in the hospital at the time of closing the case investigation. Therefore, it is likely data skewed towards longer hospitalization durations are more likely to be missing, rather than short hospitalization duration stays. This creates a concern for data bias and misrepresentation.

The epidemiology case criteria guide for Texas DSHS is reviewed each year. Updates can include and incorporate new lab technologies, such as CSF PCR. This can then result in an increase in case identification, now allowing for a probable case definition. This new case status addition did not take place until MMWR Year 2017. Therefore, no case increases or outcomes were affected by this amendment. However, as physician may have begun ordering these PCR labs in the later portion of this timeline, these were positive labs that were never counted as a

case. This is because they did not meet case definition. Therefore, this study may have not fully represented PCR cases and instead may only represent culture-based cases.

Limitations also occurred within the spatial analysis component. As an overall foundation, current research on spatial epidemiology is limited, especially for specific diseases like invasive pneumococcal disease. Therefore, the literature was abstracted as a foundation in regards to methods applied, not necessarily the specific disease of analysis. Socioeconomic and demographic data points such as household size, population density, median household income, and private health insurance coverage provided introductory information based on the case's census tract. However, this did not provide direct correlation to the case itself. This created the concern of misrepresentation of data for such cases. Furthermore, this could be especially true with census tracts with few to even single case counts. These census tracts that had minimal case distribution may also have limited statistical strength in the conducted tests for data analysis. In addition, some data did not have a normal distribution, as tested through goodness-of-fit. This resulted in conducting specific forms of statistical analysis, like ANOVA, F-test, and Bartlett's test, to account for the distribution patterns.

Recommendations

With overall findings depicting infections in vaccine-indicated populations, increased vaccination efforts and education are vital to protect high-risk populations. This especially includes those aged 18-64 years old with at least one risk factor. Once risk factors were assessed within this age group, it was identified that a strong representation of cases were eligible for vaccination but had no record of prior history of vaccination. Therefore, targeting vaccination efforts in this specific subpopulation can help prevent future disease.

It is recommended that census tract data entry be included within routine investigations for notifiable disease cases in Texas. Entering this information into the National Electronic Disease Surveillance System (NEDSS) at time of entry can assist with both outbreak and spatial analysis. This will prevent epidemiologists and researchers from having to perform post-event geocoding. Therefore, this could assist in outbreak analysis turnaround time and support analysis methods staying compliant with HIPAA and other data security protocols.

Another recommendation is to enter discharge dates of notifiable condition investigations, including invasive pneumococcal disease, that are closed prior to discharge. One concern of data validity was the discharge dates for cases where the patient was still admitted at the time of the case closing. This would also assist with case fatality identification. If a patient is admitted long term, his or her survival outcome may change, affecting the overall case fatality rate.

It is recommended that lack of vaccine history be documented. Many vaccine history entries were unknown or blank. Therefore, deciphering between a true no and unknown vaccine history was difficult. Performing this step can help in determining vaccine efficacy and coverage in both this disease as well as with other conditions and outbreaks. Furthermore, research such as this presents the limitation of missing vaccine information. Currently, Texas is an opt-in state for the state immunization registry, known as ImmTrac. This means, that once a patient turns 19, if he or she does not opt-in to the system, by signing consent, that vaccine record is removed from the system. Therefore, tracking adult vaccination history becomes difficult, creating the need for policy development for the state system to transition towards an opt-out system.

Finally, the last recommendation is to increase funding opportunities for epidemiological investigation in public health. With many health departments facing limited staff and resources,

additional funding can assist in epidemiology and surveillance staff completing investigations fully and accurately.

Researcher Reflections

Prior to this study, it was preconceived through anecdotal information that the general healthy adult population had a greater risk for infection as compared to other subgroups. This was of interest, especially since these other subgroups had a vaccine available, creating an opportunity for decreased disease incidence. Once the study was enacted and data analysis took place, it became evident that the adult population (18-64 years old) was indeed the more prominent population for case representation. However, as further analysis took place and this group was subdivided between healthy and high risk, it was identified that it was the high-risk group that was dominant. Therefore, these research findings helped identify that although the general healthy adult population was not at as increased risk as suspected, the vaccine-indicated adult group was. Furthermore, this better identifies which subgroups need to be targeted for vaccine education and public health intervention.

Suggestions for Future Research

The first suggestion for future research is to study more patient specific-variables, such as socioeconomic variables. As mentioned earlier, the socioeconomic data was collected from the United States Census Bureau and represented the case's census tract. This is because the type of data was not collected at the patient level. Future research collecting new data could collect this same information for each specific patient. By doing so, more accurate representation of case characteristics can become available.

The second suggestion for future research is to begin incorporating serotype research collaboration to identify if the border population is at risk for infection not covered by the US

vaccine. With this population of study residing along the border, it is of interest whether the serotypes of infection are aligned within the current two pneumococcal vaccines licensed in the United States. As noted in the literature review, serotypes can vary in their global distribution. Therefore, a border population, such as the one in this study, is not only exposed to binational transit but also to a global population exposure through economic development and transportation.

The third suggestion for future research is to broaden the span of the border region population of study. Future research incorporating a broader span of the border region can be beneficial in extracting additional information and details of risks for infection. Options of this can include studying the whole border region from El Paso, Texas, to Brownsville, Texas. Including the whole Texas border region will incorporate other forms of populations, such as rural areas. Doing comparative studies of all border regions of the four American states and their counterpart Mexican border states can also assist in identifying common trends and discrepancies of the populations under study. Understanding the types of infection along these areas can greatly assist in prevention of cases throughout the United States. As populations become more transient and global, international characteristics of diseases are more likely to become embedded into local populations.

The fourth suggestion for future research would be to conduct a pneumococcal vaccine acceptance study among Texas geriatric patients along the border. With a very low vaccine rate identified in this study additional research can help identify whether these lower vaccination rates are due to lack of documentation, vaccine hesitancy, limited patient education on vaccine eligibility, or limited vaccine offerings by medical providers.

Finally, the fifth suggestion in future research is to incorporate other geocoding techniques beyond census tracts. Utilizing geopoint data could help analyze further forms of outcomes and assist with other information like hospital proximity and other distance-based analysis. This includes examining and expanding to environmental factors, like access to care (i.e., hospitals, clinics, and pharmacies).

Summary and Conclusion

A total of 333 cases were received in the dataset from Hidalgo County Health and Human Services. Extensive data cleanup took place to categorize variables, input missing information, and transcribe string variables into numerical or categorical values. Four duplicated patients were identified, and their second infections were removed from the dataset, resulting in a total of 329 cases for analysis. The hypotheses in this study examined risk subgroups, hospitalization and outcomes, and population-level characteristics for infection. Overall, the findings identified the groups at increased risk for infection, groups with lower vaccination coverage, hospitalization outcomes, and spatial findings for invasive pneumococcal disease. At the population level, median household income and population density are two variables of interest, where future research can examine these potential risk factors further. With these findings, the need for research does not end here. Future areas for research has been established, targeting concerns and needs to help serve the population and support the prevention of disease.

REFERENCES

Adler, N. E., & Stewart, J. (2010). Health disparities across the lifespan: Meaning, methods, and mechanisms. *Annals of the New York Academy of Sciences, 1186*(1), 5-23. doi:10.1111/j.1749-6632.2009.05337.x

Annunziata, K., Rak, A., Del Buono, H., DiBonaventura, M., & Krishnarajah, G. (2012). Vaccination rates among the general adult population and high-risk groups in the United States. *PLoS ONE, 7*(11), e50553. doi:10.1371/journal.pone.0050553

Askim, Å., Mehl, A., Paulsen, J., DeWan, A. T., Vestrheim, D. F., Åsvold, B. O., . . . Damås, J. K. (2016). Epidemiology and outcome of sepsis in adult patients with Streptococcus pneumoniae infection in a Norwegian county 1993-2011: an observational study. *BMC Infectious Diseases, 16*, 1-9. doi:10.1186/s12879-016-1553-8

Azzari, C., Moriondo, M., Di Pietro, P., Di Bari, C., Resti, M., Mannelli, F., . . . Marchetti, F. (2015). The burden of bacteremia and invasive diseases in children aged less than five years with fever in Italy. *Italian Journal of Pediatrics, 14*, 1-8. doi:10.1186/s13052-015-0189-4

Backhaus, E., Berg, S., Andersson, R., Ockborn, G., Malmström, P., Dahl, M., . . . Trollfors, B. (2016). Epidemiology of invasive pneumococcal infections: manifestations, incidence and case fatality rate correlated to age, gender and risk factors. *BMC Infectious Diseases, 16*, 1-12. doi:10.1186/s12879-016-1648-2

Black, C. L., Williams, W. W., Warnock, R., Pilishvili, T., Kim, D., & Kelman, J. A. (2017). Pneumococcal Vaccination Among Medicare Beneficiaries Occurring After the Advisory Committee on Immunization Practices Recommendation for Routine Use Of 13-Valent Pneumococcal Conjugate Vaccine and 23-Valent Pneumococcal Polysaccharide Vaccine for Adults Aged ⩾65 Years. *MMWR: Morbidity & Mortality Weekly Report, 66*(27), 728-733. doi:10.15585/mmwr.mm6627a4

Caldwell, R., Roberts, C. S., An, Z., Chieh, I. C., & Bruce, W. (2015). The health and economic impact of vaccination with 7-valent pneumococcal vaccine (PCV7) during an annual influenza epidemic and influenza pandemic in China. *BMC Infectious Diseases, 15*(1), 1-11. doi:10.1186/s12879-015-1021-x

Center for Surveillance, E., and Laboratory Services (CSELS). (2017, March 26, 2017). Integrated Surveillance Information Systems/NEDSS. *National Notifiable Diseases Surveillance System (NNDSS).* Retrieved from https://wwwn.cdc.gov/nndss/nedss.html

Centers for Disease Control and Prevention. (2012). Lesson 1: Introduction to Epidemiology. *Principles of Epidemiology in Public Health Practice, Third Edition An Introduction to Applied Epidemiology and Biostatistics.* Retrieved from https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section8.html

Centers for Disease Control and Prevention. (2015a). Pneumococcal Disease. In A. K. Jennifer Hamborsky, Charles Wolfe (Ed.), Epidemiology and Prevention of Vaccine-Preventable Diseases (13th ed., pp. 279-296). Washington D.C.: Public Health Foundation. Retrieved from http://www.cdc.gov/vaccines/pubs/pinkbook/downloads/pneumo.pdf.

Centers for Disease Control and Prevention. (2015b). Principles of Vaccination. In A. K. Jennifer Hamborsky, Charles Wolfe (Ed.), Epidemiology and Prevention of Vaccine-Preventable Diseases. Washington D.C.: Public Health Foundation. Retrieved from http://www.cdc.gov/vaccines/pubs/pinkbook/downloads/prinvac.pdf.

Chih-Cheng, L., Sheng-Hsiang, L., Chun-Hsing, L., Wang-Huei, S., & Po-Ren, H. (2014). Decline in the incidence of invasive pneumococcal disease at a medical center in Taiwan, 2000–2012. *BMC Infectious Diseases, 14*(1), 1-15. doi:10.1186/1471-2334-14-76

Contreras, C. L., Verani, J. R., Lopez, M. R., Paredes, A., Bernart, C., Moscoso, F., . . . McCracken, J. P. (2015). Incidence of Hospitalized Pneumococcal Pneumonia among Adults in Guatemala, 2008-2012. *PLoS ONE, 10*(10), 1-14. doi:10.1371/journal.pone.0140939

Dekker, L. H., Rijnks, R. H., Strijker, D., & Navis, G. J. (2017). A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands - the Lifelines cohort study. *International Journal of Behavioral Nutrition & Physical Activity, 14*, 1-15. doi:10.1186/s12966-017-0622-8

Fay, E. E., Hoppe, K. K., Schulkin, J., & Eckert, L. O. (2016). Survey of Obstetrics and Gynecology Residents Regarding Pneumococcal Vaccination in Pregnancy: Education, Knowledge, and Barriers to Vaccination. *Infectious Diseases in Obstetrics & Gynecology*, 1-6. doi:10.1155/2016/1752379

Ferrer, H. B., Trotter, C., Hickman, M., & Audrey, S. (2014). Barriers and facilitators to HPV vaccination of young women in high-income countries: a qualitative systematic review and evidence synthesis. *BMC Public Health, 14*(1), 700-721. doi:10.1186/1471-2458-14-700

Gómez-Barreto, D., Espinosa-Monteros, L. E., López-Enríquez, C., Jiménez-Rojas, V., & Rodríguez-Suárez, R. (2010). Invasive pneumococcal disease in a third level pediatric hospital in Mexico City: epidemiology and mortality risk factors. *Salud Pública de México, 52*(5), 391-397.

Gosciminski, M., Bandy, U., & Luther, K. (2017). The Changing Epidemiology of Invasive Pneumococcal Disease after the Introduction of Pneumococcal Conjugate Vaccine, Rhode Island, 1997-2016. *Rhode Island Medical Journal, 100*(1), 57-59.

Hendrix, K. S., Sturm, L. A., Zimet, G. D., & Meslin, E. M. (2016). Ethics and Childhood Vaccination Policy in the United States. *American Journal of Public Health, 106*(2), 273-278. doi:10.2105/AJPH.2015.302952

Hill, H. A., Elam-Evans, L. D., Yankey, D., Singleton, J. A., & Dietz, V. (2016). Vaccination Coverage Among Children Aged 19–35 Months — United States, 2015. *MMWR: Morbidity & Mortality Weekly Report, 65*(39), 1065-1071.

How How, C., Phua See Chun, P., Shafi, F., & Jakes, R. W. (2016). Parental knowledge, attitudes and perception of pneumococcal disease and pneumococcal conjugate vaccines in Singapore: a questionnaire-based assessment. *BMC Public Health, 16*, 1-10. doi:10.1186/s12889-016-3597-5

Hsiao, H. J., Wu, C. T., Huang, J. L., Chiu, C. H., Huang, Y. C., Lin, J. J., . . . Hsia, S. H. (2015). Clinical features and outcomes of invasive pneumococcal disease in a pediatric intensive care unit. *BMC Family Practice, 16*(1), 85-90. doi:10.1186/s12887-015-0387-7

Jeffery, C., Ozonoff, A., & Pagano, M. (2014). The effect of spatial aggregation on performance when mapping a risk of disease. *International Journal of Health Geographics, 13*(1), 1-9. doi:10.1186/1476-072X-13-9

MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine, 33*(34), 4161-4164. doi:10.1016/j.vaccine.2015.04.036

Marcus, J. L., Baxter, R., Leyden, W. A., Muthulingam, D., Yee, A., Horberg, M. A., . . . Silverberg, M. J. (2016). Invasive Pneumococcal Disease Among HIV-Infected and HIV-Uninfected Adults in a Large Integrated Healthcare System. *AIDS Patient Care & STDs, 30*(10), 463-470. doi:10.1089/apc.2016.0165

Meade, M. S., & Emch, M. (2010). *Medical Geography*: Guilford Publications.

Muley, V. A., Ghadage, D. P., Yadav, G. E., & Bhore, A. V. (2017). Study of Invasive Pneumococcal Infection in Adults with Reference to Penicillin Resistance. *Journal of Laboratory Physicians, 9*(1), 31-35. doi:10.4103/0974-2727.187918

Nhantumbo, A. A., Gudo, E. S., Caierão, J., Munguambe, A. M., Comé, C. E., Zimba, T. F., . . . Cantarelli, V. V. (2016). Serotype distribution and antimicrobial resistance of Streptococcus pneumoniae in children with acute bacterial meningitis in Mozambique: implications for a national immunization strategy. *BMC Microbiology, 16*, 1-9. doi:10.1186/s12866-016-0747-y

Nyangoma, E. N., Arriola, C. S., Hagan, J., Socias, C., Tomczyk, S., Watkins, L. F., . . . Kim, C. (2014). Hospitalizations for Respiratory Disease Among Unaccompanied Children from Central America -- Multiple States, June-July 2014. *MMWR: Morbidity & Mortality Weekly Report, 63*(32), 698-699.

Ordóñez, J. E., & Orozco, J. J. (2015). Cost-effectiveness analysis of the available pneumococcal conjugated vaccines for children under five years in Colombia. *Cost Effectiveness & Resource Allocation, 13*(1), 1-12. doi:10.1186/s12962-015-0032-1

Prosser, L. A., Ray, G. T., O'Brien, M., Kleinman, K., Santoli, J., & Lieu, T. A. (2004). Preferences and Willingness to Pay for Health States Prevented by Pneumococcal Conjugate Vaccine. *Pediatrics, 113*(2), 283-290.

Rio Grande Valley TX. (2018). Winter Texan. Retrieved from http://riograndevalleytx.us/winter-texan-info/

Sherchan, J. B., Tandukar, S., Shrestha, L., Rai, G. K., Sharma, A., Gami, F. C., . . . Ohara, H. (2015). Streptococcus pneumoniae among children with clinical meningitis in Nepal. *Journal of Institute of Medicine, 37*(3), 47-52.

Song, J. Y., Choi, J. Y., Lee, J. Y., Bae, I.-G., Kim, Y. K., Sohn, J. W., . . . Cheong, H. J. (2013). Clinical and economic burden of invasive pneumococcal disease in adults: a multicenter hospital-based study. *BMC Infectious Diseases, 13*(1), 1-9. doi:10.1186/1471-2334-13-202

Stacevičienė, I., Petraitienė, S., Vaičiūnienė, D., Alasevičius, T., Kirslienė, J., & Usonis, V. (2016). Antibiotic resistance of Streptococcus pneumoniae, isolated from nasopharynx of preschool children with acute respiratory tract infection in Lithuania. *BMC Infectious Diseases, 16*, 1-8. doi:10.1186/s12879-016-1544-9

Stopka, T. J., Goulart, M. A., Meyers, D. J., Hutcheson, M., Barton, K., Onofrey, S., . . . Chui, K. K. H. (2017). Identifying and characterizing hepatitis C virus hotspots in Massachusetts: a spatial epidemiological approach. *BMC Infectious Diseases, 17*, 1-11. doi:10.1186/s12879-017-2400-2

Taddio, A., Ipp, M., Thivakaran, S., Jamal, A., Parikh, C., Smart, S., . . . Katz, J. (2012). Survey of the prevalence of immunization non-compliance due to needle fears in children and adults. *Vaccine, 30*(32), 4807-4812. doi:10.1016/j.vaccine.2012.05.011

Tadesse, S., Enqueselassie, F., & Hagos, S. (2018). Spatial and space-time clustering of tuberculosis in Gurage Zone, Southern Ethiopia. *PLoS ONE, 13*(6), 1-14. doi:10.1371/journal.pone.0198353

Texas Department of State Health Services. (2017). Texas Notifiable Conditions. *Infectious Disease Control - Reporting.* Retrieved from https://www.dshs.texas.gov/idcu/investigation/conditions/

U.S. Census Bureau. (2016). 2012-2016 American Community Survey 5-Year Estimates. Retrieved from https://www.census.gov/programs-surveys/acs/

U.S. Census Bureau. (2017). Quick Facts - Hidalgo County, Texas. Retrieved from https://www.census.gov/quickfacts/fact/table/hidalgocountytexas/PST045216

van der Linden, M., Falkenhorst, G., Perniciaro, S., Fitzner, C., & Imöhl, M. (2016). Effectiveness of Pneumococcal Conjugate Vaccines (PCV7 and PCV13) against Invasive Pneumococcal Disease among Children under Two Years of Age in Germany. *PLoS ONE, 11*(8), 1-15. doi:10.1371/journal.pone.0161257

Wattal, C., Goel, N., & Byotra, S. P. (2017). Prevalence of pneumococcal serotypes in adults ⩾ 50 years of age. *Indian Journal of Medical Microbiology, 35*(1), 95-100. doi:10.4103/ijmm.IJMM_16_132

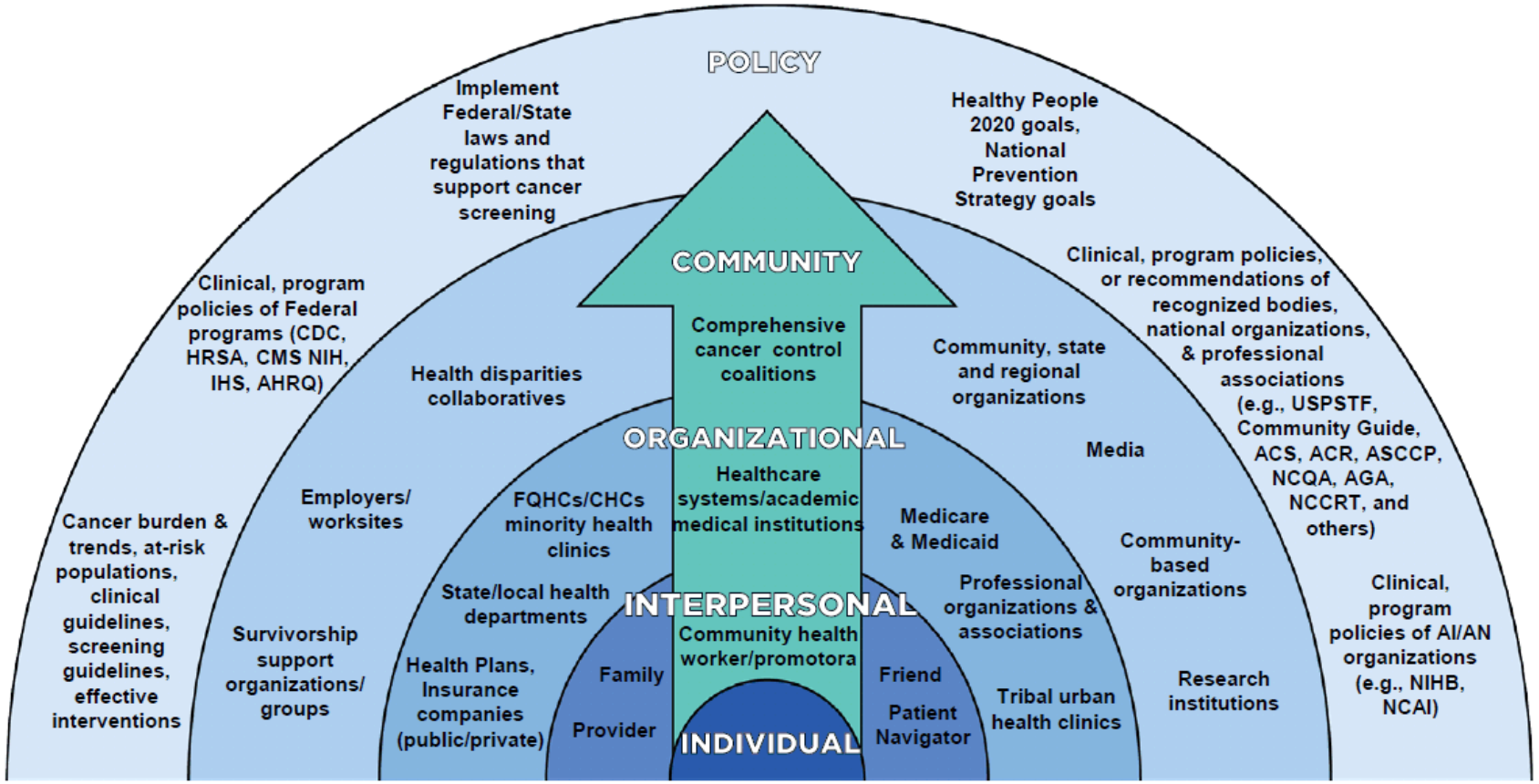
Weycker, D., Farkouh, R. A., Strutton, D. R., Edelsberg, J., Shea, K. M., & Pelton, S. I. (2016). Rates and costs of invasive pneumococcal disease and pneumonia in persons with underlying medical conditions. *BMC Health Services Research, 16*, 1-10. doi:10.1186/s12913-016-1432-4

Williams, W. W., Peng-Jun, L., O'Halloran, A., Kim, D. K., Grohskopf, L. A., Pilishvili, T., . . . Fiebelkorn, A. P. (2017). Surveillance of Vaccination Coverage Among Adult Populations -- United States, 2015. *MMWR Surveillance Summaries, 66*(11), 1-28.

Wu, H., Wang, X., Xue, M., Xue, M., Wu, C., Lu, Q., . . . Lin, J. (2017). Spatial characteristics and the epidemiology of human infections with avian influenza A(H7N9) virus in five waves from 2013 to 2017 in Zhejiang Province, China*12*, 1-17. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=124331164&site=ehost-live&scope=site doi:10.1371/journal.pone.0180763

APPENDICES

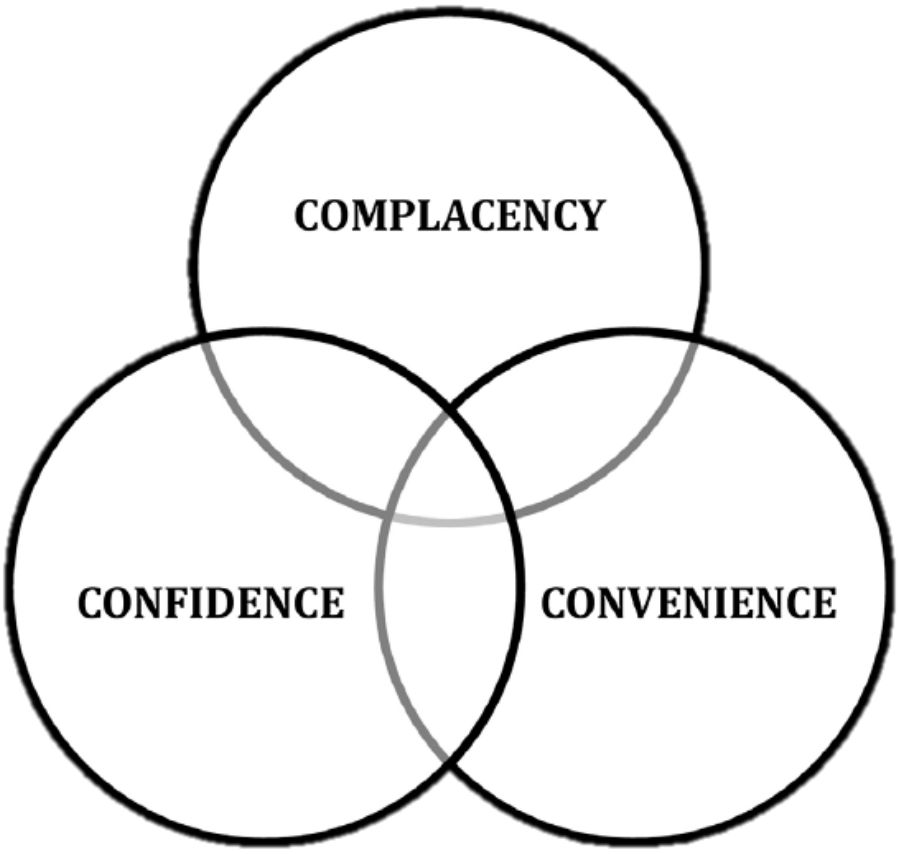
APPENDIX A: SOCIOECOLOGIC MODEL OF HEALTH



*Some groups may fit within multiple levels of this model.

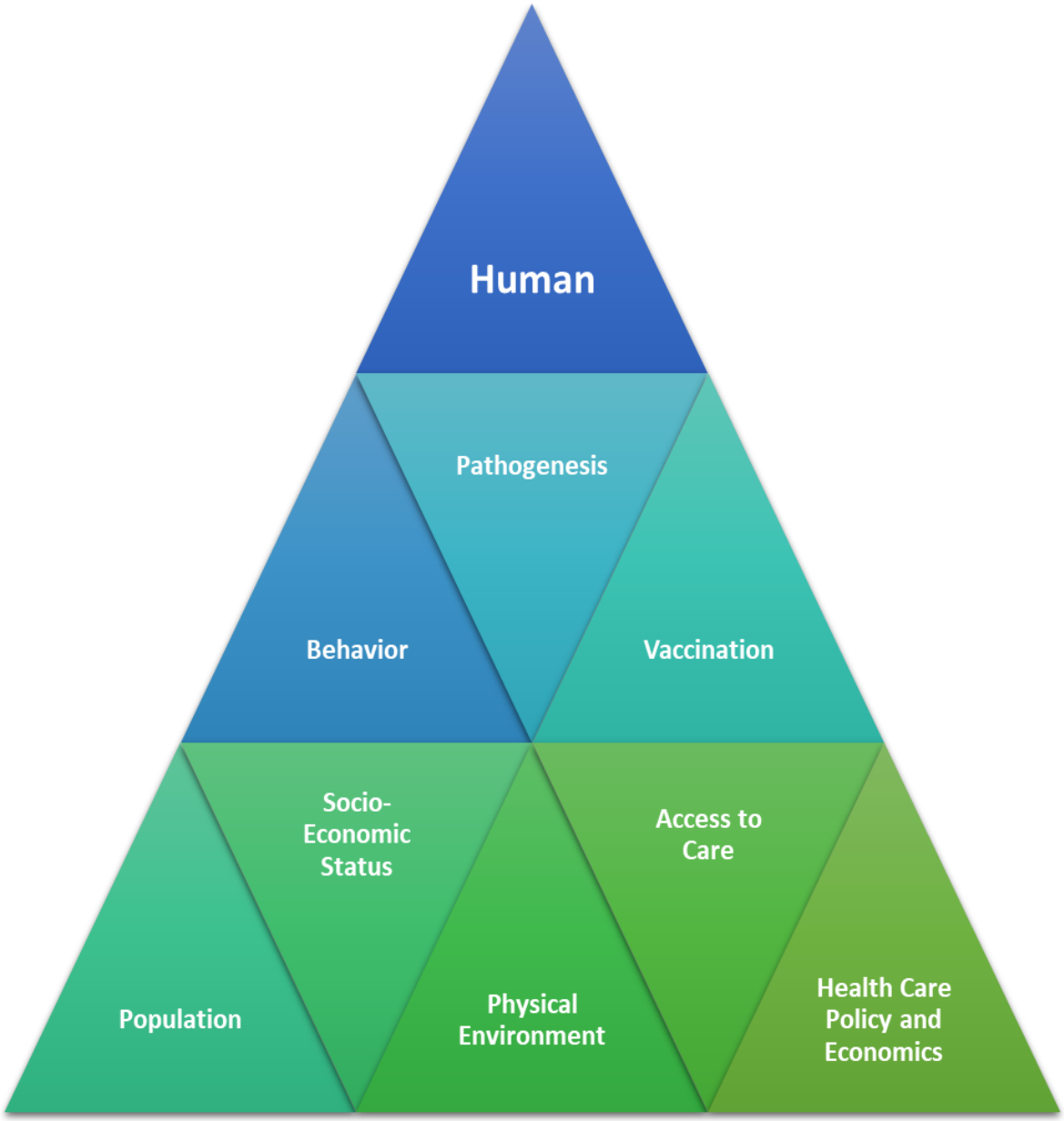
https://www.cdc.gov/cancer/crccp/sem.htm

APPENDIX B: THE 3 C's MODEL

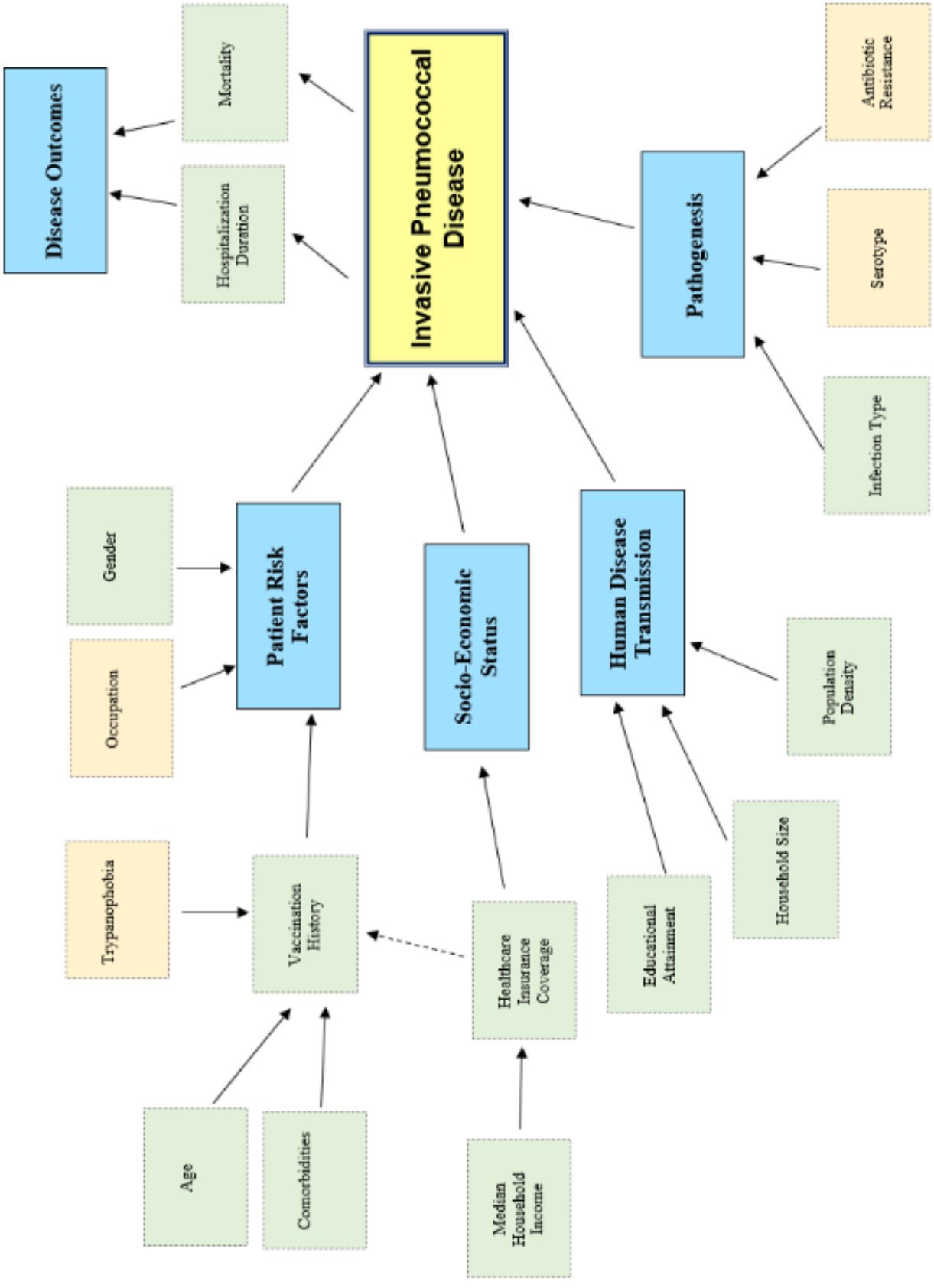


(MacDonald, 2015)

APPENDIX C: THE SOCIOECOLOGIC PYRAMID OF HUMAN DISEASE


Human
Pathogenesis
Behavior
Vaccination
Socio-
Economic
Status
Access to
Care
Population
Physical
Environment
Health Care
Policy and
Economics

APPENDIX D: CONCEPTUAL FRAMEWORK DIAGRAM


Disease Outcomes
Mortality
Hospitalization Duration
Invasive Pneumococcal Disease
Pathogenesis
Antibiotic Resistance
Serotype
Infection Type
Gender
Occupation
Patient Risk Factors
Socio-Economic Status
Human Disease Transmission
Population Density
Household Size
Educational Attainment
Trypanophobia
Vaccination History
Healthcare Insurance Coverage
Age
Comorbidities
Median Household Income

APPENDIX E: MOU BETWEEN HIDALGO COUNTY AND NORTHERN ILLINOIS UNIVERSITY

STATE OF TEXAS §
§
COUNTY OF HIDALGO §

MEMORANDUM OF AGREEMENT BETWEEN
HIDALGO COUNTY AND NORTHERN ILLINOIS UNIVERSITY

This Memorandum of Agreement is made on this the 12 day of September, 2017, by and between **HIDALGO COUNTY** by and through its Department of Health and Human Services, hereinafter referred to as "COUNTY", with administrative offices located at 1304 S 25th, Edinburg, TX 78542, and **NORTHERN ILLINOIS UNIVERSITY** by and through its College of Health and Human Sciences, hereinafter referred to as "NIU-CHHS", located at 1425 W. Lincoln Hwy., DeKalb, IL 60115.

The aim of this Agreement is to establish an exchange of information and for cooperative research between both institutions in the field of notifiable conditions.

WHEREAS, County, a political subdivision located in the State of Texas, and Northern Illinois University, an institution of Higher Education in the State of Illinois, which conducts research in the health and human sciences, desire to enter into this Agreement to collaborate and share information intended to improve the health, safety and welfare of the citizens of Hidalgo County, Texas; and

WHEREAS, the responsibilities set out herein are functions and/or services which each of the parties hereto have independent authority to perform and/or are functions and services which each of the parties hereto have independent authority to pursue, notwithstanding this Agreement;

NOW, THEREFORE, County and NIU-CHHS, in consideration of the mutual covenants expressed hereinafter, agree as follows:

I. PROVISION OF SERVICES

The COUNTY coordinates directly for health services to patients with notifiable conditions from Hidalgo County, and therefore has information of the characteristics of the patients and access to data on the infectious agents that have infected them. The faculty at NIU-CHHS has developed a comprehensive research program in the health sciences.

Through this Agreement, the parties shall develop by a cooperative agreement the following:

1. **Studies on the nature of the association between notifiable conditions and environmental, behavioral, and epidemiological health factors.** These studies will benefit all patients with notifiable conditions on the U.S.-Mexico border by providing information to the COUNTY, the public, and health providers. For each study, the NIU-CHHS researcher will forward a written narrative describing the study to the COUNTY Health Administrator who will assist in coordinating the study participation. The COUNTY, through the Health Administrator, may refuse to participate in a particular study at any time or halt data collection and participation at any time for any of these future studies.

The participation of COUNTY in this initiative shall consist of the following:

1. Providing NIU-CHHS researchers with information on the patients thought to be diseased with a notifiable condition, as well as individuals at risk of developing the condition.
2. Provide NIU-CHHS researchers with retrospective and prospective data for analysis and publication on notifiable condition and/or infectious disease trends in COUNTY.
3. Provide NIU-CHHS researchers with the relevant medical information on the Patient.

None of these studies will result in extra cost for the COUNTY. Further, Consenting Patients will not be exposed to any unnecessary or undue risk.

II. TERM OF AGREEMENT

The term of this Agreement will begin on September 12, 2017 and end on December 31, 2020.

III. COMPENSATION

COUNTY will receive no reimbursement from NIU-CHHS for any services provided. NIU-CHHS will receive no reimbursement from COUNTY for any services provided.

IV. NOTICES

All notices or other writing required under this Agreement shall be deemed to have been made when sent by certified or registered mail, return receipt request, to the following address:

TO NIU-CHHS:
James Ciesla
Associate Dean for Research and Resources
Northern Illinois University
1425 W. Lincoln Hwy.
DeKalb, IL 60115

TO COUNTY:
Hidalgo County Health & Human Services Department
Attn: Mr. Eduardo Olivarez
1304 S 25th Ave.
Edinburg, Texas 78542

V. TERMINATION

This Agreement may be terminated by either party by giving thirty (30) days written notice via certified mail, return receipt requested to the other party hereto of the intention to terminate.

VI. LAW GOVERNING VENUE

This Agreement shall be governed by and construed in accordance with the laws of the State of Texas, and, the obligations and undertakings of each of the parties to this Agreement shall be performable in Hidalgo County, Texas.

VII. IMMUNITIES

Nothing in this Agreement is intended to and County does not hereby waive, release or relinquish any right to assert any of the defenses County enjoys by virtue of the state or federal constitution, laws, rules or regulations, and any sovereign, official or qualified immunity available to County as to any claim or action of any person, entity, or individual against County.

WITNESS THE HANDS OF THE PARTIES effective as of the day and year first written above.

NOTHERN ILLINOIS UNIVERSITY

Gerald C Blazey VP Research
NIU Representative

Date: 8/28/17

HIDALGO COUNTY, TEXAS

Ramon Garcia
Ramon Garcia, County Judge

Date:

ATTEST:

Arturo Guajardo, Jr., County Clerk

COMMISSIONERS' COURT HIDALGO COUNTY, TEXAS

APPROVED BY COMMISSIONERS' COURT ON: 9/12/17

APPROVED AS TO FORM:
Office of Criminal District Attorney, Ricardo Rodriguez, Jr.

By:
Josephine Ramirez Solis
Assistant District Attorney

APPENDIX F: TEXAS DSHS IPD CASE DEFINITION

Condition/Code	Case Definition/Case Classification	Laboratory Confirmation Tests
***Streptococcus pneumoniae*, invasive disease (IPD)** 11723* *Note: Code 11717 was used prior to 2010 and for 2010 there are cases under both codes.	*Streptococcus pneumoniae* bacteria cause many clinical syndromes, depending on the site of infection (e.g., acute otitis media, pneumonia, bacteremia, or meningitis). Only invasive *Streptococcus pneumoniae* is reportable. ***Confirmed**:* A case that is laboratory confirmed ***Probable**:* A case with detection of *S. pneumoniae* from a normally sterile site using a culture independent diagnostic test (CIDT) (e.g., PCR, antigen based tests) without isolation of the bacteria Note: Positive lab results from a specimen collected more than 30 days after the collection date of a prior case should be counted as a new case. If specimen collection occurred within 30 days of the collection date of a prior case, it should not be counted as a new case.	▪ Isolation of *S. pneumoniae* from a normally sterile site (e.g., blood or cerebrospinal fluid, or, less commonly, joint, pleural, or pericardial fluid) See Normally Sterile Site and *Streptococcus* Classification

https://www.dshs.texas.gov/IDCU/investigation/epi-case-criteria-guide/Epi-Case-Criteria-Guide.doc

APPENDIX G: TEXAS DSHS IPD CASE INVESTIGATION FORM

TEXAS Department of State Health Services

Fax or mail this form to your local/regional health department
Contact your health department at 1-800-705-8868 www.dshs.state.tx.us/idcu/investigation/conditions/contacts/

Invasive Streptococcal Investigation Form

NBS Patient ID:

Patient's name: _____
Last First MI
Address: _____
City: _____ County: _____ Zip: _____
Phone 1: ( ) _____ Phone 2: ( ) _____
Date of birth: ___/___/___ Age: _____ Sex: ☐Male ☐Female ☐Unknown
Race: ☐White ☐Black ☐Asian ☐Pacific Islander ☐Native American/Alaskan
☐Unknown ☐ Other: _____ Hispanic: ☐ Yes ☐ No ☐ Unknown
Occupation: _____
Long-term care resident: ☐Yes, at: _____ ☐No ☐Unknown

Reported by: _____
Agency: _____
Phone: ( ) _____ Date reported: ___/___/___

Investigated by: _____
Agency: _____
Phone: ( ) _____
Email: _____
Investigation start date: ___/___/___

CLINICAL DATA

Physician's name: _____
Physician's phone: ( ) _____

Date of symptom onset: ___/___/___ Date illness ended: ___/___/___
Did patient die? ☐ Yes, date of death: ___/___/___ ☐ No ☐ Unknown

Type of Infection (Check all that apply): ☐ Bacteremia / Sepsis ☐ Pneumonia
☐ Toxic Shock Syndrome ☐ Necrotizing Fasciitis ☐ Meningitis ☐ Sinusitis
☐ Otis Media ☐ Endocarditis ☐ Peritonitis ☐ Septic Arthritis
☐ Other: _____

For Group B Strep investigations:
Pregnant: ☐ Yes: # weeks gestation: _____ ☐ No ☐ Unknown
Postpartum: ☐ Yes: date of delivery: ___/___/___ ☐ No ☐ Unknown

UNDERLYING HEALTH CONDITIONS

Does the patient have any underlying health conditions?
☐ Yes (check all that apply) ☐ No ☐ Unknown
☐ Asthma ☐ Chronic lung disease ☐ Cancer
☐ Cochlear implant ☐ Diabetes ☐ Heart disease
☐ Hemoglobinopathy ☐ HIV ☐ Kidney disease
☐ Organ transplant recipient ☐ Other: _____

Does the patient have high risk behaviors?
☐ Yes (check behaviors below) ☐ No ☐ Unknown
☐ Consumes raw (unpasteurized) milk/cheese
☐ Current smoker ☐ Intravenous drug user (IVDU)
☐ Alcohol abuse ☐ Other: _____

VACCINATION HISTORY (For *S. pneumoniae* investigations) **Source of vaccine history:** ☐ ImmTrac ☐ Parent ☐ Doctor ☐ School ☐ Other
Did the patient receive a pneumococcal vaccine? ☐ Yes ☐ No ☐ Unknown If yes, year vaccine was given: _____
If yes, which vaccine: ☐ Conjugate pneumococcal vaccine ☐ Polysaccharide pneumococcal vaccine ☐ Other: _____ ☐ Unknown

HOSPITALIZATION INFORMATION
Was the patient seen in an emergency room? ☐ Yes, name of hospital: _____ ☐ No ☐ Unknown
Was the patient hospitalized? ☐ Yes, name of hospital: _____ ☐ No ☐ Unknown
If yes, date of admission: / / Date of discharge: / /

LABORATORY DATA See DSHS' *Epi Case Criteria Guide* at www.idcu.org for case definitions and "normally sterile site" determination.

Date of collection	Test type	Sterile specimen source	Non-sterile specimen source	Was specimen collected during a surgical procedure?	Bacterial species identified*
___/___/___	☐ Culture ☐ Antigen ☐ PCR ☐ Antibody ☐ Other: _____	☐ CSF ☐ Blood ☐ Pericardial fluid ☐ Bone ☐ Joint fluid (no abscess) ☐Other: _____	☐ Wound ☐ Urine ☐ Skin ☐ Throat ☐ Joint fluid (abscess present) ☐Other: _____	☐ Yes ☐ No ☐ Unknown	☐ Group A Strep (*S. pyogenes*) ☐ Group B Strep (*S. agalactiae*) ☐ *Streptococcus pneumoniae* ☐ Other: _____
___/___/___	☐ Culture ☐ Antigen ☐ PCR ☐ Antibody ☐ Other: _____	☐ CSF ☐ Blood ☐ Pericardial fluid ☐ Bone ☐ Joint fluid (no abscess) ☐Other: _____	☐ Wound ☐ Urine ☐ Skin ☐ Throat ☐ Joint fluid (abscess present) ☐Other: _____	☐ Yes ☐ No ☐ Unknown	☐ Group A Strep (*S. pyogenes*) ☐ Group B Strep (*S. agalactiae*) ☐ *Streptococcus pneumoniae* ☐ Other: _____

COMMENTS

EADB Form ES9-11574 v(07/09/15) *This form is an investigation tool and is not required to be faxed to DSHS. Refer to the Epi Case Criteria for reporting requirements.

https://www.dshs.texas.gov/idcu/investigation/#