ARE WE STUDYING WHO WE THINK WE'RE STUDYING?
ROLE OF SOCIOECONOMIC STATUS IN THE VALIDITY OF ESTIMATES OF PNEUMOCOCCAL CONJUGATE VACCINE EFFECTIVENESS IN THE UNITED STATES

Ruth Link-Gelles

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2016

Approved by:

Daniel Westreich

Allison E. Aiello

Matthew R. Moore

Nong Shang

David J. Weber

ABSTRACT

Ruth Link-Gelles: Are we studying who we think we're studying? Role of socioeconomic status in the validity of estimates of pneumococcal conjugate vaccine effectiveness in the United States
(Under the direction of Daniel Westreich)

Thirteen-valent pneumococcal conjugate vaccine (PCV13) was licensed for use in children in the United States in February 2010. Shortly thereafter the Centers for Disease Control and Prevention began a post-licensure vaccine effectiveness (VE) study in 13 surveillance sites around the US. Cases were identified through active surveillance and controls were matched to cases by age (+/- 14 days) and zip code, which was used as a proxy for socioeconomic status (SES).

Due to issues locating and enrolled cases and controls in an era of increased cell phone usage, investigators were concerned that zip code may not provide adequate control for SES and that enrolled children may not be representative of eligible children, threatening both the internal and external validity of study results.

We obtained data on SES for cases and controls from a parent interview, birth certificates, and via geocoding and linkage to the US Census Bureau's American Community Survey. We used conditional logistic regression models to estimate the adjusted and unadjusted VE and to assess effect measure modification by SES of the estimated VE.

Small differences were found between enrolled cases and enrolled controls; however, these differences did not meaningfully change our estimated VE and we therefore concluded that internal validity of our estimates was high. Similarly, small differences were found between enrolled and

unenrolled cases, but we did not find effect measure modification, indicating that external validity of our estimates was high.

In addition to providing reassurance that previously published VE estimates are valid, we show that, despite not being able to contact unenrolled children, an assessment of validity of observational data is feasible.

To my mom, whose response is always “you can do it.”
Thank you for reminding me of that over and over again.

ACKNOWLEDGEMENTS

Many people made it possible for me to complete this dissertation. I am sincerely grateful for the help, support, and understanding I received along the way.

I would first like to thank Matt Moore, my supervisor at the CDC for the last six years. Matt gave me my first job out of my MPH program and fully supported my decision to go back to school two years later. He made it possible for me to continue working at CDC part-time from North Carolina, and, eventually, to use data from CDC for my dissertation. In short, without Matt, this dissertation would not exist. He humored me, celebrated my accomplishments, and was a sounding board when issues arose along the way. Beyond my dissertation work, Matt has been a fantastic mentor, leading by example, teaching me how to better communicate, and giving generously of his time. As I look forward to the next steps in my career, Matt will be a hard act to follow for future supervisors, and an example of the kind of supervisor and mentor I hope to be.

My dissertation committee chair, Daniel Westreich, has provided excellent mentorship over the last year and half that we've worked together. Daniel could have easily made my last year, commuting between Atlanta and Chapel Hill, miserable, but instead he was flexible, understanding, and always available when I needed guidance. His insights into my dissertation work were invaluable, as was his encyclopedic knowledge of the causal inference literature. His appreciation of superhero movies was welcome comic relief.

The rest of my committee, Allison Aiello, Nong Shang, and David Weber, provided much-needed expertise in their subject areas. I especially appreciated Allison's patience as I learned the world on socioeconomic research, which I knew little about when this project began. Nong was infinitely patient

with questions about biostatistics and David provided great insight into the practical implications of this work. All three vastly improved my dissertation.

I would also like to thank Annelies Van Rie, my first advisor at UNC. Annelies gave me the opportunity to learn a new area of epidemiology (TB and HIV in South Africa) and helped me craft my initial dissertation aims and first draft of my proposal. Thank you.

Nancy Colvin, Carmen Woody, Valarie Hudock, and Jennifer Moore kept and keep the Epidemiology Department running like a well-oiled machine, providing endless advice and support. I would especially like to thank Valarie for her support during a particularly difficult time in the program for me. Her understanding and encouragement were pivotal at keeping me on track.

Of course, this work would not have been possible without the children and parents who participated in the study and without the site surveillance staff who put in endless hours collecting and cleaning the data so I could analyze it.

Finally, I would like to thank my friends and family, who put up with me, encouraged me, and (when needed) reminded me why I was getting a PhD in the first place. My husband, Dan, was a constant source of encouragement, sticking with me for 2.5 years of a long distance relationship, talking me off a ledge the night before my qualifying exam, and never doubting that I could finish (even when I did). Your humor and understanding kept me sane.

TABLE OF CONTENTS

LIST OF TABLES xii

LIST OF FIGURES xiii

LIST OF ABREVIATIONS xiv

CHAPTER 1: SPECIFIC AIMS 1

CHAPTER 2: BACKGROUND AND SIGNIFICANCE 4

Role of vaccines and vaccine effectiveness evaluations on public health 4

Impact of vaccines on public health 4

Path to vaccine licensure in the US 4

Importance of post-licensure evaluations for estimating the impact of vaccines 5

Pneumococcal disease and vaccine in the US 6

The burden of pneumococcal disease in the pre-conjugate vaccine era 6

Licensure, introduction and effectiveness of PCV7 in the US 7

Serotype replacement and the need for PCV13 8

Licensure, and recommendations for use and uptake of PCV13 8

Socioeconomic disparities, pneumococcal disease and vaccination in the US 10

Socioeconomic health disparities 10

Disparities and risk of invasive pneumococcal disease 10

Socioeconomic disparities, vaccination, and the vaccines for children program 11

Remaining disparities in vaccination coverage 11

Individual and neighborhood level measures of socioeconomic status 12

Past research on representativeness by socioeconomic status in VE studies in the US 13

CHAPTER 3: PARENT STUDY: PCV13 VACCINE EFFECTIVENESS STUDY 15
Active Bacterial Core surveillance system and PCV13 VE extended area surveillance 15
IPD case definition 16
Case enrollment procedures 16
Control matching 16
Exposure assessment – provider follow-up and IIS 17
Calculation of vaccine effectiveness and primary objective 17
CHAPTER 4: RESEARCH METHODS 20
Population included in SES analysis 20
Covariate assessment/definitions 20
ABCs case report form 20
Birth certificate data 22
Parent interview/provider follow-up 23
American Community Survey 26
Ethics approval 27
CHAPTER 5: ANALYTIC METHODS 32
Aim 1 32
Identification of confounders 32
Comparison groups 32
Control for residual confounding 33
Aim 2 34
Socioeconomic Position Index 35
Choosing variables for modification assessment 35
Predictive model 36

Assessment of effect measure modification 37

CHAPTER 6: BIAS WITH RESPECT TO SOCIOECONOMIC STATUS: A CLOSER LOOK AT ZIP CODE MATCHING IN A PNEUMOCOCCAL VACCINE EFFECTIVENESS STUDY 40

Overview 40

Introduction 40

Methods 42

Results 46

Discussion 49

CHAPTER 7: GENERALIZABILITY OF VACCINE EFFECTIVENESS ESTIMATES: AN ANALYSIS OF CASES INCLUDED IN A POST-LICENSURE EVALUATION OF 13-VALENT PNEUMOCOCCAL CONJUGATE VACCINE 58

Overview 58

Introduction 58

Methods 60

Results 66

Discussion 68

CHAPTER 8: DISCUSSION 79

Overview 79

Summary of findings 80

Aim 1 81

Aim 2 82

Strengths 82

Limitations 83

Future directions 83

Conclusions 84

APPENDIX 1: PNEUMOCOCCAL VACCINES USED IN THE US SINCE 2000 85

APPENDIX 2: 2014 ABCS CASE REPORT FORM 86

APPENDIX 3: MEDICAL AND VACCINE HISTORY FORM 88

REFERENCES 90

LIST OF TABLES

Table 4.1. Variables, sources of information, and availability by case and enrollment status 29

Table 4.2. American Community Survey variables included in prediction models for enrollment..... 30

Table 5.1. Operational definitions of ACS variables included in final prediction model. 39

Table 6.1. Number of unique zip codes and census tracts for eligible and enrolled children, by case status and serotype of disease 53

Table 6.2. Characteristics of eligible cases and matched controls. Data come from (a) the parent interview/medical provider, (b) birth certificates, or (c) American Community Survey. 56

Table 6.3. Comparison of results of original model vs. models adjusted for minimally sufficient subsets (MSS) for effectiveness against PCV13-type and non-PCV13-type disease. 57

Table 7.1. Enrollment numbers, reasons for non-enrollment, and geocoding status for cases 71

Table 7.2. Characteristics of eligible cases by enrollment status. (a) ABCs case report form (cases only); (b) birth certificate variables; (c) American Community Survey. 72

Table 7.3. Prediction models for case enrollment..... 75

Table 7.4. Operational definitions of ACS variables included in final prediction model 76

LIST OF FIGURES

Figure 3.1. Map of surveillance area for PCV13 VE Study, by year added to ABCs 18

Figure 3.2. Enrollment procedures for cases and controls in the PCV13 VE Study 19

Figure 4.1. Enrollment numbers for SES analysis 28

Figure 5.1. Causal diagram showing the relationship between vaccination status, IPD, and confounders. 38

Figure 7.1. Vaccine effectiveness (and 80% CIs) for characteristics retained in the final predictive model, stratified as indicated. (a) Individual variables; (b) ACS variables, divided at median. 77

LIST OF ABREVIATIONS

ABCs	Active Bacterial Core surveillance system
ACIP	Advisory Committee on Immunization Practices
ACS	American Community Survey
AIC	Akaike information criteria
APNCU	Adequate Prenatal Care Utilization Index
BLA	Biologic License Application
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CRF	Case report form
CT	Census Tract
DAG	Directed acyclic graph
DTaP	Diphtheria, tetanus, and acellular pertussis vaccine
DTP	Diphtheria, tetanus, and whole cell pertussis vaccine
EMM	Effect measure modification
FDA	Food and Drug Administration
ICU	Intensive Care Unit
IIS	Immunization information system
IPD	Invasive pneumococcal disease
IPV	Inactivated Polio Vaccine
IQR	Interquartile range
IRB	Institutional Review Board
MMR	Measles, Mumps, and Rubella Vaccine
MOR	Matched odds ratio

MSS	Minimally sufficient set
NIS	National Immunization Survey
PCV7	7-valent pneumococcal conjugate vaccine
PCV13	13-valent pneumococcal conjugate vaccine
PPSV23	23-valent pneumococcal polysaccharide vaccine
SEP	Socioeconomic Position Index
SES	Socioeconomic status
RCT	Randomized controlled trial
RDD	Random digit dialing
VAERS	Vaccine Adverse Events Reporting System
VFC	Vaccines for Children
VE	Vaccine effectiveness
WHO	World Health Organization

CHAPTER 1: SPECIFIC AIMS

In the past decade, many vaccines approved for children in the United States[1-4] have been licensed on the basis of observational immunogenicity and safety data, rather than on the basis of randomized controlled trials (RCT) of vaccine efficacy. In such cases, post-licensure observational studies of vaccine effectiveness (VE) are often the only opportunity to evaluate whether a vaccine is working as expected. While RCTs estimate vaccine efficacy under carefully controlled use and administration and typically have high internal validity, they are often subject to problems with generalizability to the broader population (external validity).[5] Post-licensure observational studies, however, provide the opportunity to assess VE under typical use in the general population (increased external validity), while issues of confounding are of greater concern than in RCTs (decreased internal validity).

One of the most common types of observational VE study designs is the case-control study, with cases identified through active surveillance, and with retrospective ascertainment of vaccination status for both cases and healthy controls. In order to increase statistical efficiency when the disease outcome under study is rare (as is the case for most vaccine preventable diseases), controls are frequently matched to cases based on factors which are likely to confound the relationship between vaccination and disease, age being the most common matching factor. While the ideal case-control VE study would enroll 100% of cases and a random sample of the target population as controls, this is often not feasible. Requirements such as interviewing participants to obtain behavioral risk factors and vaccine histories can result in study populations that differ from target populations. If differences between the study population and the target population include factors which modify the effect of vaccine on disease, external validity is reduced.[6]

Socioeconomic status (SES), generally measured as a combination of income, education, and occupation, is likely to be related to whether or not an individual is located and enrolled in an observational study.[7,8] Because less affluent individuals may move or change telephone numbers more frequently, they may be harder to find than more affluent individuals. Less affluent individuals may have less access to or trust in the medical system, and may not be fluent in English, all factors which decrease their likelihood of consenting to participate in a study.[9-11] At the same time, SES is also an increasingly recognized risk factor for many vaccine preventable diseases since it changes exposure patterns, increases stress, and decreases access to care;[12-17] as such, it may be an effect measure modifier of the impact of vaccine on the disease. Therefore, generalizability with respect to SES may be reduced.

In addition to being a risk factor for both selection into a study and disease outcome status, SES is *also* associated with vaccination status,[18-22] meaning that SES may confound the vaccine/disease relationship. To deal with such potential confounding, studies often match cases to controls on neighborhood (using zip code in the US), a common proxy for SES.[23,24] Little research to date has focused on understanding how SES may reduce generalizability and/or confound the relationship between vaccination and disease, and whether matching on zip code (or other neighborhood indicators) eliminates this confounding.

We propose to use data from a VE study of 13-valent pneumococcal conjugate vaccine (PCV13) in the US[25] to assess how SES may affect both the internal and external validity of case-control studies of VE. Using these data, we will evaluate the **internal** and **external** validity of the estimate of VE from this study in two separate aims. Specifically, we will:

Aim 1. Evaluate whether zip code matching results in a similar distribution of SES in cases and controls; estimate the amount of residual confounding and appropriately adjust the main VE estimate.

Hypothesis: Although matching by zip code likely controls for some confounding by SES, we expect individual (as measured on the birth certificate) and neighborhood (as measured by census tract) level SES indicators to differ between enrolled cases and controls, indicating the potential for residual confounding. We hypothesize this residual confounding to meaningfully change the VE estimate, which for purposes of this analysis will be a 5% absolute change in the estimated VE.

Aim 2. Assess generalizability in a population-based case-control study by:

(a) assessing whether enrollment rates differ with respect to SES, and

(b) determining whether SES is an effect measure modifier of the estimated vaccine effectiveness.

Hypothesis: Enrolled cases will differ substantially from the target population by important characteristics, including maternal race/ethnicity, crowding in the household, and prenatal care quality, all markers of SES. Additionally, because SES is related to immune system status (through nutrition, stress, comorbidities, etc.) and increased risk factors for IPD (e.g., crowding, smoking exposure),[26] effect measure modification by SES, and therefore reduced generalizability, is likely.

Together these aims will help improve both the internal and external validity of future VE studies for IPD, with potential applications to other VE studies.

CHAPTER 2: BACKGROUND AND SIGNIFICANCE

Role of vaccines and vaccine effectiveness evaluations on public health

Impact of vaccines on public health

In 1999, the Centers for Disease Control and Prevention listed universally recommended vaccines for children as one of the top ten achievements in public health of the past century.[27] Of the 10 diseases for which universal vaccination was recommended at some point in the 20th century, cases in the United States have declined by 96-100% since introduction.[28] Smallpox has been eradicated and polio is near eradication.[29]

Vaccines were again highlighted as a top achievement in 2011, when CDC estimated that vaccination of each US birth cohort under the current schedule prevents approximately 42,000 deaths and 20 million cases of disease, with corresponding cost savings in the tens of billions.[29] Worldwide, vaccines were estimated to have prevented 2.5 million deaths in children under age five in between 2001-2010.[30]

Path to vaccine licensure in the US

Development and licensure of new vaccines in the US is a lengthy process, often lasting a decade or more.[31] The first step toward licensure is a Phase I trial of the vaccine to evaluate safety. These trials are generally extremely small and do not attempt to make conclusions about the efficacy or effectiveness. Phase II clinical trials include a few hundred to a few thousands participants and aim to evaluate immunogenicity, efficacy, and safety. Finally, for products showing efficacy, but lacking severe adverse events, Phase III trials are conducted. These trials can include up to tens of thousands of individuals and are the last step before the manufacturer files a Biologic License Application (BLA) with

the Food and Drug Administration (FDA).[31] The BLA is then reviewed by the FDA and may be presented to the Vaccines and Related Biological Products Advisory Committee if data on safety, efficacy, and appropriate use need review. During this process the manufacturer works with the FDA to refine the indications for use and other statements that will be included in the package insert at the time of licensure. Post-licensure changes in the package insert can result from additional human trials or safety concerns.[31]

After licensure, CDC's Advisory Committee on Immunization Practices (ACIP) reviews the available data and makes recommendations for the actual use of the vaccine in the US. Usually, these recommendations mirror those in the package insert approved by FDA, however, ACIP can make "off label" recommendations.[32] Recommendations are formally issued through CDC's *Morbidity and Mortality Weekly Report*.

Importance of post-licensure evaluations for estimating the impact of vaccines

Once a vaccine is introduced, post-licensure monitoring of safety, impact and effectiveness begin. FDA and CDC maintain the Vaccine Adverse Events Reporting System (VAERS) for tracking safety. CDC also generally tracks impact and effectiveness through active and passive surveillance and observational vaccine effectiveness (VE) studies. Post-licensure VE studies have become increasingly important as new vaccines are licensed to replace existing vaccines and it becomes unethical to conduct long-term pre-licensure studies when a recommended vaccine is already in use.

Such observational studies have a number of strengths over pre-licensure studies. First, they assess real-world effectiveness (i.e., few exclusion criteria, administration of doses on altered or delayed schedules, herd protection), rather than the controlled environment of an RCT.[33] Second, they sometimes help identify groups where the ACIP recommendations are not being fully implemented, and can also identify vaccine failures.[34,35] Third, post-licensure studies allow estimate of direct and indirect benefits,[36] whereas traditional, individually randomized controlled trials only allow estimation of direct

benefits.[37] Limitations of post-licensure observational studies compared to pre-licensure RCTs include a potential for uncontrolled confounding, issues with selection bias, and misclassification of exposure (vaccination status).

Most post-licensure VE studies conducted in the US are population-based, nested case-control studies, often with matching to efficiently control for confounding by age and other factors.[38-40] Since 2000, CDC has repeatedly conducted post-licensure VE studies for vaccines against influenza,[41] pneumococcal disease,[42] and meningococcal disease[43] within defined geographic active-surveillance areas. Other groups have used individuals in health maintenance organizations as the target population, especially for influenza studies.[44] In both cases matching on zip code is a common way to attempt to control for confounding by socioeconomic status.[42-44] This study design, in turn, has been replicated around the world with neighborhood, rather than zip code, matching.[45-49] Studying the validity of these methods is the focus of the present work.

Pneumococcal disease and vaccine in the US

The burden of pneumococcal disease in the pre-conjugate vaccine era

Streptococcus pneumoniae (pneumococcus) is a major cause of morbidity and mortality worldwide. Prior to the development and introduction of 7-valent pneumococcal conjugate vaccine (PCV7) in the early 2000s, pneumococcus was the leading cause of pneumonia, bacteremia, sinusitis, and acute otitis media worldwide, and also a major cause of bacterial meningitis.[2] The World Health Organization (WHO) estimates that in 2000, pneumococcus caused 14.5 million cases and 735,000 deaths in HIV-negative children under the age of five, 7% of all-cause mortality in this age group.[50] In 1999, the year before PCV7 was introduced in the United States, the Centers for Disease Control and Prevention (CDC) estimates there were 64,000 cases (23.63/100,000 persons) and 7,300 deaths in all age groups in the United States.[51]

Licensure, introduction and effectiveness of PCV7 in the US

PCV7 (Prevnar™, Wyeth Lederle) was licensed by the US Food and Drug Administration in February 2000 and covers the seven serotypes that accounted for approximately 80% of invasive pneumococcal infections in the United States in children under six in the pre-vaccine era: 4, 6B, 9V, 18C, 14, 19F, and 23F. Licensure of PCV7 was based primarily on a prospective double-blind randomized controlled trial (RCT) among 37,830 healthy children in a large health maintenance organization in California.[52] The trial showed PCV7 efficacy against IPD caused by one of the seven serotypes included in the vaccine was 93.9% (95% Confidence Interval [CI]: 79.6-98.5%) for one or more doses and 97.4% (95% CI: 82.7-99.9%) for the full, four dose, schedule.[2,52] Additionally, the study found that children who received at least one dose of PCV7 had fewer episodes of less-specific outcomes such as all-cause pneumonia and acute otitis media.[2,52]

Other factors considered during licensure of PCV7 were double-blind RCTs of a similarly-formulated 9-valent pneumococcal vaccine among toddlers attending day care in Israel and infants in Soweto, South Africa, as well as immunogenicity and safety data from.[2,53-61] Both the Israeli and South African studies found significant declines in nasopharyngeal carriage of vaccine-included serotypes in children who received vaccine compared with children who received placebo (or a non-pneumococcal vaccine).[53,55]

In the spring of 2000, CDC's Advisory Committee on Immunization Practices (ACIP) recommended PCV7 for routine use in the infant immunization series on a "3+1" schedule, that is a three-dose primary series at two, four, and six months of age, with a booster recommended between 12 and 15 months of age. For children aged 24-59 months with specific underlying conditions, ACIP recommended two doses of PCV7, followed at least two months later by a dose of the older 23-valent pneumococcal polysaccharide vaccine (PPSV23) which, while not efficacious against otitis media or non-bacteremic pneumonia, covers a broader range of serotypes (Appendix 1).[2]

From January 1, 2001 through May 31, 2004 CDC conducted a matched case-control study to estimate the effectiveness of PCV7 in the United States in children under five years of age.[42] One or more doses of PCV7 was found to be 72% effective (95% CI: 65-75%) against all serotypes. Against the seven serotypes included in the vaccine, VE was found to be 96% (95% CI: 93-98%) in otherwise healthy children and 81% (95% CI: 57-92%) in children with comorbid conditions such as cancer and sickle cell anemia.[42] The RCT, which only enrolled healthy infants, found an efficacy against IPD of 97.4% (CI: 82.7-99.9%).[52] The full, 3+1 schedule, was found to be 100% effective (95% CI: 94-100%) compared with no doses.[42]

Serotype replacement and the need for PCV13

Incidence of vaccine-type IPD decreased in the US rapidly after vaccine introduction in both children, due to direct effects, and adults, due to indirect effects.[62-65] Unfortunately, however, as rates of vaccine-type IPD decreased, rates of non-vaccine-type disease began to increase, a phenomenon termed "serotype replacement" that has also been seen to varying degrees in other countries after introduction of PCV7.[66-68] While serotype replacement in the US did not reach high enough levels to cancel out the gains since PCV7, replacement was considered problematic enough to warrant the licensure of a 13-valent vaccine (PCV13).

Licensure, and recommendations for use and uptake of PCV13

PCV13, (Prevnar13™, Pfizer) includes the 7 serotypes in PCV7 and 6 additional serotypes, 1, 3, 5, 6A, 7F, and 19A. PCV13 was licensed by FDA on the same schedule as PCV7, i.e., a three-dose primary series at two, four, and six months of age and a booster dose at 12-15 months of age. Additional doses were recommended for older children with specific underlying conditions.[25] Because PCV13 is manufactured in exactly the same way as PCV7 it was thought to be as safe and immunogenic as PCV7, a hypothesis confirmed by Phase I and II trials. The extra six serotypes in PCV13 meant that the vaccine

would, at minimum, be no less effective than PCV7 and it was therefore not considered ethical to undertake a large RCT of the efficacy of PCV13. Instead, the new vaccine was licensed based on safety and immunogenicity data, with the requirement that post-licensure effectiveness studies would follow introduction.[69] Safety was assessed through 13 separate clinical trials, including a total of 4,729 infants and toddlers.[70] Immunogenicity was evaluated in a double-blind RCT of 663 infants. The study assessed whether or not PCV13 elicited pneumococcal IgG antibody concentrations noninferior to those elicited by PCV7. A cutoff of ≥0.35 μg/mL 1 month after a dose was considered noninferior. Functional responses via opsonophagoocytosis assay (OPA) were measured for a subset of the study population. The primary noninferiority criteria were met for all serotypes, except serotype 3, which failed to elicit a noninferior response after the fourth dose.[68]

Based on these results, PCV13 was licensed by the FDA on February 24, 2010 for the prevention of IPD caused by one of the 13 serotypes included in the vaccine and for prevention of otitis media caused by one of the 7 serotypes included in PCV7.[25] ACIP approved PCV13 the same day and recommended it replace PCV7 for routine.[25]

Uptake of PCV13 in young children in the US was rapid owing to the fact that the vaccine replaced an existing vaccine in the routine schedule.[71] By 2012, 18 months after PCV13 was introduced, coverage amongst 19-35-month-olds was 92.3% (+/- 0.8%) for three doses and 81.9% (+/- 1.1%) for four doses, only two percentage points below the corresponding coverage estimates for diphtheria, tetanus, and acellular pertussis (DTaP) vaccine.[72]

Almost immediately after licensure, CDC began conducting a case-control study of PCV13 to assess vaccine effectiveness against IPD in children in the US[46,69,73] CDC's VE Study began enrolling cases on May 1, 2010 in 13 sites around the country and enrollment concluded on May 31, 2014. This study is further explained below as the parent study for the proposed research (see Chapter 3).

Socioeconomic disparities, pneumococcal disease and vaccination in the US

Socioeconomic health disparities

Healthy People 2020 defines health disparities as differences in health that are "closely linked with social, economic, and/or environmental disadvantage."[74] In particular, disparities "adversely affect groups of people who have systematically experienced greater obstacles to health" based on characteristics including racial/ethnic group, socioeconomic status (SES), geographic location, etc.[74] Research has shown wide disparities in health indicators ranging from cancer incidence and survival, to mental health, to diabetes, to infectious diseases.[24,60,75,76] Theories for these disparities include *in utero* and childhood living conditions, lifetime stress, and differential access to preventive health services and adequate nutrition.[12,16,26,76,77]

Disparities and risk of invasive pneumococcal disease

Racial disparities in IPD incidence have been a persistent problem in the US[78,79] While incidence rates dropped after introduction of PCV7 across all racial and ethnic groups, disparities remained. Indeed, while rates dropped by 89% (95% CI: 85-92%) in blacks under two years of age by 2002 compared with only 77% (95% CI: 72-81%) in whites of the same age group, IPD rates in blacks remained substantially higher than rates in whites.[78] Disparities remained through 2009, with the rate ratio comparing blacks to whites actually *increasing*, due primarily to rates of non-PCV7-type IPD increasing faster in blacks than whites.[79] While PCV13-types accounted for 71% of cases in white children under age 5 in 2009, types included in the vaccine accounted for only 58% of cases in black children in this age group, meaning racial disparities are likely to persist after PCV13 introduction.[79]

These racial disparities may be proxies for socioeconomic disparities, including differences in the prevalence of underlying conditions, higher poverty rates amongst blacks, differences in health insurance coverage, smoking rates, and crowding.[15,79-81] Identifying and understanding these disparities

and their relationship with IPD is an important component of future disease control efforts, such as targeted immunization campaigns.

Socioeconomic disparities, vaccination, and the vaccines for children program

Historically, immunization rates for children of lower income households in the US for the entire childhood immunization schedule have lagged behind rates in children from wealthier households, especially amongst the youngest children.[18] A major resurgence of measles in the US between 1989 and 1991 was primarily concentrated in unvaccinated children less than five years of age, disproportionately from low-income, black, and Hispanic families.[19-21] As a result, the Presidential Childhood Immunization Initiative was developed in 1993 and Congress authorized the Vaccines for Children (VFC) program the following year.[19] Since then, the VFC program has provided free vaccines to children who are Medicaid-eligible, uninsured, American Indian/Alaska Native, or underinsured. Implemented and managed by CDC since its inception, the VFC program is able to buy vaccines at a discount price and distribute them to Federally Qualified Health Centers (registered private physicians' offices and public health clinics).[19] In 2011, a total of 54% of US children aged 19-35 months were VFC eligible, with the largest contingent qualifying due to Medicaid eligibility. Of these VFC eligible children, 52% belonged to a racial/ethnic group other than non-Hispanic white.[19]

Remaining disparities in vaccination coverage

A 2014 analysis of National Immunization Survey (NIS) data showed that while disparities in coverage of measles, mumps, and rubella vaccine (MMR), polio vaccine (IPV), and DTaP have decreased since the VFC program began, some disparities remained in 2011, especially when comparing non-Hispanic whites and non-Hispanic blacks. In particular, the authors found that, "the disparities in estimated DTP/DTaP coverage between non-Hispanic black and non-Hispanic white children did not decrease significantly during 1995-2011."[19]

These disparities were seen for pneumococcal vaccine in the early 2000s, when a number of temporary shortages of PCV7 resulted in amended ACIP recommendations and reduced-dose schedules. Children most likely to be affected by the shortages included Hispanics, non-Hispanic blacks, non-Hispanic American Indians, children who had received prior doses from all public, hospital, military, or mixed-type providers (vs. private providers), children whose mother was never married or widowed/divorced/separated, children whose mother had not completed college, and children living in a household with an annual income less than 135% of the federal poverty level.[82]

In addition to truly lower vaccination coverage amongst racial and ethnic minorities, some research has shown that ascertainment of vaccination histories is lower amongst children reporting two or more vaccine providers (e.g., transient children, or those with unstable access to care), adding to perceived disparities.[22] However, even after adjusting for these ascertainment issues, the authors reported that disparities in coverage persisted.[22]

Individual and neighborhood level measures of socioeconomic status

Socioeconomic status (SES) is a measure of how an individual or family's economic position relates to others, usually in the same geographic area.[83] While no universal definition exists, individual-level SES is generally measured as a combination of income, education, and occupation.[7,8] Race/ethnicity are sometimes also considered as they correlate with income level, are often associated with health status, and are frequently easier to obtain for research than income.[13,16,84-86] Additional characteristics, such as health insurance status, have also been used.[87]

In addition to a mix of individual-level characteristics, SES can be measured at the neighborhood or area level, using factors such as unemployment rate, median area income, crime statistics, food scarcity, and crowding in the household, etc.[7,12,13,23,24]

In the US, the most common way to obtain neighborhood-level SES is through census areas, such as county, zip code, or census tract.[88] Zip code has been frequently used in the past; however, zip

codes are defined by the US Postal Service for convenient mail delivery, are often heterogeneous with respect to SES, and change frequently.[23] Researchers have moved toward using census tracts and blocks, which are defined by the US Census Bureau.[88] Census tracts (CTs), the most common area-level SES measure in the US, generally have between 1,200 and 8,000 individuals and are maintained consistently over long periods of time for ease of statistical comparisons from one census to the next.[88]

Past research on representativeness by socioeconomic status in VE studies in the US

A commonly listed benefit of observational studies compared to RCTs is increased representativeness of the source population.[89-91] RCTs generally have numerous exclusion criteria, whereas observational studies usually have more limited exclusion criteria.[89-91] However, this benefit may be limited when observational studies fail to enroll a specific subset of their cases and/or controls, due to either an inability to locate them or to high refusal rates. If the reason a certain subset of the population was not enrolled is a modifier of the main exposure/disease relationship under study, then the results will not be generalizable to the entire target population. For example, if lower SES individuals are not enrolled, due to frequent moves or a lack of telephone, and the intervention under study is less effective in less affluent people, generalizability will be reduced.[92]

Research into generalizability in VE studies is extremely limited and has focused primarily on adult influenza vaccine. These studies have little application to childhood vaccines as the issues arise from enrollment of individuals living in long-term care facilities, healthy user bias, and vaccine affordability issues, which are mitigated by the VFC program.[93-95]

A few studies have looked at methods, such as random digit dialing (RDD) and commercial databases, for identifying and recruiting individuals into observational studies in areas other than vaccine effectiveness. The most common methods for identifying and recruiting controls are birth certificates, random digit dialing (RDD), commercial databases, and neighborhood/friend controls. Each has potential benefits and drawbacks, with RDD likely the least useful in recent years since area codes

no longer indicate that an individual lives in a certain region.[96-98] Commercial databases often have the most accurate contact information, but only include a subset of otherwise eligible individuals, making the source population difficult to identify.[96] Birth certificate controls are likely to be the most representative, but the data can be difficult to acquire and contact information may be out of date, especially for older children, making it difficult to actually locate the individual.[98] Neighborhood/friend controls should, in theory, allow easy matching for SES factors, but past research has found this is not always the case.[97]

While rare for studies in most research areas to report differences between enrolled and unenrolled subjects, researchers of childhood leukemia and magnetic field exposure have taken an interest in generalizability as a way to explain significant effects found in observational studies that cannot be explained by a biological mechanism. A few studies have attempted to assess differences in participation by SES and case status and have found lower participation, particularly by controls, in census tracks or neighborhoods with indicators of lower SES.[99-102]

All enrollment methods may be susceptible to differential enrollment by SES factors if these factors are associated with researchers' ability to locate controls and/or obtain their consent.[103] One recent study used a commercial database that classified neighborhoods into "lifestyle" clusters, using demographics augmented by data on consumer purchases. The "lifestyle" controls allowed the authors to compare SES characteristics of enrolled and unenrolled controls, finding higher enrollment rates amongst the wealthier clusters. In the least affluent groups, the proportion of individuals who consented was much higher in cases than controls.[97]

Although a lack of representativeness is likely an issue in many observation studies, most relegate these issues to a single sentence in the discussion of limitations. Likewise, SES both as it relates to selection, and as a potential modifier of the effect of vaccination, is not generally a focus of VE studies.

CHAPTER 3: PARENT STUDY: PCV13 VACCINE EFFECTIVENESS STUDY

Active Bacterial Core surveillance system and PCV13 VE extended area surveillance

Since 1998, CDC has conducted active, population- and laboratory-based surveillance for invasive bacterial infections through the Active Bacterial Core surveillance (ABCs) system.[104] In addition to IPD, the ABCs system includes Group A *Streptococcus*, Group B *Streptococcus*, *Haemophilus influenzae*, Methicillin-resistant *Staphlococcus aureaus*, and *Neisseria meningitides*. The surveillance populations differ slightly for each pathogen and year, but the total population under surveillance for IPD in 1998 was 17.4 million persons in eight states and had expanded to 30.4 million persons in 10 states by 2012. (Figure 1.1)[105]

Cases of IPD are defined as persons with *Streptococcus pneumoniae* isolated from a normally sterile site, such as blood or cerebrospinal fluid, who are residents of the surveillance areas. ABCs collects standard demographic, disease course, and basic risk factor data on all cases (see Appendix 2). Additionally, isolates are sent to CDC for antimicrobial resistance testing and serotyping.

The high burden of invasive disease, particularly an increase in antimicrobial resistant disease, identified through ABCs was one of the factors leading to the development of PCV7 in the late 1990s.[2] When PCV7 was introduced in the US in 2000, CDC conducted a matched case-control VE study within ABCs surveillance sites.[42] This study formed the basis for the PCV13 VE Evaluation, which began in May 2010, two months after the US release of PCV13.[73] Due to relatively lower case counts in 2010 as compared to 2000, four additional areas were added to the catchment area for the PCV13 VE Evaluation: Los Angeles County (713,000 children <5 years of age), New York City (518,000 children <5 years of age), 18 additional counties in upstate New York (66,000 children <5 years of age), and all members of Intermountain Health Care in Utah (213,000 children <5 years of age).

Throughout ABCs and the PCV13 VE Study extended areas, site personnel routinely contact hospitals and laboratories that serve residents of the surveillance catchment areas. Case and isolate lists are requested and periodic audits ensure that hospitals and laboratories report all cases.

IPD case definition

Eligible cases for the PCV13 VE Study were children aged 2-59 months of age who had *S. pneumoniae* isolated from a normally sterile site, who resided in one of the surveillance areas on the date their culture was obtained, and had a pneumococcal serotype available.

Case enrollment procedures

Enrollment began on May 1, 2010 and went through May 31, 2014. Once a case was identified by surveillance, cases were reviewed for study eligibility. Cases meeting the study IPD definition were excluded if complete vaccination histories could not be obtained (Figure 3.2). Additionally, if the case had a previous diagnosis of IPD or the parent refused consent, the child was excluded. For eligible cases, site personnel conducted routine medical record reviews to complete the ABCs Case Report Form (CRF, Appendix 2). Site personnel then attempted to contact the parents/guardians of cases to obtain consent and conduct a detailed interview, which included questions about IPD risk factors such as smoking exposure and crowding. Parents were also asked to provide contact information for any providers who had administered vaccines to their child. Sites would then attempt to contact all providers and have them complete the Medical and Vaccine History Form (Appendix 3), which includes a detailed vaccination history and questions about underlying health conditions and recent antibiotic exposure.

Control matching

Controls were individually matched to cases on the basis of age and zip code, with a goal of matching four controls to each case. After enrolling a case, site personnel obtained a list of controls born within 14 days of the case and whose mother's residence at the time of birth was within the same zip

code as the case's residence at time of culture. Controls were then contacted, beginning with those controls born closest to the case and moving further out on each side of the case's birth date. If insufficient numbers of controls were available within the case's zip code, sites selected controls from the zip codes immediately contiguous to the case. Parents of controls were asked for consent, and then interviewed in the same manner as case parents. Case status could not be blinded due to logistic requirements at the surveillance sites.

Exposure assessment – provider follow-up and IIS

The main exposure for the PCV13 VE Evaluation was dichotomous – whether or not the child received one or more doses of PCV13. Secondary analyses considered two, three, and four doses, as well as mixed schedules (e.g., first two doses of PCV7, followed by two doses of PCV13). For cases (and matched controls) with IPD caused by one of the serotypes included in PCV7, vaccination with either PCV7 or PCV13 was included in the vaccine history.

Vaccination histories came primarily from two sources: medical records and state immunization information systems (IIS). Doses were considered valid if administered more than four weeks after the previous dose. Any doses administered before March 15, 2010 were automatically considered to be PCV7 as PCV13 had not yet shipped in the US. For doses administered on or after March 15, 2010, researchers used lot numbers provided by Pfizer to determine if a dose was PCV7 or PCV13.

Calculation of vaccine effectiveness and primary objective

The primary objective for the parent study was estimation of the VE of one or more doses of PCV13 against IPD caused by one of the serotypes included in the vaccine.[106] VE was calculated as (1-matched odds ratio)*100.[42] The matched odds ratio (mOR) was calculated from a conditional logistic regression model. Because IPD is a rare disease and controls were chosen using incidence density sampling, the odds ratio should estimate the rate ratio.[107]

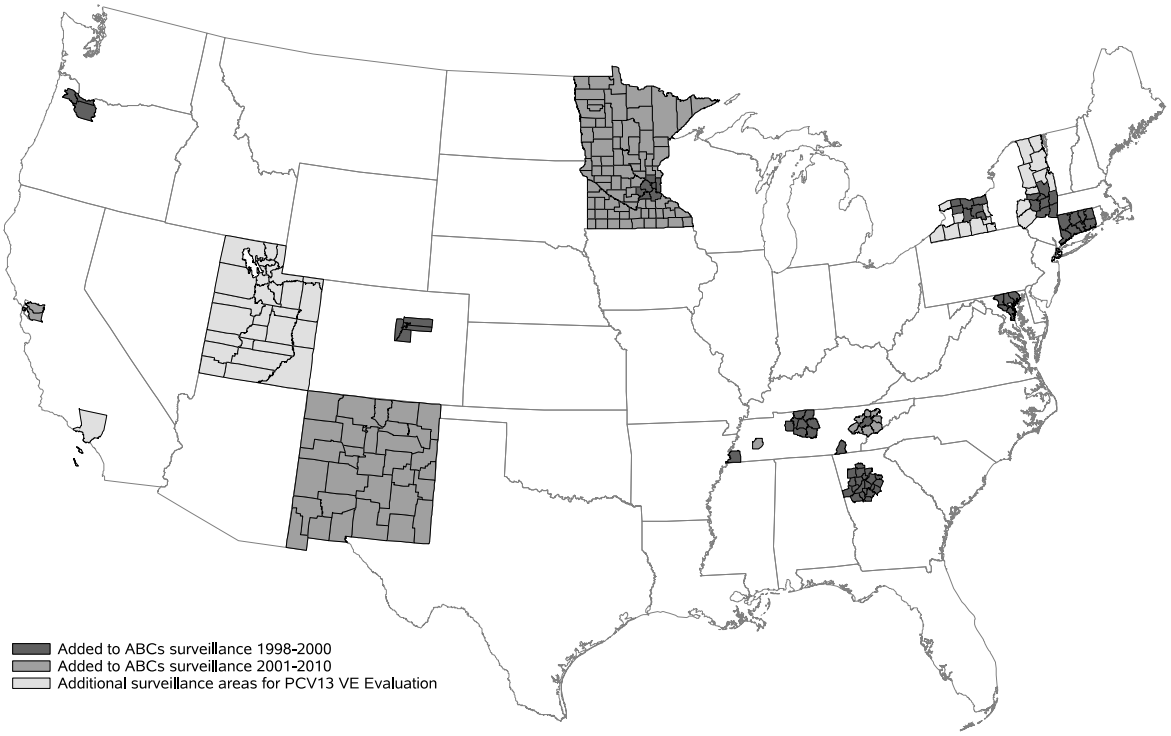


Figure 3.1. Map of surveillance area for PCV13 VE Study, by year added to ABCs

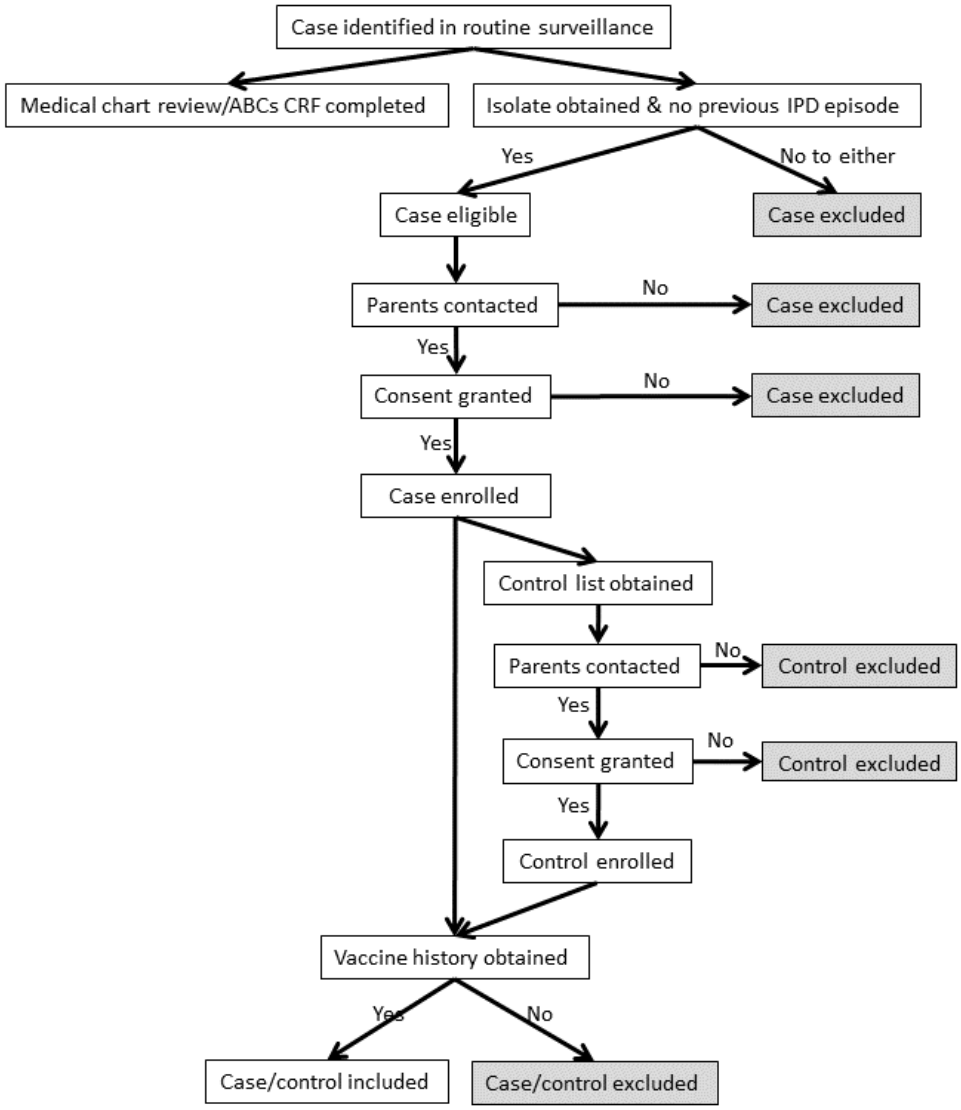


Figure 3.2. Enrollment procedures for cases and controls in the PCV13 VE Study

CHAPTER 4: RESEARCH METHODS

Population included in SES analysis

Two surveillance sites, Colorado and Maryland, were not able to obtain individual-level SES variables and cases (and matched controls) from these two sites and were therefore excluded for purposes of the SES analysis (although they were included in the parent study analysis). Complete enrollment numbers for purposes of the SES analysis are outlined in Figure 4.1.

Covariate assessment/definitions

Covariates were collected from a number of sources, although not all sources were available for every child. A list of the variables, sources, and availability is in Table 4.1.

ABCs case report form

Certain SES-related characteristics are routinely collected on the ABCs CRF[108] from medical record reviews and are available for all cases, regardless of enrollment in the PCV13 VE study. Some (e.g., insurance status at IPD culture, underlying condition status) are related to SES or are potentially useful for predicting enrollment.

Outcome

Outcome of the IPD episode was a binary variable, survived or died, and was collected through medical record review or subsequent matching with state death records.

Severity

Severity of IPD episode was calculated based on hospitalization and intensive care unit (ICU) admittance status. Cases who were admitted to the ICU were given a score of 3, the most severe. Cases

who were admitted to the hospital, but not the ICU, were given a score of 2. Cases who were not hospitalized were given a score of 1.

Underlying condition status

Chronic underlying chronic conditions of interest captured on the ABCs CRF are: diabetes, heart failure, chronic lung disease, cochlear implant, and neuromuscular disorder. Immunocompromising conditions captured on the ABCs CRF are: immunoglobulin deficiency, sickle cell anemia, congenital or acquired asplenia, leukemia, lymphoma, immunosuppressive therapy, complement deficiency, HIV/AIDS, cerebrospinal fluid leak, dialysis, nephrotic syndrome, solid organ malignancy, and bone marrow or organ transplant. Presence of either a chronic or immunocompromising condition resulted in the case being classified as having an underlying condition.[25]

Insurance status at IPD

The following insurance status choices are available on the ABCs CRF: private, Medicaid/state assistance, Medicare, military, Indian Health Service, incarcerated, uninsured, other (with a specify field). For the purposes of this analysis, Medicare, Medicaid, Indian Health Service, and incarcerated were grouped as "public/state" insurance.

Race/ethnicity

Races captured on the ABCs CRF are: white, black, American Indian/Alaska Native, Asian, Native Hawaiian/Other Pacific Islander, and unknown. Ethnicity is captured as either Hispanic or non-Hispanic. For purposes of this analysis, we created joint race/ethnicity categories: white, non-Hispanic, black non-Hispanic, other (including American Indian/Alaska Native, Asian, and Native Hawaiian/Other Pacific Islander) non-Hispanic, and Hispanic.

Age

Child's age was calculated in months as culture date – birth date.

Birth certificate data

We collected demographic and prenatal care data from the birth certificates of enrolled and unenrolled cases and controls. Variables collected were chosen for both their reported association with SES and for their completeness. For example, we collected maternal race, ethnicity, and education level, but not paternal race ethnicity, and education. While these factors have been shown to be related to SES for both parents, the maternal characteristics were close to 100% complete in most study sites, while the paternal characteristics were frequently closer to 90% complete, if collected at all. Although income is not available on the birth certificate, demographic characteristics, such as race/ethnicity and education, correlate with income and can therefore be used to understand an individual's SES.

Payment status for birth

We collected the principal source of payment/insurance status for the delivery, which is divided into four categories on the standard birth certificate: private insurance, Medicaid, self-pay, and other.[109]

Adequacy of Prenatal Care Utilization Index

We also collected number and timing of prenatal care visits and gestational age, which will allow us to estimate the Kotelchuck Index, also known as the Adequacy of Prenatal Care Utilization (APNCU) index.[110,111] The APNCU is a four-category variable which summarizes both how early in the pregnancy a woman initiated prenatal care and how many total prenatal care visits the woman received. First, timing of initiation of prenatal care is divided into four categories (1-2 months gestation, 3-4 months gestation, 5-6 months gestation, and 7-9 months gestation). Second, the expected number of prenatal care visits is calculated based on gestational age using an algorithm from the American College of Obstetrics and Gynecology (ACOG): one visit per month through 28 weeks' gestation, one visit every two weeks through 36 weeks' gestation, and one visit each week thereafter.[111] The ratio of the actual number of prenatal care visits to the expected number is then calculated. For example, a woman who initiated care

during the first two months of gestation and had 100% of the expected prenatal care visits would fall into the "adequate" category. A woman who did not initiate care until the last trimester would fall into the inadequate category, regardless of the number of prenatal care visits she received.

Maternal race/ethnicity

Maternal race/ethnicity was captured on birth certificates with a minimum of the following categories: white, black, American Indian/Alaska Native, Asian, Native Hawaiian/Other Pacific Islander, and unknown. Ethnicity was captured as Hispanic or non-Hispanic. As with the CRF race/ethnicity variable, we created joint race/ethnicity categories: white, non-Hispanic, black non-Hispanic, other (including American Indian/Alaska Native, Asian, and Native Hawaiian/Other Pacific Islander) non-Hispanic, and Hispanic.

Maternal education

Maternal education was divided into four categories: less than high school, high school equivalent, some college, or college graduate and above.

Parent interview/provider follow-up

For enrolled cases and controls, site surveillance personnel conducted a detailed parent/guardian interview, assessing demographics, IPD behavioral risk factors, and vaccine and medical history information. Parents/guardians were also asked to provide contact information for vaccine providers, who were then contacted to obtain vaccine histories.

Income

Parents/guardians were asked to report total household income before taxes, which was collected as a categorical variable with five categories in addition to "Refused" and "Don't know." The income categories were: $0-$15,000, >$15,000-$30,000, >$30,000 to $45,000, >$45,000 to $60,000, and >$60,000.

Caregiver education

Parents/guardians were asked what best describes the highest level of education the primary caregiver completed. Categories corresponded closely with categories on the standard US Birth Certificate: No high school, some high school, high school graduate/GED, technical school, some college, college graduate, postgraduate/professional, and don't know/refused. These categories were combined to match the categories for the birth certificates as follows: technical school was combined with some college. College graduate and postgraduate/professional were combined.

Race/ethnicity

Race and ethnicity were categorized in an identical way to the CRF.

Smoking exposure

During the interview, parents/guardians were asked a series of questions about the child's exposure to secondhand smoke during the 30-day reference period (the 30 days before culture for cases or the corresponding month for controls). These questions included whether or not people living or staying in the house smoked, how many cigarettes were smoked per day, and whether the child's primary or secondary caregiver smoked in a room or car with the child. Given the difficulties associated with accurately assessing amount of smoking and the potential for recall bias, we dichotomized smoking exposure (any vs. none).

Crowding

Parents/guardians were asked about the type of house (e.g., single family, apartment, etc.), the number of bedrooms, and the number of individuals living in the house during the reference period. A child was considered to have lived in a crowded household during the exposure period if the number of people per bedroom was greater than two.

Daycare attendance

Parents/guardians were also asked whether or not their child attended daycare and, if so, what type of daycare (e.g., daycare center or home daycare), for how many hours a week, for how many months before the reference period, and how many children were in the same classroom or home daycare. As with smoking exposure, daycare attendance was dichotomized (any vs. none) to avoid potential recall bias issues.

Breastfeeding

Breastfeeding was also a dichotomous variable, counted if the parent/guardian reported any history of breastfeeding.

Insurance coverage

Parents/guardians were asked to report insurance coverage, which was categorized, as follows: private, Medicaid/state assistance, Medicare, uninsured, other (with a specify field).

Influenza vaccination

Influenza vaccination (including both seasonal and H1N1) status was collected as part of the provider follow-up. We counted any vaccination received in the six months before the case's culture date.

Influenza infection

Influenza infection within the 30 days before the case's culture date was also collected via provider report.

Underlying conditions

As part of the provider follow-up, we asked providers to report if a child had even been diagnosed with the same underlying conditions requested on the CRF. As with the CRF data, we

considered a child to have an underlying condition if they had either a chronic or immunocompromising condition. Both the parent interview and the provider forms included questions about these underlying conditions. When a discrepancy between the parent and provider existed, study staff used the provider's report as this would be more likely to be complete and accurate.

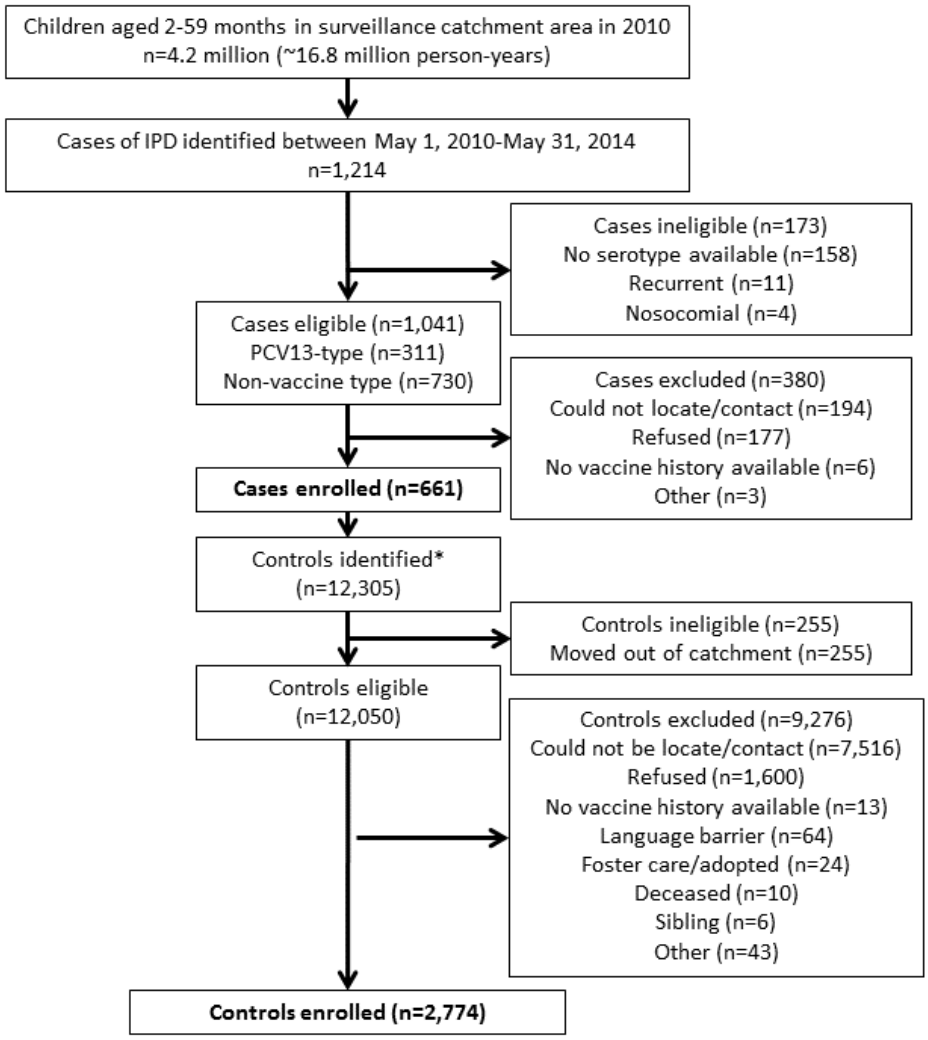
American Community Survey

Cases and controls were geocoded, regardless of enrollment status. Geocoding was conducted using ArcGIS software, with a standard geocoding protocol. Census tract was merged with data from the American Community Survey (ACS). The ACS is an annual survey conducted by the US Census Bureau. Surveys are distributed to approximately 250,000 randomly selected households each month, totaling three million households a year. The ACS replaced the long form of the decennial census, which was last administered in 2000. The first full ACS was conducted in 2005 and, by 2010, the ACS was able to produce estimates for areas of all population sizes, using information collected from January 2005-December 2009.[106] We used the five year estimates from 2009-2013.

The ACS includes demographic information, such as the distribution of racial and ethnic groups with a census tract, the number and percent of individuals indicating multiple races, age and gender distributions, distribution of educational level and primary and secondary languages, and percent of households with a female head of household. It also includes number and percent of poor, as well as distribution of income brackets and median household income, percent with private, public, and no insurance, distribution of different job classes (e.g., sales, professional, executive-level, etc.), employment rates and number of households on public assistance, number of renter vs. owner-occupied housing units, percent who own a car, rental or mortgage price, and distribution of crowding. Detailed information on variables used in each aim will be provided below, as appropriate.

Ethics approval

Both the parent study and the SES analyses were approved as non-research by the Institutional Review Board at CDC as they was part of an assessment of the implementation of a public health intervention. The SES analyses were also approved by the Institutional Review Board of the University of North Carolina, Chapel Hill, USA.



* Includes controls up to and including the last enrolled controls. Children that were not needed because 4 controls had already been enrolled for the case, are excluded.

Figure 4.1. Enrollment numbers for SES analysis

Table 4.1. Variables, sources of information, and availability by case and enrollment status

Variables	Source	Cases		Controls	
		Enrolled	Unenrolled	Enrolled	Unenrolled
IPD severity, outcome, underlying condition status, insurance status at IPD, race/ethnicity, gender, age	ABCs CRF	✓	✓		
Income, caregiver education, race/ethnicity, smoking exposure, crowding, daycare attendance, breastfeeding, insurance status at IPD, influenza vaccination, influenza infection, underlying conditions	Parent interview & provider follow-up	✓		✓	
Payment status for birth, number and timing of prenatal care visits, gestational age, maternal education, maternal race/ethnicity	Birth certificate	✓	✓	✓	✓
Numerous	ACS (geocoding)	✓	✓	✓	✓

Table 4.2. American Community Survey variables included in prediction models for enrollment.

Category	Variable name	ACS Table
Demographics	White, non-Hispanic	0005
	Hispanic, any race	0005
	Citizen, born in US	0009
	Immigrant, entered US in 2010 or later	0010
	Born in North America	0010
	Born in Europe	0010
	Born in Asia	0010
	Born in Africa	0010
	Born in Latin America	0010
	Child living in 2-parent household	0012
	Child is native-born	0012
	Citizen, naturalized in 2010 or later	0012
	Born in same state as current residence	0013
	Speaks English well	0014
	Never married	0014
	Lacks high school diploma (of those ≥25 years)	0043
	Veteran (of those ≥25 years)	0073
	Private insurance	0116
Wealth	Median income	0015
	Living below 100% of poverty line	0015
	Lived in same house 1 year ago	0016
	Lived abroad 1 year ago	0016
	Received public assistance in past 12 months	0034
	Lived in poverty in past 12 months	0048
	Earns < $30,000 per year	0059
	Earns ≥$60,000 per year	0059
	Gini Index	0063
	Per capita income in past 12 months	0065
Work	Drives alone to work, workers ≥16 years of age	0023
	Workplace in same county as residence	0023
	Night shift	0023
	Commute time ≥ 60 minutes	0023
	No vehicle available for commute	0023
	Disabled	0058
	Mean hours worked	0078
	Unemployed	0079
	Working class (includes workers in sales/office, service, production, transportation, moving, construction, and maintenance)	0080
Household	Owner occupied housing units	0019
	Size of household	0028
	Number of workers in household	0028
	Female head of household in household with children <18 years, no husband	0034
	2-parent household, children <18 years	0034
	Unmarried partner living in household	0034
	Child lives with biological parent	0034

Household with no children < 18 years of age	0060
Householder < 25 years of age	0060
Householder ≥ 65 years of age	0060
Occupied housing units	0103
Housing units for rent	0103
Housing units for migrant workers	0103
House with > 1.0 occupant per room	0103
Median rooms in housing units	0104
Median year structure built	0105
Median year householder moved into unit	0105
Housing unit with complete plumbing facilities	0105
Housing unit with complete kitchen facilities	0105
Median rent	0106
Median rent as percent of income	0106
Median monthly housing costs	0108
Owner-occupied homes worth < $300,000	0107

CHAPTER 5: ANALYTIC METHODS

Aim 1

Evaluate whether zip code matching results in a similar distribution of SES in cases and controls; estimate the amount of residual confounding and appropriately adjust the main VE estimate.

Identification of confounders

Confounders for adjustment were identified using a directed acyclic graph (DAG). DAGitty.net (version 2.2) software was used to identify minimally sufficient confounding subsets (Figure 5.1).[112-115] Zip code matching ensured enrolled cases and eligible controls had similar aggregate SES at the zip code level, but not at the census tract or at the individual level. Our DAG, however, included these census tract and individual SES measures to determine if zip code was an adequate proxy. Multiple minimally sufficient confounding subsets were identified, with substantial overlap between them. We selected one minimally sufficient subset for our primary analysis based on the completeness of the variables included (i.e., least number of matched pairs dropped due to missing data).

Comparison groups

We explored the potential for residual confounding in two ways, both of which explored differences between enrolled cases and a group of controls. First, using available data on all enrolled cases and all *eligible* controls (regardless of enrollment) we assessed whether differences existed between the groups. If no differences exist between enrolled cases and *eligible* controls, this would indicate zip code matching *theoretically* controlled for measured confounders. In other words, if the two groups are similar, this indicates that, in the absence of selection issues, matching on zip code resulted in controls that were exchangeable with cases, with respect to measured SES characteristics. If however,

differences between enrolled cases and eligible controls exist, this would indicate zip code matching had failed to control for individual SES.

Second, we restricted our analysis only to enrolled cases and *enrolled* controls. This allowed us to assess how selection issues, such as failure to locate or enroll controls in the study, affected our final study population. The meaning of the results of this step is dependent on the results of the first step. If enrolled cases were similar to eligible controls *and* enrolled controls, this would indicate that zip code matching was successful (i.e., enrolled cases were similar to eligible controls) and there was no selection bias. Our study population would therefore be exchangeable with respect to measured SES. If, however, enrolled cases were similar to eligible controls, but not to enrolled controls, this would indicate selection bias. If zip code matching failed (enrolled cases were not similar to eligible controls), any similarity between enrolled cases and enrolled controls would likely be due to chance.

Control for residual confounding

We used the same primary outcome, VE of one or more doses of PCV13 against one of the serotypes included in the vaccine, as the parent study. VE against serotypes not included in PCV13 was also calculated as a negative control. Assuming no cross-reactivity with vaccine serotypes, VE against non-vaccine types should by zero, so an estimate significantly different from zero would indicate a problem with the analysis.

Unadjusted VE was calculated via conditional logistic regression with only vaccination status included in the model. Adjusted (full) models were run, which included vaccination status and each of the minimally sufficient confounding subsets identified from our DAG. Since each of the full models included numerous variables, any one of which could be missing for a particular child, a reduction in sample size was expected compared to the unadjusted model. We therefore chose as our primary minimally sufficient subset the model with the most discordant pairs.

We considered an absolute change in the VE of 5% between the full and unadjusted models (e.g., a change from 95% to 90% effectiveness) to be an indication of meaningful change (i.e., meaningful residual confounding). Experience with tracking individual PCV7 serotypes after PCV7 was introduced indicated that a difference of 7% in the VE could result in large differences in post-vaccine cases. For example, serotypes 14 and 19F, both of which were included in PCV7, had pre-vaccine rates of 63.3 and 21.8 cases per 100,000, respectively, in the US. VE for serotype 14 was estimated as 94% (95% CI: 81-98%) and VE for serotype 19F was estimated as 87% (95% CI: 65-95%). By 2004, only 14 breakthrough cases of serotype 14 were identified in ABCs, compared with 45 cases of serotype 19F.[34,42,66] This difference, especially given the much higher rates of serotype 14 in the pre-vaccine era, indicated a difference of 7% would be important, so to be conservative for purposes of the current analysis, we considered a slightly lower difference to be meaningful.

Aim 2

Assess generalizability in a population-based case-control study by:

(a) assessing whether enrollment rates differ with respect to SES, and

(b) determining whether SES is an effect measure modifier of the estimated vaccine effectiveness.

External validity, or generalizability, is the measure of how well results from a study pertain to individuals in the target population.[116-118] Study results may fail to generalize to a target population under several circumstances.[6,116] A key reason that studies fail to be generalizable is when both of the following criteria are met: (a) selection/enrollment into the study is differential with respect to variable X (i.e., the study population is not representative of the source population with respect to variable X, which is usually a potential confounder), and (b) variable X modifies the exposure/outcome relationship under study.[6,116]

It is important to note that in matched case-control studies, controls need not be representative of the source population with respect to the matching factor(s). In fact, if matching is done correctly,

controls should resemble enrolled cases on the matching factor. This analysis, therefore, focuses on representativeness of enrolled cases with respect to eligible cases and leaves questions concerning controls for other analyses.

Socioeconomic Position Index

In addition to the ACS variables used in Aim 1, we calculated a composite index for SES based on work done by the Public Health Disparities Geocoding Project.[24] The Socioeconomic Position (SEP) Index includes ACS variables measuring working class, unemployment, poverty, education, home prices, and median family income. Together these are meant to capture the major SES constructs of wealth, education, and occupation. The SEP Index is created by first calculating the percent of each of the ACS variables included. Then, a standardized z-score is calculated for each variable:

$$Z_{ij} = (X_{ij} - m_j)/s_j$$

Where X_{ij} is the value of variable *j* for area *i*, m_j is the mean of variable *j* across all areas, and s_j is the standard deviation of variable *j* over all areas.[24] The sum of the Z scores for each variable is then the value of the composite index.

Choosing variables for modification assessment

Because we had access to many SES-related variables through the case report form, birth certificates, and ACS, it was not practical to assess every variable for effect measure modification (EMM). Instead, we chose variables to assess for EMM in two ways. First, we decided *a priori* to assess as modifiers the SEP Index and the any individual-level variables which had p-values <0.2 in the exploratory analyses for differences between enrolled and unenrolled. P-values were calculated for continuous and categorical predictors via two-sided Wilcoxon rank sum test and Fisher's exact test, respectively.

Second, to ensure we did not miss any variables for the EMM assessment that may be strong modifiers *and* related to representativeness, we used predictive modeling to select additional ACS variables to assess for EMM.

Predictive model

Many of the available ACS variables measured similar metrics (both to other ACS variables and to the individual variables); therefore, we fit a series of models to narrow the selection. All models fit were logistic regression, using backward selection and retaining variables with a p-value of <0.2.

Individual model

We first fit a single predictive model including all individual-level variables: severity, outcome, underlying condition status, child's race/ethnicity, insurance status at IPD, mother's education, mother's race/ethnicity, APNCU Index, source of payment for birth, and interaction terms between child's race ethnicity and maternal education and between child's race/ethnicity and APNCU Index. The model was required to retain severity, outcome, underlying condition status, child's race/ethnicity, and insurance status at IPD.

ACS models

We divided the ACS variables into four categories – demographics, wealth, work, and household characteristics – for inclusion in separate models to reduce the number of ACS variables used. Four models were run, including all individual-level variables retained from the first model and each category of the ACS variables. Lastly, we fit the final predictive model, which included all the individual-level variables and interaction terms (regardless of whether they were retained in the first model) and any variables retained in the ACS models. As with the first model, we required the model to retain the case report form variables.

Assessment of effect measure modification

Any variables included in the final model predicting enrollment were assessed as potential modifiers, in addition to the SEP Index and individual-level variables chosen *a priori*. All the individual-level variables were already categorical. We did not include variables (i.e., severity of IPD, outcome status) that are not applicable to controls. Additionally, variables obtained from the case report form for cases (child's race/ethnicity, underlying condition status, and insurance status at IPD) were not available for controls (since no case report form is completed for controls); however, these variables were collected as part of the parent interview/provider follow-up for enrolled children and could therefore be assessed for EMM. The ACS variables were continuous, so we dichotomized them at the median for the EMM analysis (Table 5.1). To assess variables as modifiers, we fit conditional logistic regression models for each variable with IPD caused by one of the 13 serotypes included in the vaccine as the outcome, receipt of one or more doses of PCV13 as the exposure, and an interaction term between each variable of interested and PCV13 receipt.[119] Because power to assess interaction terms is reduced, we used a p-value <0.2 as our cutoff for the likelihood ratio chi-square value for the interaction term and then did a Bonferroni adjustment by the total number of variables assessed for EMM to account for multiple comparisons. Therefore, $\alpha<0.015$ for the likelihood ratio test was considered an indication of modification and cause for concern about generalizability.

Figure 5.1. Causal diagram showing the relationship between vaccination status, IPD, and confounders.

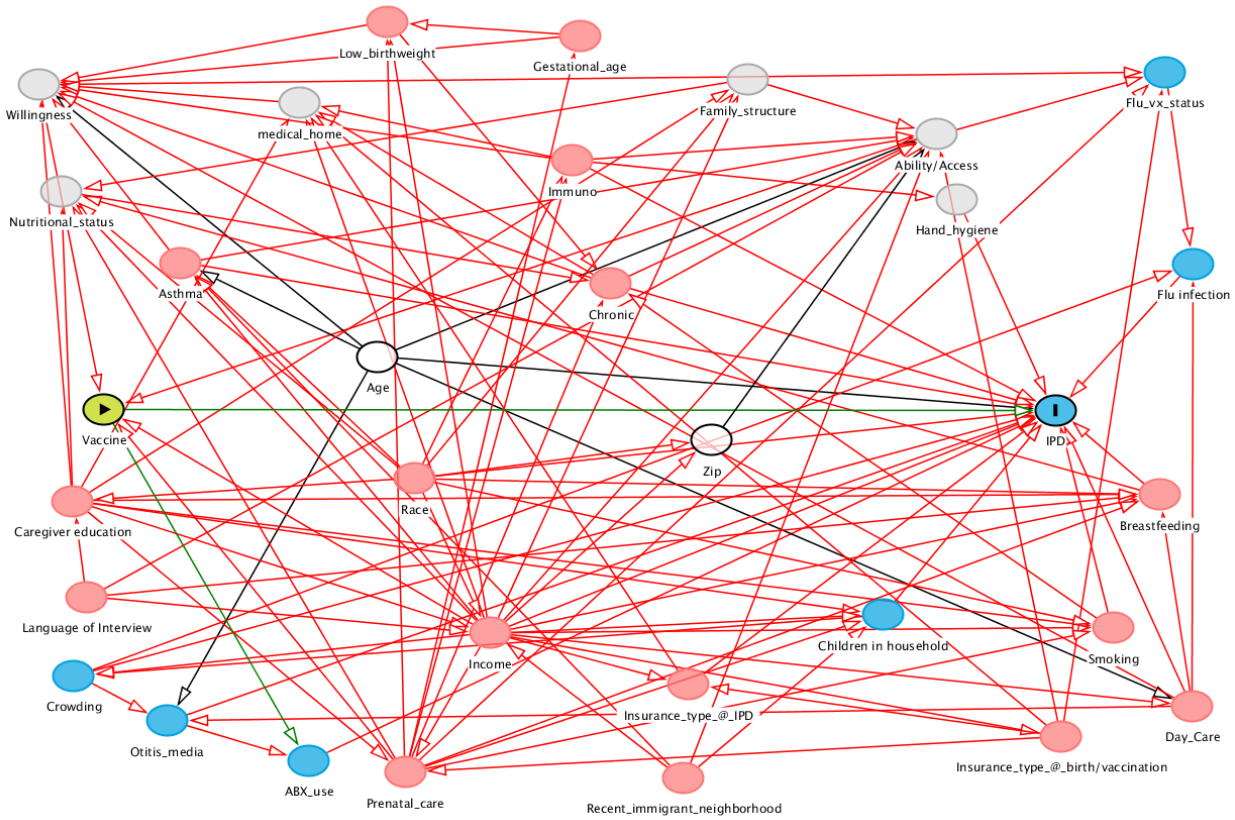


Table 5.1. Operational definitions of ACS variables included in final prediction model.

Category	Variable name	Variable description	Cut point (median) for EMM analysis
Demographic	White race	Percent of individuals identifying as white, non-Hispanic	48.3%
Wealth	Household income <$30,000	Household income in the past 12 months (in 2013 inflation-adjusted dollars)	28.0%
	Household income ≥$60,000	Household income in the past 12 months (in 2013 inflation-adjusted dollars)	40.4%
Work	Working class	Percent of individuals reporting jobs in sales, service (except protective), production, transportation, and material moving, natural resources, construction and maintenance	61.4%
	Disabled	Percent of individuals with a disability	10.3%
	Mean usual hours worked	Mean number of hours worked annually	37.9 hours
Household	Units occupied	Percent of occupied housing units	92.3%

CHAPTER 6: BIAS WITH RESPECT TO SOCIOECONOMIC STATUS: A CLOSER LOOK AT ZIP CODE MATCHING IN A PNEUMOCOCCAL VACCINE EFFECTIVENESS STUDY

Overview

In 2010, 13-valent pneumococcal conjugate vaccine (PCV13) was introduced in the US for prevention of invasive pneumococcal disease in children. Individual-level socioeconomic status (SES) is a potential confounder of the estimated effectiveness of PCV13 and is often controlled for in observational studies using zip code as a proxy. We assessed the utility of zip code matching for control of SES in a post-licensure evaluation of the effectiveness of PCV13 (calculated as [1-matched odds ratio]*100). We used a directed acyclic graph to identify subsets of confounders and collected SES variables from birth certificates, geocoding, a parent interview, and follow-up with medical providers. Cases tended to be more affluent than eligible controls (for example, 48.3% of cases had private insurance vs. 44.6% of eligible controls), but less affluent than enrolled controls (52.9% of whom had private insurance). Control of confounding subsets, however, did not result in a meaningful change in estimated vaccine effectiveness (original estimate: 85.1%, 95% CI 74.8-91.9%; adjusted estimate: 82.5%, 95% CI 65.6-91.1%). In the context of a post-licensure vaccine effectiveness study, zip code appears to be an adequate, though not perfect, proxy for individual SES.

Introduction

Socioeconomic status (SES) is increasingly understood to be a fundamental cause of disease due to the persistent association between low SES and poor health outcomes, despite substantial advances in prevention and treatment of disease.[7,86,120] This association is concerning, especially in the US, where substantial differences in access to healthcare, nutritious foods, and physical activity exist between more and less affluent individuals and neighborhoods.[7,15-17,26,75,86,120] While no single definition of SES is

universally accepted, individual-level SES is generally measured as a combination of income, education, and occupation, which in turn provide surrogate measures of resources, prestige, knowledge, and power.[7,8,12,115,120-122] Race, ethnicity, and health insurance status may also be considered markers of SES, because these factors provide insights into access to resources, knowledge and power, and are frequently easier to obtain for research than income or education levels.[13,16,84-87]

When SES is not the exposure of interest and instead measured to control for potential confounding of an exposure-disease relationship, most researchers will simply match on SES or control for SES during analysis, depending on the study design. Whether SES is examined as an exposure or confounder, it is paramount that the variable serve as an accurate surrogate of the construct that one intends to measure. For example, if neighborhood-level income is being used as a surrogate for individual level income level, one must be confident that this cross-level inference is valid.[123,124]

Because SES is often clustered geographically and individual-level data can be difficult to obtain, researchers often assess SES ecologically, for example by using neighborhood-level measures, such as prevalence of poverty by zip code.[125-127] For example, research conducted using cases identified through disease surveillance systems frequently uses zip code as a proxy for individual SES. Surveillance systems generally incorporate addresses, but rarely include characteristics such as personal or household income, educational attainment, or occupation, which require follow-up with individual cases.[12,126] Using zip code is a relatively easy way to measure SES, but requires the assumption that zip code is an adequate proxy for individual or household level SES.[124,128]

One type of study in which potential confounding by SES is of concern is post-licensure vaccine effectiveness studies, frequently conducted after a vaccine is introduced and typically using a case-control study design. Because both the exposure (vaccination) and outcome (infectious disease) may be associated with SES, the potential for confounding may exist and researchers therefore frequently match on zip code.[17-22,42,44,75,129,130] Zip code matching, however, only ensures that eligible controls are

similar to enrolled cases at the zip code level. Differences may remain between the groups at smaller area levels (i.e., census tract) or at the individual level. Thus, even after matching on zip code, confounding by individual SES may remain. To date, little research has explored whether matching on zip code provides adequate control for individual SES in vaccine effectiveness studies in the US.[23,42,44,129]

We were concerned about confounding by individual SES in a zip code-matched case-control study of 13-valent pneumococcal conjugate vaccine (PCV13) effectiveness.[119] PCV13 was licensed for use in children in the US in February 2010 and replaced the effective, but more limited, 7-valent vaccine (PCV7).[68,73] SES, including income, educational attainment, and related factors (e.g., asthma, smoking exposure), has been frequently shown to be associated with both vaccination status and risk of IPD and is therefore of concern as a potential confounder.[18,19,21,22,78,79,82] Zip code matching was used to control for SES. The purpose of the present study was to determine whether this approach provided adequate control for confounding at the census tract and individual levels or if additional control of confounding was necessary.

Methods

Enrollment methods

Details of the vaccine effectiveness study and results of the primary analysis have been previously published.[119] Briefly, cases of IPD were identified through the Centers for Disease and Control and Prevention's (CDC) Active Bacterial Core surveillance, an active population- and laboratory-based surveillance system for invasive bacterial diseases in ten sites around the US.[105] Three other sites with similar case identification methods were added to increase numbers of cases: New York City, Los Angeles County, and the State of Utah. Eligible case-children were identified through routine surveillance between May 1, 2010 and May 31, 2014 who were 2-59 months of age with a pneumococcal serotype available.[73] Informed consent was obtained for all enrolled cases and controls. Both the parent study and the current analysis were approved by institutional review boards (IRB) at

CDC and the surveillance sites. The current analysis was also approved by the University of North Carolina, Chapel Hill IRB.

Enrollment procedures for case and controls have been described previously.[119] Briefly, study staff contacted parents/guardians of case and control children via telephone to obtain consent, ascertain information on factors potentially related to disease, and gather contact information for vaccine providers; providers were then asked for detailed medical and vaccine history information.[14,42] Once a case-child was enrolled, staff obtained from local birth registries a list of 20-40 children born in the case-child's zip code within 14 days of the case-child's birth. If four controls could not be enrolled from with a case-child's zip code, additional controls were obtained from adjacent zip codes. Controls were then enrolled in order, starting with the control-child whose birth date was closest to the case and then ranked alphabetically. At least 10 attempts to enroll a control were made at different times of the day and on different days of the week before moving on to the next potential control.

The main analysis excluded children who could not be located, whose parents refused, whose vaccination history could not be verified, who had a recurrent IPD episode (cases only), were in foster care (controls only), had died for any reason (controls only), or were the sibling of a previously enrolled child (controls only), and residents of long-term care facilities. Finally, for the purposes of this analysis, cases and controls from two surveillance sites, Colorado and Maryland, were excluded because individual-level birth certificate data were not available to investigators.

Identification of confounders

To identify confounders for adjustment in analytic model of vaccine effectiveness, we constructed a directed acyclic graph (DAG).[112,113,115] Potential confounders were identified from past research, and DAGitty.net (version 2.2) software[114] was used to identify minimally sufficient confounding subsets for adjustment. Zip code matching ensured that enrolled cases and eligible controls had similar aggregate SES at the zip code level, but not at the census tract or at the individual level. Our DAG, however,

included these census tract and individual SES measures to determine if zip code was an adequate proxy. Multiple minimally sufficient confounding subsets were identified, with substantial overlap between them. We selected one minimally sufficient subset for our primary analysis based on the completeness of the variables included (i.e., fewest matched pairs dropped due to missing data). In addition to confounders identified by our DAG, we also assessed distributions of other SES-related characteristics available from birth certificates and the US Census Bureau's American Community Survey (ACS) by case-control status.

Values of confounders were identified from three sources. First, we used the parent interview and provider follow-up to obtain information on smoking exposure and daycare attendance (any vs. none in the 30 days before the case's culture date), influenza vaccination or infection within the previous six months, household income, primary caregiver education, insurance status at time of IPD culture, underlying condition status (asthma, chronic lung or heart disease, diabetes, cerebrospinal fluid leak, cochlear implant, sickle cell disease, congenital or acquired asplenia, HIV/AIDS, chronic renal failure, nephrotic syndrome, malignant neoplasm, leukemia, lymphoma, solid organ transplant, congenital immunodeficiency[25]), breastfeeding (ever vs. never), presence of other children in the household, and household crowding (>2 people per room).[14,17,22,79,131] Because only enrolled children were interviewed, these variables were not available for unenrolled controls.

The second source of confounder information was data from birth certificates of enrolled and unenrolled children. These variables included timing of initiation of prenatal care and gestational age (which were used to calculate the Adequacy of Prenatal Care Utilization Index[110,111]), maternal race/ethnicity, maternal education, and insurance status at birth.[109] Prenatal care, while not a typical SES measure, is likely to be related to access to and utilization of health services. Finally, eligible cases and controls were geocoded, allowing linkage with census tract information obtained via the ACS, which includes such neighborhood measures as income, racial/ethnic distribution, and proportion living below

the poverty line, among many others.[106] Of these, residence in a neighborhood with >25% foreign born individuals was included on our DAG.

Comparison groups

We explored the potential for residual confounding in two ways, both of which compared differences between enrolled cases and a group of controls. First, using available data on all enrolled cases and all eligible controls (regardless of enrollment) we assessed whether differences existed between the groups. If no differences existed between enrolled cases and eligible controls, this would indicate zip code matching theoretically controlled for measured confounders. In other words, if the two groups were similar, this indicates that, in the absence of selection issues, matching on zip code resulted in controls who were exchangeable with cases with respect to measured SES characteristics. If, however, differences between enrolled cases and eligible controls existed, this would indicate zip code matching had failed to control for individual-level SES.

Second, we restricted our analysis to enrolled cases and enrolled controls, allowing us to assess how selection issues such as failure to locate or enroll controls in the study affected our final study population. The meaning of the results of this step is dependent on the results of the first step. If enrolled cases were similar to eligible controls and enrolled controls, this would indicate that zip code matching was successful (i.e., enrolled cases were similar to eligible controls) and there was no selection bias. Our study population would therefore be exchangeable with respect to measured SES. If, however, enrolled cases were similar to eligible controls, but not to enrolled controls, this would indicate selection bias. If zip code matching failed (enrolled cases were not similar to eligible controls), any similarity between enrolled cases and enrolled controls would likely be due to chance.

Statistical methods

Descriptive analyses of univariate distributions in cases and controls were assessed for confounders identified in the minimally sufficient confounding subset, as well as related characteristics available from birth certificates and geocoding. Most variables collected from the parent interview, provider follow-up, and birth certificates were categorical in nature and left in this form in the initial analysis. Categories were combined for modeling purposes when sample sizes in individual strata were too low. The results of conditional logistic regression models with enrolled children only, including all confounders identified in the minimally sufficient confounding subsets, were compared to the original model, which included only the matching factors of age and zip code. The exposure was receipt of one or more doses of PCV13 at least 14 days before pneumococcal culture (or the matched case's culture date for controls).

The primary outcome for the parent study was PCV13-type IPD, which was also the focus of the current analysis. We used cases caused by serotypes not included in PCV13 as negative controls. That is, assuming no cross-reactivity with vaccine-types, vaccine effectiveness against non-vaccine types should be zero, so a high (or low) significant estimate would indicate a problem with the methods or analysis. Vaccine effectiveness is calculated as (1 – matched odds ratio)*100% for a rare disease, such as IPD.[42] An absolute difference in the vaccine effectiveness of 5% between the full and original models was considered an indication of meaningful confounding (e.g., a change from 95% to 90% effectiveness).

Results

Enrollment

Of 1,040 eligible cases, we enrolled 661 (63.6%) children. We identified 12,305 potential controls, of whom, 255 were excluded because they had moved out of the surveillance area by the time of the corresponding case's IPD diagnosis and were therefore ineligible for enrollment. Of the 12,050 eligible controls, 2,774 (23.0%) were enrolled. The primary reasons for non-enrollment were an inability

to locate/contact the parent/guardian (7,516, 81.0%) and refusal (1,600, 17.2%). In addition, 160 (1.7%) were not enrolled for other reasons, including the lack of a vaccine history, a language barrier, or being in foster care.

The 661 enrolled cases came from 557 zip codes and 632 census tracts (Table 4.1). Of the 12,050 eligible controls, the majority, 8,690 (72.1%), came from the same zip code as their matched case. However, only 1,250 (10.4%) came from the same census tract as their matched case. A similar pattern was seen among enrolled controls, with, 1,921 (69.3%) coming from the same zip code as their matched case and 271 (9.8%) coming from the same census tract as their matched case.

Differences between enrolled cases and eligible controls

Based on birth certificate data, enrolled cases tended to have slightly more affluent mothers than eligible controls. For cases, 44.1% of mothers had no college education, compared with 49.3% of mothers of mothers of eligible controls (Table 4.2). Additionally, 48.2% of cases had private insurance at birth, compared with 44.9% of controls. Mothers of cases and eligible controls were similarly likely to have had at least adequate prenatal care utilization (70.4% of cases vs. 69.8% of controls).

A similar, though less pronounced, pattern was seen for neighborhood level characteristics. In census tracts of enrolled cases, a median of 15.4% of individuals lived below the poverty level, compared with a median of 16.5% in census tracts of eligible controls. In addition, 31.3% of cases came from census tracts with more than a quarter of the population being foreign born, compared to 34.8% of eligible controls. We did not find substantial differences in median income, crowding, or income inequality (as measured by Gini Index) between cases and eligible controls (Table 4.2).

Differences between enrolled cases and controls

Unlike eligible controls, enrolled controls had a higher SES than enrolled cases. Based on information collected during the parent interview, 53.8% of cases came from households with incomes

above $30,000/year, compared to 62.1% of controls. Less than half (44.4%) of cases had private insurance at the time of IPD diagnosis, vs. 52.7% for controls. Primary caregivers of cases were slightly less likely to have at least some college education (67.3% of cases vs. 70.6% of controls). Enrolled controls were also more likely to have breastfed and less likely to have an underlying condition, have attended daycare, be passively exposed to smoking (Table 4.2). The birth certificate variables showed fewer differences between enrolled cases and controls. The two groups had a similar distribution of prenatal care utilization (70.4% of cases vs. 72.5% of controls with adequate or adequate plus prenatal care utilization), while cases were slightly less likely to have had private health insurance at the time of birth (48.2% of cases vs. 52.9% of controls).

DAG analysis and adjusted models

In the main analysis, the unadjusted vaccine effectiveness against PCV13-type disease was 86.0% (95% CI: 75.5 to 92.3%).[119] Once we excluded the children from Maryland and Colorado, the original estimate (controlling for only the matching variables) was 85.1% (95% CI: 73.8 to 91.9%), similar to that from the main analysis. We identified four minimally sufficient confounding subsets. The subset including age, asthma, breastfeeding, presence of children in the household, underlying condition status, influenza vaccination status, household income, insurance type at IPD diagnosis, race, smoking exposure and zip code had the fewest missing values and was chosen for the primary analysis (Table 3). The adjusted vaccine effectiveness estimate was 83.5% (95% CI: 67.3 to 91.6%). The remaining three subsets yielded estimates of vaccine effectiveness between 81.2 and 83.1%, with 95% CIs ranging from 55.1 to 93.6% (Table 3). None of the vaccine effectiveness point estimates from the adjusted models differed by an absolute value of 5% or more from the original model, so we used the original model as our “final” model. As expected, our negative control (vaccine effectiveness against non-vaccine types) yielded low point estimates, with wide confidence limits, all of which crossed the null value (vaccine effectiveness=0).

Discussion

We assessed the use of zip code matching to control for individual-level SES in a matched case-control study of the vaccine effectiveness of PCV13 in children less than five years of age in the US. We found enrolled cases to be slightly more affluent than eligible controls, but slightly less affluent than enrolled controls, as measured by census tract and individual SES variables from parent interviews, provider follow-up, and birth certificates. Adjustment for these variables, however, did not substantially change our estimate of vaccine effectiveness, indicating that zip code matching was an adequate proxy for individual SES in our study and that our previously-published unadjusted estimates should be valid with respect to individual SES.

We assessed a number of SES-related variables beyond those identified as confounders in our DAG. SES is a general term encompassing numerous aspects of an individual or neighborhood and cannot be perfectly measured by any one or any series of characteristics. The exact mechanism(s) by which SES is related to IPD risk is unknown, but clearly multifaceted (i.e., related to conventional SES measures such as household income and crowding, but also to less conventional measures, such as smoking exposure and asthma). Therefore, the potential for unmeasured confounding could be substantial, so exploring a broader subset of SES characteristics is ideal.

Our finding that enrolled cases were slightly more affluent than eligible controls was expected, given that enrolled cases are the subset of the population of eligible cases we were able to locate and enroll, whereas eligible controls represent the entire area. More affluent individuals may be more likely to have landlines or retain a single telephone number over time (making them easier to reach) and may have increased use of and trust in the medical system (making them more likely to agree to enrollment).[5,9,132] The differences indicate that (as expected) zip code may not be a perfect proxy for individual SES in our population. However, the differences did not have a substantial effect on our estimate of vaccine effectiveness. Thus, zip code may suffice for matching purposes for SES, especially if,

as in this study, data are available to assess differences and adjust for or interpret results appropriately. Our second comparison explored the differences between enrolled cases and enrolled controls, which takes into account both zip code matching and our ability to locate and enroll controls. In this analysis, we found that enrolled cases were slightly less affluent than enrolled controls. This may be because parents of cases were easier to locate (medical records from the IPD episode provide more current contact information) and had an incentive to participate (their child recently had a major illness), and therefore enrolled cases may have been more representative of all eligible cases whereas enrolled controls may have represented only the most affluent of eligible controls who were successfully located and contacted and gave consent for participation.

Differences in both comparisons were smaller when census tracts were compared as opposed to individual-level data (either from the parent interview or birth certificates). This likely reflects the fact that census tract is an ecologic measure and thus represents the average for a geographic area, rather than individual differences. Additionally, there was overlap in census tracts, blunting the differences between groups.

Adjustment for the primary minimally sufficient confounder subset resulted in little change in the vaccine effectiveness point estimate (1.6% absolute change). Similarly, none of the vaccine effectiveness point estimates from the additional confounder subsets identified reached the 5% absolute change we decided a priori to be meaningful. This suggests that our original (unadjusted except for the matching factors) estimate of vaccine effectiveness was not substantially biased – and therefore that traditional zip code matching was adequate for control of individual-level SES. Less than expected confounding by SES may also be due to the success of the Vaccines for Children program, which has operated since 1994 and has reduced immunization coverage disparities in many routine childhood vaccines.[133]

Our study had limitations. We were not able to conduct interviews with unenrolled controls and therefore had to rely on data from geocoding and birth certificates to assess SES. Census tracts, while more granular than zip codes, still provide only a group-level estimate of SES. Census tract income, for example, may not be an adequate proxy for individual income and may be simultaneously measuring the effect of low individual income and living in a poorer neighborhood. Birth certificates, meanwhile, provide individual-level information, but their accuracy can vary by state.[134-136] Additionally, birth certificate variables were not available for cases born outside the state where they lived at the time of their IPD episode. While more information was available for enrolled children, data from parent interviews (i.e., behavioral risk factors) could be subject to recall bias. We attempted to mitigate this by using measurements less prone to poor recall (e.g., any smoking exposure instead of number of cigarettes per day), but this could potentially result in other forms of misclassification.

Our study had a number of strengths, including multiple measures of SES at both the neighborhood- and individual-level from the parent/guardian, birth certificate, and census tract. Because we had access to SES information on unenrolled controls, we were able to assess both the theoretical use of zip code as a proxy for individual SES, as well as effects of selection methods on the real world study population. And while birth certificates and geocoding may not be the ideal way to estimate individual SES, they provide more information on eligible children than is usually available to researchers, especially in such a large surveillance system. Such data can provide insight into the study population and how selection may affect internal validity, as well as potentially helping identify SES-related risk factors for disease.

In summary, we found that, despite some differences between cases and controls, zip code matching achieved its intended purpose and our estimated vaccine effectiveness is internally valid with respect to individual-SES. Our results should be broadly generalizable to other vaccine effectiveness

studies in the US, as well as studies of other health outcomes utilizing similar control identification and participant enrollment methods.

Table 6.1. Number of unique zip codes and census tracts for eligible and enrolled children, by case status and serotype of disease

	Enrolled cases	**Eligible controls**	**Enrolled controls**
Total N	661	12,050	2,774
Unique zip codes	557	1,209	577
Unique Census Tracts	632	4,835	2,126

(a) Parent interview/medical provider

Characteristic	Enrolled cases (n=661)	Eligible controls (n=12,050)	Enrolled controls (n=2,774)
Median age, months (range)	21 (2-59)		21 (2-60)
Asthma, n (%)	128 (19.4)		321 (11.6)
Chronic condition, n (%)	51 (7.7)		32 (1.2)
Immunocompromising condition, n (%)	111 (16.8)		82 (3.0)
Breastfeeding, n (%)			
Ever breastfed	480 (73.2)		2224 (80.4)
Currently breastfed	52 (7.9)		303 (11.0)
Crowding (>2 people per bedroom)	111 (16.8)		414 (15)
Day care attendance, n (%)	313 (47.5)		957 (34.6)
Smoking exposure, n (%)	134 (20.5)		443 (16.1)
Recent influenza infection, n (%)	20 (3.2)		25 (1)
Influenza vaccination in last 6 months, n (%)	184 (27.8)		830 (30)
Household income, n (%)			
≤$15,000	166 (27.9)		474 (18.5)
>$15,000 to ≤$30,000	100 (16.8)		455 (17.7)
>$30,000 to ≤$45,000	53 (8.9)		259 (10.1)
>$45,000 to ≤$60,000	65 (10.9)		286 (11.1)
>$60,000	192 (32.3)		975 (38)
Refused	19 (3.2)		119 (4.6)
Unknown	66		206
Insurance type at IPD, n (%)			
Private	288 (44.4)		1449 (52.7)
Public	344 (53.1)		1227 (44.7)
Uninsured	15 (2.3)		54 (2.0)
Other	0		4 (0.1)
Refused	1 (0.2)		13 (0.5)
Unknown	13		27
Race/ethnicity, n (%)			
White, non-Hispanic	259 (39.4)		1368 (49.4)
Black, non-Hispanic	165 (25.1)		469 (16.9)
Hispanic	64 (9.7)		191 (6.9)
Other, non-Hispanic	169 (25.7)		739 (26.7)
Unknown	4		7
Primary caregiver education level, n (%)			
Less than high school	78 (12.0)		289 (10.6)
High school equivalent	134 (20.7)		515 (18.8)
Some college	193 (29.8)		661 (24.2)
College degree or more	243 (37.5)		1272 (46.5)
Unknown	13		37

(b) Birth certificate

Characteristic	Enrolled cases (n=661)	Eligible controls (n=12,050)	Enrolled controls (n=2,774)
Maternal race/ethnicity, n (%)			
White, non-Hispanic	268 (46.7)	5055 (43.2)	1488 (54.7)
Black, non-Hispanic	127 (22.1)	2257 (19.3)	420 (15.4)
Hispanic	52 (9.1)	1072 (9.2)	186 (6.8)
Other, non-Hispanic	127 (22.1)	3312 (28.3)	628 (23.1)
Unknown	87	354	52
Maternal education level, n (%)			
Less than high school	110 (20.2)	2516 (22.9)	420 (16.8)
High school equivalent	130 (23.9)	2905 (26.4)	529 (21.2)
Some college	146 (26.8)	2851 (26.0)	690 (27.6)
College degree or more	158 (29.0)	2714 (24.7)	857 (34.3)
Unknown	117	1064	278
Source of payment for birth, n (%)			
Private	228 (48.2)	4238 (44.9)	1120 (52.9)
Public/state	223 (47.1)	4707 (49.9)	893 (42.1)
Uninsured	8 (1.7)	232 (2.5)	41 (1.9)
Other	14 (3.0)	264 (2.8)	65 (3.1)
Unknown	188	2609	655
Adequacy of Prenatal Care Utilization Index, n (%)			
Adequate Plus	183 (34.9)	3377 (31.4)	814 (33.0)
Adequate	186 (35.5)	4130 (38.4)	972 (39.5)
Intermediate	67 (12.8)	1360 (12.6)	272 (11.0)
Inadequate	88 (16.8)	1888 (17.6)	405 (16.4)
Unknown	137	1295	311

Table 6.2. Characteristics of eligible cases and matched controls. Data come from (a) the parent interview/medical provider, (b) birth certificates, or (c) American Community Survey.

(c) American Community Survey

Characteristic	**Enrolled cases (n=661)**	**Eligible controls (n=12,050)**	**Enrolled controls (n=2,774)**
Not successfully geocoded, n (%)	9 (1.4)	179 (1.5)	20 (0.7)
Median income, n (%)			
≤$15,000	37 (5.7)	796 (6.7)	150 (5.4)
>$15,000 to ≤$30,000	374 (57.4)	6802 (57.4)	1485 (53.9)
>$30,000 to ≤$45,000	191 (29.3)	3315 (28)	849 (30.8)
>$45,000 to ≤$60,000	37 (5.7)	711 (6.0)	204 (7.4)
>$60,000	13 (2.0)	232 (2.0)	66 (2.4)
Crowding, median % (IQR)			
0.50 or less occupants per room	68.2 (52.8,77)	67.1 (50.6,77)	69.3 (55,78.1)
0.51 to 1.00 occupants per room	28.6 (20.9,38.5)	29.7 (22,38.5)	27.5 (19.8,37.4)
1.01 to 1.50 occupants per room	2.2 (1.1,5.5)	2.2 (1.1,6.6)	2.2 (0,5.5)
1.51 to 2.00 occupants per room	0 (0,2.2)	0 (0,2.2)	0 (0,2.2)
2.01 or more occupants per room	0 (0,0)	0 (0,1.1)	0 (0,0)
Poverty, median % (IQR)			
<100% of poverty level	15.4 (7.7,25.3)	16.5 (8.8,26.4)	13.2 (7.7,24.2)
100-149% of poverty level	9.9 (5.5,14.3)	9.9 (5.5,15.4)	9.9 (5.5,14.3)
≥150% of poverty level	73.7 (59.4,85.8)	72.6 (58.3,84.7)	75.9 (61.6,86.9)
Gini Index*, n (%)			
0.2 to <0.3	17 (2.6)	148 (1.2)	56 (2.0)
0.3 to <0.4	255 (39.1)	4537 (38.3)	1146 (41.6)
0.4 to <0.5	325 (49.8)	6092 (51.4)	1324 (48.1)
0.5 to <0.6	52 (8.0)	1030 (8.7)	219 (8.0)
0.6 to <0.7	3 (0.5)	48 (0.4)	9 (0.3)
0.7 to <0.8	0	0	0
Census tract is > 25% foreign born, n (%)	204 (31.3)	4120 (34.8)	804 (29.2)

* Measure of income inequality for a geographic area where zero indicates absolute equality and one indicates total inequality.

Table 6.3. Comparison of results of original model vs. models adjusted for minimally sufficient subsets (MSS) for effectiveness against PCV13-type and non-PCV13-type disease.

Model±	VE (95% CI) PCV13	VE (95% CI) NVT*	PCV13-type discordant pairsβ	Absolute % difference in VE vs. unadjusted for PCV13-types
Original (unadjusted, except for matching factors)	85.1 (73.8 – 91.9%)	21.4 (-18.8 – 47.7%)	96	Referent
Primary minimally sufficient confounding subset				
MSS1¥: other children in household, influenza vaccination in the year before culture	83.5 (67.3 – 91.6%)	32.6 (-12.7 – 59.7%)	80	-1.6%
Additional minimally sufficient confounding subsets				
MSS2¥: other children in household, crowding, influenza infection in 30 days before culture	81.2 (62.9 – 90.4%)	35.6 (-8.3 – 61.7%)	76	-3.9%
MSS3¥: caregiver education, crowding, influenza infection in 30 days before culture, prenatal care utilization, recent immigrant neighborhood	83.1 (55.1 – 93.6%)	39.2 (-6.5 – 65.3%)	52	-2.0%
MSS4¥: caregiver education, influenza vaccination in the year before culture, prenatal care utilization, recent immigrant neighborhood	82.4 (55.3 – 93.0%)	35.2 (-13.6 – 63.0%)	54	-2.7%

* PCV13 = 13-valent pneumococcal conjugate vaccine; NVT = non-vaccine types; MSS = Minimally Sufficient confounding Subset.

± All models include adjustment for the matching variables, age and zip code. MSS1 was considered the primary subset due to less missing data (most discordant pairs retained).

β Because this is a conditional (matched) analysis, only matched sets which have discordant vaccination status (i.e., vaccinated case/unvaccinated control[s] or unvaccinated case/vaccinated control[s]) contribute to the analysis.

¥ All MSSs included adjustment for: matching factors (age and zip code), asthma, breastfeeding, underlying condition, daycare attendance, household income, insurance type at culture, race/ethnicity, and smoking exposure. Additional variables included in each subset indicated in table.

CHAPTER 7: GENERALIZABILITY OF VACCINE EFFECTIVENESS ESTIMATES: AN ANALYSIS OF CASES INCLUDED IN A POST-LICENSURE EVALUATION OF 13-VALENT PNEUMOCOCCAL CONJUGATE VACCINE

Overview

External validity, or generalizability, is the measure of how well results from a study pertain to individuals in the target population. We assessed generalizability, with respect to socioeconomic status, of estimates from a matched case-control study of 13-valent pneumococcal conjugate vaccine effectiveness for prevention of invasive pneumococcal disease in children in the US. Cases were identified from active surveillance and controls were age and zip code-matched. We enrolled 54.6% of eligible cases and found a trend toward enrolled cases being more affluent than unenrolled cases. Enrolled cases were slightly more likely to have private insurance at birth ($p=0.08$) and have mothers with at least some college education ($p<0.01$). Enrolled cases also tended to come from more affluent census tracts. Despite these differences, our best predictive model for enrollment yielded a concordance statistic of only 0.703, indicating mediocre predictive value. Variables retained in the final model were assessed for effect measure modification and none were found to be significant modifiers of vaccine effectiveness. We conclude that, although enrolled cases are somewhat more affluent than unenrolled cases, our estimates are externally valid with respect to socioeconomic status. Our analysis provides evidence that this study design can yield valid estimates and the assessing generalizability of observational data is feasible, even when unenrolled individuals cannot be contacted.

Introduction

External validity, or generalizability, is the measure of how well results from a study pertain to individuals in the target population.[116-118] Study results may fail to generalize to a target population

under several circumstances.[6,116] A key reason that studies fail to be generalizable is when both of the following criteria are met: (a) selection/enrollment into the study is differential with respect to variable X (i.e., the study population is not representative of the source population with respect to variable X, which is usually a potential confounder), and (b) variable X modifies the exposure/outcome relationship under study.[6,116]

Increased external validity can be a major benefit of observational studies as compared to randomized controlled trials (RCTs).[6,89,91,92,117,118,137,138] Unlike in RCTs, where strict exclusion criteria are often used, observational studies usually enroll a broader subset of the population.[118] This may be especially true in case-control studies, in which researchers commonly attempt to enroll *every* case of disease in a given surveillance population (e.g., geographic area or hospital system).[139] With the assumption that the study source population (in this case the surveillance area) does not vary in a meaningful way from the target population (e.g., an entire country), then results from observational studies may be more generalizable than those from RCTs.[139]

Unfortunately, due to issues related to tracking, contacting, and enrolling participants, and in some cases obtaining needed biological samples from participants, generalizability in observational studies may not be achieved, even when *a priori* exclusion criteria are minimal. In particular, the study population may differ from the source population by socioeconomic status (SES), which is frequently associated with characteristics that may affect enrollment probability.[120] For example, less affluent individuals may lack a landline or frequently switch cell phone numbers, making them harder to reach; they may hold multiple jobs or have long commutes, making them less likely to answer their phone; they may have diminished use of or trust in the medical system, making them less likely to agree to enrollment; they may seek care at underfunded (e.g., public) hospitals which lack the resources to enroll participants and preserve specimens.[5,9,99,132,140,141] On the other hand, lower SES individuals may be more

likely to join a study that provides a monetary incentive for participating versus those of higher SES for which the monetary incentive is lower.

While not always a problem, lack of representativeness by SES may reduce external validity if SES modifies the effect of the exposure/outcome relation under study.[100,107,142] One way to explore generalizability as it relates to SES is to evaluate differences in SES indicators between those selected into a study and those who were eligible, but not selected. Although some individual SES constructs, such as use of or trust in the medical system, may not be measured accurately, others, such as educational attainment, are more readily available. Additionally, other characteristics, such as recent immigration, insurance status, and utilization of prenatal care, can serve as proxies for harder to measure SES constructs, providing some context for the demographic characteristics of the underlying source population and their participant population.

Many studies are not able to access extensive SES information for individuals who are not enrolled. We present an analysis of the external validity of estimates from a study with access to relatively detailed information on unenrolled cases, providing a rare opportunity to assess differences in key SES variables between enrolled and unenrolled cases and, for variables with substantial differences, assessed effect measure modification (EMM), using data from a large matched case-control study of 13-valent pneumococcal conjugate vaccine (PCV13).

Methods

Parent study and study population

PCV13 was introduced in the US in 2010 as part of the routine childhood immunization schedule.[25] Licensure was largely based on immunogenicity and safety data, making a post-licensure evaluation the best opportunity to measure vaccine effectiveness.[25,119] The primary aim of the parent study was to assess the effectiveness of 1 or more doses of PCV13 in preventing invasive pneumococcal disease (IPD) caused by one of the 13 serotypes included in the vaccine.[119] Cases of IPD were identified

through Active Bacterial Core surveillance (ABCs), part of the Centers for Disease Control and Prevention's (CDC) Emerging Infections Program, or through one of three additional surveillance sites with similar case identification methods.[105,119] The source population consisted of IPD cases in children aged 2-59 months with culture dates between May 1, 2010 and May 31, 2014. Only children with a pneumococcal serotype available were eligible for enrollment. Additionally, children were deemed ineligible if the IPD episode was recurrent or if they were living in a long-term care facility at the time of culture.

In addition to the exclusion criteria for the parent study, we excluded cases from two ABCs sites, Colorado and Maryland, because the necessary SES data were not available from those sites.

In the current study, almost 45% of cases were not enrolled because they lacked a specimen for serotyping, could not be located, or refused enrollment, leading to concerns about representativeness with respect to SES (Table 7.1). Further, past studies of IPD and pneumococcal vaccines have provided some evidence for differential risk of IPD at different levels of individual SES, potentially due to the association of lower SES with IPD risk factors (e.g., existing chronic or immunocompromising conditions, smoking exposure, household crowding).

Enrollment procedures

Enrollment procedures have been previously published.[119] Briefly, once a case was identified via routine surveillance, study staff attempted to locate a correct phone number for the case. If a case was successfully located, staff began the enrollment process, which included attempting to interview a parent/guardian to gain consent and to ascertain medical history, IPD risk factors (e.g., smoking exposure, household crowding), and contact information for vaccine providers. A $20 gift card was offered as an incentive. Study staff then contacted vaccine providers to ascertain vaccine and medical histories. Staff also reviewed any available vaccine histories through state or city immunization information systems. If a parent/guardian could not be contacted or refused enrollment, if a vaccine

history could not be obtained, or if the parent/guardian did not speak English or Spanish, the case was excluded. Both the parent study and the current analysis were approved by institutional review boards at CDC and surveillance sites. The current analysis was also approved by the IRB of the University of North Carolina – Chapel Hill.

Lists of eligible controls were obtained from local birth registries for children born in the same zip code in which the case resided when diagnosed with IPD and born within 14 days of the case. Controls were called in order, beginning with those closest in age to the case, until 4 controls were enrolled for every case. Enrollment proceeded for controls in an identical manner to cases. Additionally, controls were excluded if they had moved out of the surveillance area by the case's culture date, they had been previously approached for enrollment, they were the sibling of a previously enrolled child, their birth mother no longer had custody, or if they had died for any reason. Detailed information on eligible and enrolled controls has been previously published.[143]

Covariate assessment

Every IPD case identified through ABCs has a standard case report form completed by surveillance staff.[108] We included the following variables from the case report form: hospitalization status, length of hospitalization and ICU admission (if hospitalized), outcome (survived/died), presence of an underlying condition which is a known risk factor for IPD, and insurance status at IPD culture.[25,108] Hospitalization and intensive care unit (ICU) admission were combined into a three-level severity index, with a child who was not hospitalized being the least severe, followed by a hospitalized child who was not admitted to the ICU, and then a child admitted to the ICU.

Individual-level SES indicators

For cases and controls, study staff obtained variables from the US standard birth certificate.[144] Variables collected included: mother's race, ethnicity, and education, source of payment for the birth,

and timing/initiation of prenatal care and gestational age, which were used to calculate the Adequacy of Prenatal Care Utilization (APNCU) Index.[110,111] The APNCU is a four-category variable which summarizes both how early in the pregnancy a woman initiated prenatal care and how many total prenatal care visits the woman received compared to how many are recommended by the American College of Obstetrics and Gynecology for a pregnancy of a given length.[110,111] Insurance status and APNCU Index were collected as measures of the individual's access and use of the medical system.[120]

Neighborhood-level SES indicators

All enrolled and unenrolled children were geocoded by surveillance site personnel, using a standard geocoding protocol (available from the authors upon request). For children whose locations could not be confirmed, surveillance staff geocoded the most recent known address (from the birth certificate, state or city immunization registry, or medical chart). Census tract data were merged with the 2013 5-year estimates from the American Community Survey (ACS), an annual survey conducted by the US Census Bureau, which includes demographic and SES indicators, such as median household income, employment rates, and distribution of crowding. These census tract-level estimates can be a useful tool for understanding the context in which a child lives by complementing individual-level SES indicators.[123,128]

Comparison groups

It is important to note that in matched case-control studies, controls need not be representative of the source population with respect to the matching factor(s). In fact, if matching is done correctly, controls should resemble enrolled cases on the matching factor. This analysis, therefore, focuses on representativeness of enrolled cases with respect to eligible cases and leaves questions concerning controls for other analyses.[143]

Statistical methods

Exploratory analyses looked at univariate differences in individual-level characteristics and ACS variables between enrolled and unenrolled cases. We also used the ACS variables to calculate a composite index of SES, based on the Socioeconomic Position Index (SEP Index) created by Krieger et al. for the Public Health Disparities Geocoding Project.[24] The SEP Index measures the major SES constructs of wealth, education, and occupation at the neighborhood level. Specifically, it includes: working class, unemployment, poverty, education, home prices, and median family income (with home prices and median income reversed, so a higher SEP score indicates lower SES).[145] P-values were calculated for continuous and categorical predictors via two-sided Wilcoxon rank sum test and Fisher's exact test, respectively.

Because we had access to many SES-related variables through the case report form, birth certificates, and ACS, it was not practical to assess every variable for effect measure modification (EMM). Instead, we chose variables to assess for EMM in two ways. First, we decided *a priori* to assess as modifiers the SEP Index and the following individual-level variables which had p-values <0.2 in the exploratory analyses for differences between enrolled and unenrolled: maternal education, insurance status at birth, and APNCU. Second, to ensure we did not miss any variables for the EMM assessment that may be strong modifiers *and* related to representativeness, we used predictive modeling to select additional ACS variables to assess for EMM. Many of the available ACS variables measured similar metrics (both to other ACS variables and to the individual variables); therefore, we fit a series of models to narrow the selection. All models fit were logistic regression, using backward selection and retaining variables with a p-value of <0.2.

We first fit a single predictive model including all individual-level variables (i.e., case report form and birth certificate variables), requiring the model to retain the following case report form variables: severity, outcome, underlying condition status, child's race/ethnicity, and insurance status at IPD. This

model also included interaction terms between child's race/ethnicity and maternal education and between child's race/ethnicity and APNCU. We then divided the ACS variables into four categories – demographics, wealth, work, and household characteristics – for inclusion in separate models to reduce the number of ACS variables used. Four models were run, including all individual-level variables retained from the first model and each category of the ACS variables. Lastly, we fit the final predictive model, which included all the individual-level variables and interaction terms (regardless of whether they were retained in the first model) and any variables retained in the ACS models. As with the first model, we required the model to retain the case report form variables.

Any variables included in the final model predicting enrollment were assessed as potential modifiers, in addition to the SEP Index and individual-level variables chosen *a priori*. All the individual-level variables were already categorical. We did not include variables (i.e., severity of IPD, outcome status) that are not applicable to controls. Additionally, variables obtained from the case report form for cases (child's race/ethnicity, underlying condition status, and insurance status at IPD) were not available for controls (since no case report form is completed for controls); however, these variables were collected as part of the parent interview/provider follow-up for enrolled children and could therefore be assessed for EMM. The ACS variables were continuous, so we dichotomized them at the median for the EMM analysis. To assess variables as modifiers, we fit conditional logistic regression models for each variable with IPD caused by one of the 13 serotypes included in the vaccine as the outcome, receipt of one or more doses of PCV13 as the exposure, and an interaction term between each variable of interested and PCV13 receipt.[119] Because power to assess interaction terms is reduced, we used a p-value <0.2 as our cutoff for the likelihood ratio chi-square value for the interaction term and then did a Bonferroni adjustment by the total number of variables assessed for EMM to account for multiple comparisons. Therefore, $\alpha<0.015$ for the likelihood ratio test was considered an indication of modification and cause for concern about generalizability.

Results

Enrollment

The annual population of children under age five in the catchment area for the parent study included approximately 4.7 million children, which was reduced to 4.2 million with the exclusion of sites that could not obtain the necessary variables for purposes of this analysis. We identified 1,214 cases of IPD in children in the catchment area used for this analysis. Three children were initially miss-categorized as ineligible by surveillance personnel and enrollment was not attempted, leaving 1,211 cases in our analysis, 661 (54.6%) of whom were enrolled (Table 7.1). Of the 550 cases not enrolled, 194 (35.3%) could not be located/contacted, 177 (32.2%) refused, and 158 (28.7%) did not have a pneumococcal isolate available for serotyping. The remaining 21 children were not enrolled for other reasons.

Differences between enrolled and unenrolled cases

Enrolled and unenrolled cases were of similar age (median age 21-22 months) and race/ethnicity (42%-43% white, non-Hispanic, p=0.41; Table 7.2a). Enrolled children were slightly more likely to have private insurance coverage at the time of IPD culture (49.9% vs. 47.7%), to be hospitalized (68.8% vs. 66.4%) and, if hospitalized, more likely to have been admitted to the ICU (31.9% vs. 29.0%), although none of these differences were statistically significant. Enrolled children were slightly more likely to survive their illness (98.8% vs. 97.0%, p=0.04), and slightly less likely to have an underlying chronic or immunocompromising condition, although differences in underlying conditions were not statistically significant.

As with the children themselves, mothers of enrolled and unenrolled children had a similar racial/ethnic distribution (Table 7.2b). Differences in insurance status at birth were more pronounced than those at culture, with 48.2% of enrolled children having private insurance at birth compared with only 39.8% of unenrolled children (p=0.08). Maternal education was also somewhat different between

the two groups, with 44.1% of mothers of enrolled children having no college education compared to 53.9% of mothers of unenrolled children ($p<0.01$). Mothers of enrolled and unenrolled children had relatively similar APNCU Index scores ($p=0.16$); however, when broken into its individual components, enrolled children had earlier prenatal care initiation ($p=0.01$).

Univariate differences between enrolled and unenrolled cases in ACS variables retained in the predictive model are in Table 7.2c. The SEP Index was similar in enrolled and unenrolled cases ($p=0.07$). A number of characteristics differed between the two groups, including proportion with an income above $60,000, proportion disabled, and proportion of households occupied ($p<0.01$ for all).

Enrollment prediction model

Full results of the predictive models are in Table 7.3. In the individual-level model, three variables – maternal education, APNCU, and the interaction between race/ethnicity and maternal education – were retained in addition to the five variables that we required. The model yielded a concordance (c-) statistic of 0.639, indicating only marginal predictive ability (0.5 is equal to a coin flip).[146] Retaining these eight variables and adding demographic, wealth, work, and household variables from the ACS yielded similarly low c-statistics of 0.632, 0.652, 0.643, and 0.655, respectively. The final predictive model included IPD severity, outcome, underlying condition status, child's race/ethnicity, insurance status at IPD, maternal education, APNCU, race/ethnicity and maternal education interaction, as well as the following census tract variables: percent of tract of white race, earning <$30,000, earning ≥$60,000, disabled, working class, occupied households, and mean hours worked (Table 7.4). This model had a concordance statistic of 0.703, indicating only slightly better predictive ability than the single ACS category models.

Effect measure modification

None of the individual-level variables retained in the prediction model were found to be significant modifiers of the effect of PCV13 receipt on IPD caused by one of the 13 serotypes in the vaccine (Figure 7.1a). Likewise, none of the ACS variables (including SEP Index) assessed as modifiers had a p-value<0.015 for the likelihood ratio test (Figure 7.1b).

Discussion

We assessed the external validity of vaccine effectiveness estimates from a post-licensure evaluation of PCV13 in children in the US. Despite a small trend toward enrolled cases being more affluent than unenrolled cases, the differences were minimal and most did not meet *a priori* definitions of significance. Additionally, we did not find EMM of the vaccine effectiveness by any of SES variables and therefore conclude that lack of generalizability to the broader source population is of minimal concern in the current study. Our results provide some evidence that study designs based on population-based surveillance systems can be generalizable – in particular, when researchers conduct extensive investigations into case contact information and provide incentives (in our case, a $20 gift card) to encourage enrollment.

Generalizability can be a major benefit of observational vaccine studies over RCTs[92,147]; however, it is rare for reports on either study type to provide extensive information on how their study population differs from their source or target population.[116,148] Our study used a novel approach to assessing generalizability with respect to SES, assessing a mix of variables chosen *a priori* and through predictive modeling. We had access to more detailed information on unenrolled cases than is typical; however, the data we assessed were collected without patient contact and therefore could be gathered more routinely in observational studies. Our analysis shows that assessing and reporting on generalizability may be feasible even when it is not possible to interview unenrolled cases.

Post-licensure vaccine effectiveness studies are common tools to assess the "real world" effectiveness of vaccines and, as with similar studies, cases were identified from a disease surveillance system.[41,42,46,105] Generalizability of results with respect to SES from vaccine studies is an important component of assessing the overall quality of a study, as well as the utility of results beyond the study population. Thus, although we found little cause for concern in the current analysis, this conclusion may not be applicable for vaccines for diseases, such as rotavirus, which are spread differently and therefore may have different risk factors and potential modifiers. Likewise, if case identification or enrollment methods are different, representativeness to the source population could be lacking. Since SES has been identified as a significant risk factor for many vaccine preventable diseases, other post-licensure vaccine studies may find more substantial cause for concern. However, careful selection of case and control populations, regardless of the disease under study, can yield results, such as these, which are generalizable to the underlying source population.

Our study had some limitations. We did not have access to variables collected during the parent interview (e.g., household income) for unenrolled cases and so could not explore differences between enrolled and unenrolled cases for these variables. We may have also lacked sufficient power to identify EMM in our study. For example, the confidence limits for vaccine effectiveness amongst those without insurance were extremely wide. Likewise, a vaccine effectiveness study of the 7-valent pneumococcal conjugate vaccine (a more limited valency, but otherwise identical, vaccine) in a similar population to the current analysis found clear evidence of EMM by underlying condition status.[42] We did not identify EMM, but only had two matched sets with underlying conditions and discordant vaccination status (required to contribute to a conditional analysis) in our study. Additionally, we were not able to compare cases eligible for our study (i.e., the source population) to cases throughout the US (i.e., the target population) because active, population-based surveillance for IPD does not exist nationwide.

We identified minimal external validity concerns in a PCV13 vaccine effectiveness study in children in the US. Nevertheless, future vaccine effectiveness studies should take care to enroll as broad a subset of cases as possible, especially focusing on children from lower SES areas. Additionally, every effort should be made to gather at least basic SES information so that study populations can be compared to source populations and estimates can be understood in context.

Table 7.1. Enrollment numbers, reasons for non-enrollment, and geocoding status for cases

Enrollment status	Cases, n (%)	Could not be geocoded, n (%)
Annual surveillance population <5 years of age*	4,249,724	N/A
IPD cases identified	1,211±	21 (1.8)
Enrolled	661 (54.6)	9 (1.4)
Not enrolled	550 (45.4)	12 (2.2)
Could not locate/contact	194 (35.3)	4 (2.1)
Refused	177 (32.2)	5 (2.8)
No serotype available	158 (28.7)	2 (1.3)
Recurrent	11 (2.0)	1 (9.1)
Vaccine history not available	6 (1.1)	0
Resident of long-term care facility	4 (0.7)	0

* Average population <5 years of age in the catchment area.

± Excludes 3 children originally miss-classified as ineligible and excluded from this analysis.

Table 7.2. Characteristics of eligible cases by enrollment status. (a) ABCs case report form (cases only); (b) birth certificate variables; (c) American Community Survey.

(a) ABCs case report form

Characteristic	Enrolled (n=661)	Not enrolled (n=550)	p-value*
Median age, median months (IQR)	21 (11-37)	22 (12-37)	0.27
Child's race			
White, non-Hispanic	262 (42.2)	217 (43.1)	0.41
Black, non-Hispanic	157 (25.3)	139 (27.6)	
Other, non-Hispanic	64 (10.3)	55 (10.9)	
Hispanic	138 (22.2)	92 (18.3)	
Unknown	40	47	
Insurance status at IPD			
Private	207 (49.9)	205 (47.7)	0.83
Public	254 (47.0)	208 (48.4)	
Military	2 (0.4)	2 (0.5)	
Uninsured	15 (2.8)	15 (3.5)	
Other/unknown	120	120	
Hospitalized	454 (68.8)	354 (66.4)	0.38
Median length of stay in days, (IQR)	5 (3-10)	5 (2-10)	0.65
ICU	124 (31.9)	89 (29.0)	0.46
Survived	648 (98.8)	511 (97.0)	0.04
Chronic condition	9 (1.4)	5 (0.9)	0.59
Immunocompromising condition	64 (9.7)	64 (11.6)	0.30

* P-value is from Wilcoxon rank sum test for continuous variables and Fisher's exact test for categorical variables.

IQR=Interquartile range; IPD=Invasive pneumococcal disease; ABCs=Active Bacterial Core surveillance; ICU=Intensive care unit

(b) Birth certificates

Characteristic	Enrolled (n=661)	Not enrolled (n=550)	p-value*
Maternal race/ethnicity, n (%)			
White, non-Hispanic	268 (46.7)	206 (44.8)	0.22
Black, non-Hispanic	127 (22.1)	127 (27.6)	
Hispanic	52 (9.1)	36 (7.8)	
Other, non-Hispanic	127 (22.1)	91 (19.8)	
Unknown	87	90	
Maternal education level, n (%)			
Less than high school	110 (20.2)	99 (22.7)	<0.01
High school equivalent	130 (23.9)	136 (31.2)	
Some college	146 (26.8)	115 (26.4)	
College degree or more	158 (29.0)	86 (19.7)	
Unknown	117	114	
Source of payment for birth, n (%)			
Private	228 (48.2)	151 (39.8)	0.08
Public/state	223 (47.2)	209 (55.2)	
Uninsured	8 (1.7)	9 (2.4)	
Other	14 (3.0)	10 (2.6)	
Unknown	188	171	
Adequacy of Prenatal Care Utilization, n (%)			
Adequate Plus	183 (34.9)	135 (32.8)	0.16
Adequate	186 (35.5)	157 (38.1)	
Intermediate	67 (12.8)	37 (9.0)	
Inadequate	88 (16.8)	83 (20.2)	
Unknown	137	138	
Gestational age, median weeks (IQR)±	39 (38-40)	39 (38-40)	0.57
Prenatal care initiation, median month (IQR) ±	2 (2-3)	3 (2-4)	0.01
Total prenatal care visits, median (IQR) ±	12 (10-16)	12 (10-18)	0.94

* P-value is from Wilcoxon rank sum test for continuous variables and Fisher's exact test for categorical variables.

± Included in calculation of Adequacy of Prenatal Care Utilization Index.[110,111]

IQR=Interquartile range

(c) American Community Survey. Unless otherwise noted, all measurements are the median percent (IQR) by census tract.

Characteristic	Enrolled (n=661)	Not enrolled (n=550)	p-value (univariate)*	p-value (multivariable model)
Successfully geocoded, n (%)	652 (98.6)	538 (97.8)	0.28	N/A
White race	56.1 (17.6, 81.9)	58.9 (19.8, 83.9)	0.26	<0.01
Household income <$30,000	28.1 (17.3, 38.5)	29.9 (18.9, 42.1)	0.03	<0.01
Household income ≥$60,000	41.1 (29.4, 57.8)	38.2 (25.1, 53.8)	<0.01	<0.01
Mean hours worked (annually)	38.1 (36.8, 39.3)	38.0 (36.7, 39.2)	0.39	0.05
Disabled workers	10.1 (7.5, 13.5)	11.2 (8.2, 16.0)	<0.01	<0.01
Working class	60.3 (52.6, 69.8)	62.3 (52.5, 69.8)	0.36	<0.01
Occupied houses	92.4 (88.1 ,95.8)	91.3 (86.4, 94.7)	<0.01	0.07
SEP Index±, median (IQR)	1.5 (-1.4, 4.5)	2.1 (-0.8, 4.9)	0.07	N/A

* P-value is from Wilcoxon rank sum test for continuous variables and Fisher's exact test for categorical variables.

± Socioeconomic position index. Higher value indicates higher level of deprivation.[24] IQR=Interquartile range

Table 7.3. Prediction models for case enrollment

Model	# of variables in model		AIC	C-statistic
	Start	After selection		
Individual-level	11	7 (IPD severity, outcome, underlying condition status, race/ethnicity, insurance at IPD, maternal education, race/ethnicity*maternal education)	717.840	0.639
ACS: demographics	26	9 (individual-level from above + percent: white race, speaks English)	927.995	0.632
ACS: wealth	18	11 (individual-level from above + percent: living abroad 1 year ago, earning <$30,000, earning ≥$60,000, per capita income)	927.838	0.652
ACS: work	16	10 (individual-level from above + percent: mean hours worked, disabled, working class)	932.599	0.643
ACS: household	30	14 (individual-level from above + percent: owner-occupied housing units, households with one resident, households without children <18 years, crowded households, households with plumbing, households with kitchens, occupied housing units)	896.747	0.655
Final (all individual-level, plus retained ACS variables)	27	15 (individual-level from above + APNCU Index, percent: white race, earning <$30,000, earning ≥$60,000, disabled, working class, occupied houses, mean hours worked)	690.606	0.703

ACS=American Community Survey; IPD=Invasive pneumococcal disease; APNCU=Adequacy of Prenatal Care Utilization Index; AIC=Akaike information criterion

Table 7.4. Operational definitions of ACS variables included in final prediction model

Category	Variable name	Variable description	ACS Table Number	Cut point (median) for EMM analysis
Demographic	White race	Percent of individuals identifying as white, non-Hispanic	B03002	48.3%
Wealth	Household income <$30,000	Household income in the past 12 months (in 2013 inflation-adjusted dollars)	B19001	28.0%
	Household income ≥$60,000	Household income in the past 12 months (in 2013 inflation-adjusted dollars)	B19001	40.4%
Work	Working class	Percent of individuals reporting jobs in sales, service (except protective), production, transportation, and material moving, natural resources, construction and maintenance	C24010	61.4%
	Disabled	Percent of individuals with a disability	B18135	10.3%
	Mean usual hours worked	Mean number of hours worked annually	B23020	37.9 hours
Household	Units occupied	Percent of occupied housing units	B25002	92.3%

(a) Individual-level variables

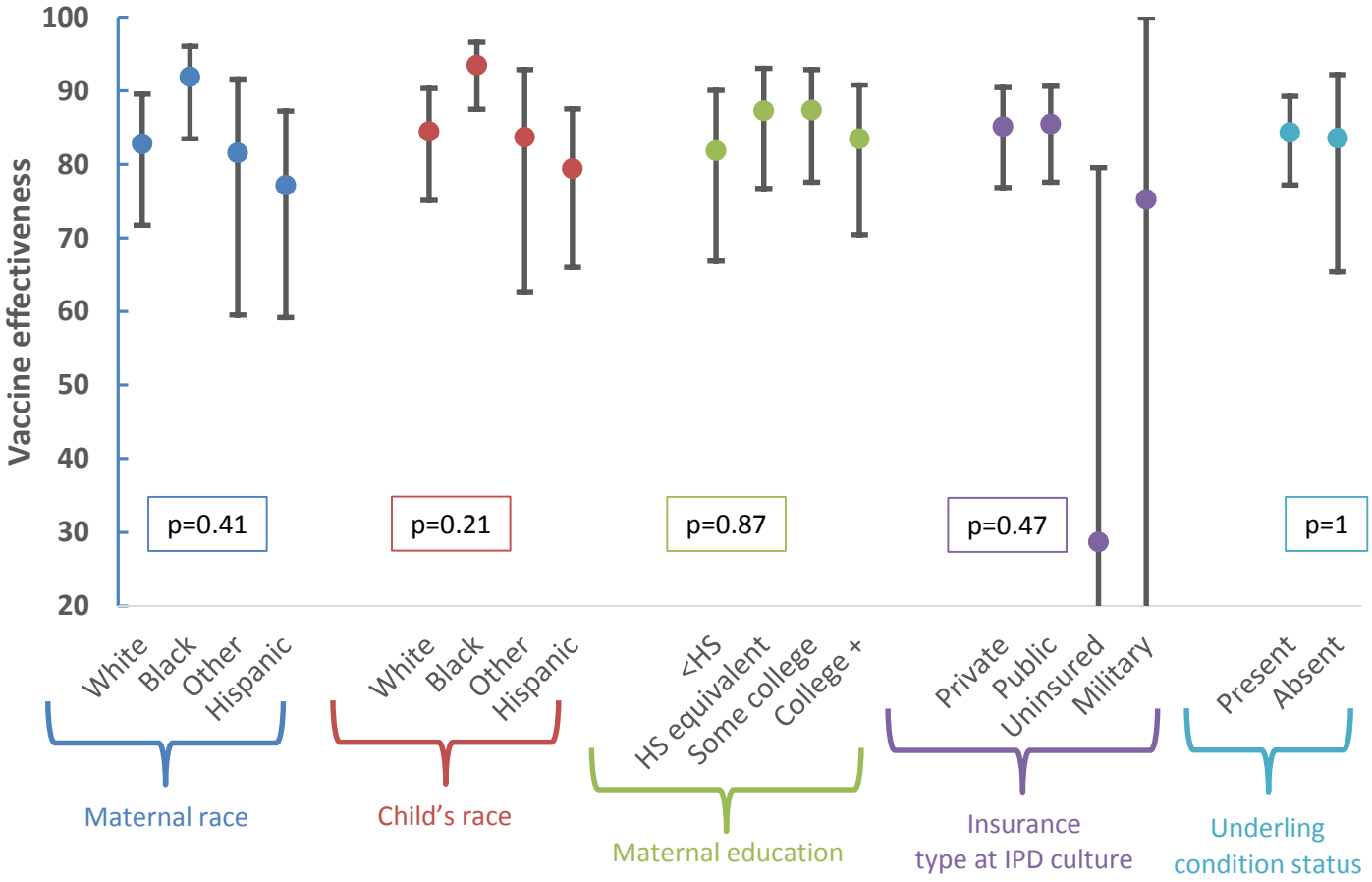
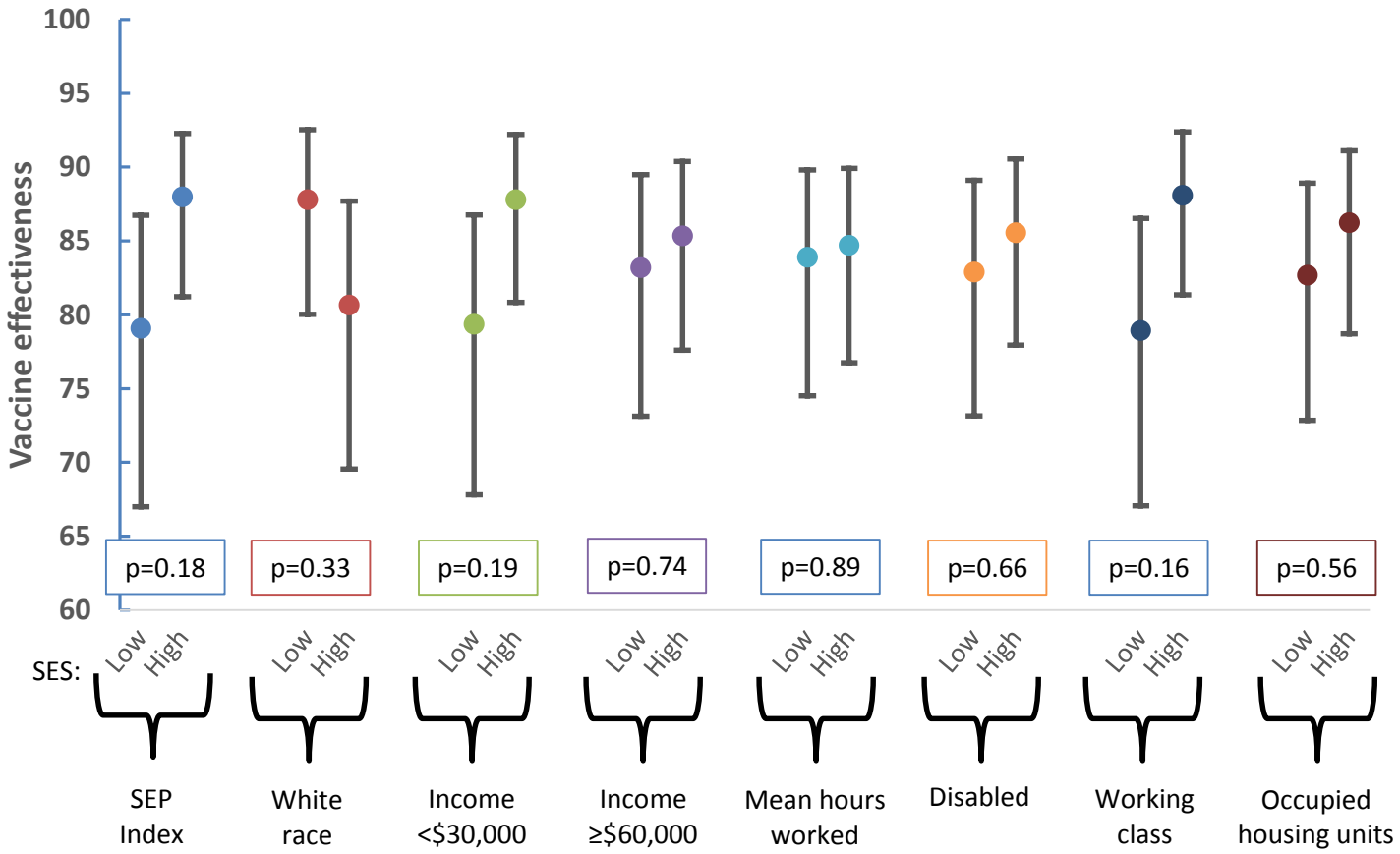


Figure 7.1. Vaccine effectiveness (and 80% CIs) for characteristics retained in the final predictive model, stratified as indicated. (a) Individual variables; (b) ACS variables, divided at median.

(b) ACS variables



CHAPTER 8: DISCUSSION

Overview

In the years following the introduction of PCV7, the burden of disease caused by *S. pneumoniae* in the US was reduced, but remained substantial.[66,67] In 2004, four years after PCV7 was introduced, direct medical costs of disease (including invasive and noninvasive disease) were estimated to be approximately $3.5 billion, including over one million episodes in children under five years of age and 1.1 million courses of antibiotics.[149] Costs likely changed little in the intervening years before PCV13 introduction, due to replacement disease and increasing medical costs in general.[66,149] The introduction of PCV13 in 2010 thus represented a significant opportunity to impact rates of disease, antibiotic usage (and thus, resistance), and healthcare costs. However, given remaining racial disparities in disease burden after PCV7 introduction, concerns about differential impact by race and/or socioeconomic status were raised and when CDC began an evaluation of the effectiveness of PCV13 in 2010, a secondary objective included evaluating the effectiveness in white vs. black children.[73,79,119]

Early in the enrollment process for the PCV13 VE study, it became apparent that study sites were having a more difficult time locating and enrolling children than had been the case during the PCV7 VE study a decade before. Site surveillance personnel raised concerns about the increased use of cell phones, higher proportions of recent immigrant families who were more transient and harder to locate, and recognized that there appeared to have been an increase in the number of parents/guardians of eligible children who did not speak English or Spanish and therefore could not be enrolled. Minnesota and New York City were particularly concerned about non-English/Spanish speaking populations, while Los Angeles County found it difficult to even contact parents/guardians of many eligible children (personal communications from C. Holtzman, J. Rosen, T. Motala, and R. Guevera).

In the interest of exploring how these issues might impact VE estimates, and adjusting our estimates if necessary, we began collecting additional data – geocodes and birth certificate variables – on eligible children. The overarching goal was to assess the validity of estimates from the current study with respect to SES and to make recommendations for changes in the study design for future VE studies. We hypothesized that 1) controls would be more affluent than cases, and 2) that enrolled children (cases and controls) would be more affluent than unenrolled children. The most likely change to the study design would be tighter matching criteria (either excluding adjacent zip codes or matching on census tract instead of zip code), although more drastic changes, such as matching via a pediatrician's office or using birth certificate variables to match, were also discussed. Simultaneously, CDC was drafting the protocol for an adult PCV13 VE study, and it was hoped that findings from the pediatric study could help inform methods for the adult study.

Summary of findings

In general, our hypotheses that controls would be more affluent than cases and that enrolled children would be more affluent than unenrolled children were confirmed in our univariate analyses. Controls tended to have better educated mothers and primary caregivers, live in households with higher family incomes, and were more likely to have private health insurance (both at birth and at time of IPD culture). Enrolled children (cases and controls) were similarly more affluent than their unenrolled counterparts (although family income for unenrolled children was not available).

Differences were less obvious at the census tract level, as census tracts are neighborhood measures that may or may not apply directly to individuals. There was, however, a slight trend toward controls coming from less crowded, wealthier census tracts with a lower percentage of recent immigrants compared to cases. Likewise, enrolled controls were slightly better off than eligible controls. No clear trend was seen in differences between enrolled and unenrolled cases.

Aim 1

While our general hypothesis was that controls would be less affluent than cases, we were interested in deconstructing the factors that go into a control sample in a matched study. Specifically, we wanted to separate the effects of zip code matching from those of selection and determine both if zip code matching worked *in theory* and if the composition of our actual control sample was altered by selection. While the combined answers to these questions dictate whether we should be concerned about internal validity, each question individually determines if changes should be made to the study design for future vaccine effectiveness studies.

The first part of this analysis involved a comparison between enrolled cases and eligible controls, with the aim of determining whether zip code matching worked to control for SES, in the absence of problems locating, contacting, and enrolling controls. Specifically, if zip code matching worked, we would expect eligible cases to look similar to enrolled cases with respect to characteristics, such as income, education, and insurance status, which measure SES. Using the birth certificate and ACS variables, we found that eligible controls were actually slightly *less* affluent that enrolled cases, as measured by maternal education and insurance status at birth.

The second part of the analysis compared enrolled cases to enrolled controls, with the goal of determining whether selection into the study resulted in a control group that was more or less similar to enrolled cases than the eligible controls. We found enrolled controls to be slightly more affluent than enrolled cases.

When all confounders from our minimally sufficient subsets were included in models, we did not find substantial cause for concern about reduced internal validity. Interestingly, while there were differences between our enrolled cases and controls, they did not have a meaningful change in our overall estimate. One possible explanation for the lack of effect on the overall estimate is that, while SES is related to IPD risk, it may be less related to vaccination than has historically been the case. The

passage and implementation of the Vaccines for Children (VFC) program has no doubt reduced disparities in vaccination coverage. Additionally, since PCV13 simply replaced PCV7 in the immunization schedule (rather than adding an additional vaccine to the schedule), acceptance may have been higher among parents, further reducing disparities.

Aim 2

As with Aim 1, differences between groups were seen in individual variables, but did not result in significant findings for multivariable models. In the case of Aim 2, despite differences between enrolled and unenrolled children, our predictive models were mediocre, implying that we could not fully predict enrollment based on a child's SES. This is an important finding, which should allay some concerns about selection bias and reinforce the notion that generalizability can be a major strength of observation studies.

Strengths

This study was the first to evaluate concerns about internal and external validity due to SES in a pediatric case-control vaccine effectiveness study in the US. Few studies of any kind have such comprehensive information on those cases and controls who are eligible for the study, but not enrolled, and we were thus be in a unique situation to assess the potential for bias. Further, with the exception of a handful of variables collected via parent interview, the majority of our SES indicators were collected without patient contact, demonstrating that routine assessment of validity in observational studies need not require exhaustive follow-up beyond study enrollment procedures. Additionally, strong existing disease surveillance and a catchment area of greater than 4 million children meant relatively high numbers of cases. Finally, as neighborhood matching and enrollment via telephone are common design components in observational studies, our results may inform future studies, both of vaccine effectiveness and other disease endpoints.

Limitations

ABCs surveillance sites are not randomly chosen and our results may therefore not be generalizable to the entire US population. Thus, we focused our generalizability analyses on whether our enrolled cases represented eligible cases *within ABCs,* rather than all cases within the US. Because active, population-based surveillance for IPD does not exist throughout the US, we could not address generalizability of cases within ABCs with respect to cases throughout the US. However, because of the wide geographic spread of ABCs, and the clear geographic boundaries of the catchment area, in the future we will be able to explore differences in our catchment area population as compared to the rest of the country. These analyses are ongoing within CDC.

Some misclassification of location was possible, due to incorrectly reported addresses and/or issues with geocoding procedures or software. This misclassification should be minor as sites followed standard geocoding procedures, including hand-checking any addresses that software initially failed to geocode. Further, small issues with specific addresses (e.g., typos in the street number) are unlikely to change the census tract, mitigating problems arising from errors in data entry. Further, because we collected both address at time of birth and IPD culture, if one address for a child was invalid (e.g., a P.O. Box), we frequently had a second address that was valid.

Finally, although the variables available in this study are likely better proxies for individual-level SES than zip code, they are unlikely to be perfect. For example, we did not have access to the actual income level of families not enrolled in the study, an important measure of SES; however, since perfect SES measures are unlikely to be available in most large studies, it is of substantial interest to better understand and utilize the available measures, such as those used in our study.

Future directions

As previously discussed, work is ongoing to explore the representativeness of the ABCs catchment area with respect to the entire US. Those findings will help us understand whether results

from both routine surveillance (e.g., disease trends) and special studies (e.g., vaccine effectiveness, risk factor analyses, etc.) are applicable to the entire US population.

The adult PCV13 vaccine effectiveness study, which is just beginning enrollment at the time of this writing (winter 2015/2016), will include, for the first time in an ABCs special study, routine, prospective geocoding of all eligible cases and controls, regardless of enrollment status. Identification of individual-level variables for eligible cases and controls (adults 65 years and older) was more difficult than for children, given the lack of a universal registry for elderly adults; however, some data may be available via Medicare enrollment databases from the Centers for Medicare and Medicaid Services. Discussions of this possibility are ongoing.

Conclusions

Our results provide assurance that previously published vaccine effectiveness estimates are both internally and externally valid. Further, we did not identify substantial need to change the basic case/control identification and enrollment methods in future studies. Rather, we showed that observational studies can (and should) evaluate validity in quantitative, rather than qualitative ways.

APPENDIX 1: PNEUMOCOCCAL VACCINES USED IN THE US SINCE 2000

	PPSV23	PCV7	PCV13
Licensure in US	1983	2000	2010
Tradename	Pneumovax 23	Prevnar	Prevnar13
Manufacturer	Merck & Co.	Wyeth Lederle*	Pfizer
Serotypes included	1		1
	2		
	3		3
	4	4	4
	5		5
			6A
	6B	6B	6B
	7F		7F
	8		
	9N		
	9V	9V	9V
	10A		
	11A		
	12F		
	14	14	14
	15B		
	17F		
	18C	18C	18C
	19A		19A
	19F	19F	19F
	20		
	22F		
	23F	23F	23F
	33F		

* Wyeth Lederle was purchased by Pfizer in 2009.

APPENDIX 2: 2014 ABCS CASE REPORT FORM

– ACTIVE BACTERIAL CORE SURVEILLANCE CASE REPORT –

Patient's Name: _____ (Last, First, MI.) Phone No.:( ) _____

Address: _____ (Number, Street, Apt. No.) Patient Chart No.: _____

_____ (City, State) _____ (Zip Code) Hospital: _____

– Patient identifier information is not transmitted to CDC –

DEPARTMENT OF HEALTH AND HUMAN SERVICES
CENTERS FOR DISEASE CONTROL
AND PREVENTION
ATLANTA, GA 30333

2014 Active Bacterial Core Surveillance (ABCs) Case Report

A CORE COMPONENT OF THE EMERGING INFECTIONS PROGRAM NETWORK

OMB No. 0920-0978

– SHADED AREAS FOR OFFICE USE ONLY –

1. STATE: *(Residence of Patient)* ☐☐

2. STATE I.D.: ☐☐☐☐☐☐☐

3. DATE FIRST POSITIVE CULTURE COLLECTED *(Date Specimen Collected)* Mo. ☐☐ Day ☐☐ Year ☐☐☐☐

4. Date reported to EIP site: Mo. ☐☐ Day ☐☐ Year ☐☐☐☐

5. CRF Status: 1 ☐ Complete 2 ☐ Incomplete 3 ☐ Edited & Correct 4 ☐ Chart unavailable after 3 requests

6. COUNTY: *(Residence of Patient)* _____

7a. HOSPITAL/LAB I.D. WHERE CULTURE IDENTIFIED: ☐☐☐☐☐

7b. HOSPITAL I.D. WHERE PATIENT TREATED: ☐☐☐☐☐

8. DATE OF BRTH: Mo. ☐☐ Day ☐☐ Year ☐☐☐☐

9a. AGE: ☐☐☐

9b. Is age in day/mo/yr? 1 ☐ Days 2 ☐ Mos. 3 ☐ Yrs.

10. SEX: 1 ☐ Male 2 ☐ Female

11a. ETHNIC ORIGIN: 1 ☐ Hispanic or Latino 2 ☐ Not Hispanic or Latino 9 ☐ Unknown

11b. RACE: *(Check all that apply)* 1 ☐ White 1 ☐ Black 1 ☐ American Indian or Alaska Native 1 ☐ Asian 1 ☐ Native Hawaiian or Other Pacific Islander 1 ☐ Unknown

12a. BACTERIAL SPECIES ISOLATED FROM ANY NORMALLY STERILE SITE:
1 ☐ *Neisseria meningitidis* 3 ☐ Group B *Streptococcus* 5 ☐ Group A *Streptococcus*
2 ☐ *Haemophilus influenzae* 4 ☐ *Listeria monocytogenes* 6 ☐ *Streptococcus pneumoniae*

12b. OTHER BACTERIAL SPECIES ISOLATED FROM ANY NORMALLY STERILE SITE: *(specify)* _____

13. STERILE SITES FROM WHICH ORGANISM ISOLATED: *(Check all that apply)*
1 ☐ Blood 1 ☐ CSF 1 ☐ Peritoneal fluid 1 ☐ Pericardial fluid 1 ☐ Bone 1 ☐ Muscle/Fascia/Tendon 1 ☐ Joint 1 ☐ Pleural fluid
1 ☐ Other normally sterile site (specify) _____ 1 ☐ Internal body site *(specify)* _____

14. OTHER SITES FROM WHICH ORGANISM ISOLATED: *(Check all that apply)*
1 ☐ Placenta 1 ☐ Amniotic fluid 1 ☐ Wound 1 ☐ Middle ear 1 ☐ Sinus

INFLUENZA 15. Did this patient have a positive flu test 10 days prior to or following any ABCs positive culture? 1 ☐ Yes 2 ☐ No 9 ☐ Unknown

16. WAS PATIENT HOSPITALIZED? 1 ☐ Yes 2 ☐ No
If YES, date of admission: Mo. ☐☐ Day ☐☐ Year ☐☐☐☐
Date of discharge: Mo. ☐☐ Day ☐☐ Year ☐☐☐☐

17. If patient was hospitalized, was this patient admitted to the ICU during hospitalization? 1 ☐ Yes 2 ☐ No 9 ☐ Unknown

18a. Where was the patient a resident at time of initial culture?
1 ☐ Private residence 2 ☐ Long term care facility 3 ☐ Long term acute care facility 4 ☐ Homeless 5 ☐ Incarcerated 6 ☐ College dormitory 7 ☐ Non-medical ward 8 ☐ Other *(specify)* _____ 9 ☐ Unknown

18b. If resident of a facility, what was the name of the facility? _____

19a. Was patient transferred from another hospital? 1 ☐ Yes 2 ☐ No 9 ☐ Unknown

19b. If YES, hospital I.D.: ☐☐☐☐☐

20a. WEIGHT: _____lbs_____ oz OR _____ kg OR ☐ Unknown

20b. HEIGHT: _____ft _____ in OR _____ cm OR ☐ Unknown

20c. BMI: ___ ___ . ___ OR ☐ Unknown

21. TYPE OF INSURANCE: *(Check all that apply)*
1 ☐ Private 1 ☐ Medicare 1 ☐ Medicaid/state assistance program 1 ☐ Military 1 ☐ Indian Health Service (IHS) 1 ☐ Incarcerated 1 ☐ Other *(specify)* _____ 1 ☐ Uninsured 1 ☐ Unknown

22. OUTCOME: 1 ☐ Survived 2 ☐ Died 9 ☐ Unknown

23. If patient died, was the culture obtained on autopsy? 1 ☐ Yes 2 ☐ No 9 ☐ Unknown

24a. At time of first positive culture, patient was:
1 ☐ Pregnant 2 ☐ Postpartum 3 ☐ Neither 9 ☐ Unknown

24b. If pregnant or postpartum, what was the outcome of fetus:
1 ☐ Survived, no apparent illness 2 ☐ Survived, clinical infection 3 ☐ Live birth/neonatal death 4 ☐ Abortion/stillbirth 5 ☐ Induced abortion 6 ☐ Still pregnant 9 ☐ Unknown

25. If patient <1 month of age, indicate gestational age and birth weight. If pregnant, indicate gestational age of fetus, only.
Gestational age: ☐☐ *(wks)* Birth weight: ☐☐☐☐ *(gms)*

26. TYPES OF INFECTION CAUSED BY ORGANISM: *(Check all that apply)*
1 ☐ Bacteremia without Focus 1 ☐ Meningitis 1 ☐ Otitis media 1 ☐ Pneumonia 1 ☐ Cellulitis 1 ☐ Epiglottitis 1 ☐ Hemolytic uremic syndrome (HUS) 1 ☐ Abscess (not skin) 1 ☐ Peritonitis 1 ☐ Pericarditis 1 ☐ Septic abortion 1 ☐ Chorioamnionitis 1 ☐ Septic arthritis 1 ☐ Osteomyelitis 1 ☐ Empyema 1 ☐ Endocarditis 1 ☐ Endometritis 1 ☐ STSS 1 ☐ Necrotizing fasciitis 1 ☐ Puerperal sepsis 1 ☐ Septic shock 1 ☐ Other *(specify)* _____ 1 ☐ Unknown

Public reporting burden of this collection of information is estimated to average 10 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection information, including suggestions for reducing this burden to CDC, CDC/ATSDR Reports Clearance Officer, 1600 Clifton Road, MS D-74, Atlanta, GA 30333, ATTN: PRA(0920-0978). **Do not send the completed form to this address.**

CDC 52.15A REV. 12-2013

– ACTIVE BACTERIAL CORE SURVEILLANCE CASE REPORT –

Page 1 of 2

– IMPORTANT – PLEASE COMPLETE THE BACK OF THIS FORM –

27. UNDERLYING CAUSES OR PRIOR ILLNESSES: *(Check all that apply* OR *if* NONE *or* CHART UNAVAILABLE,*check appropriate box)* 1☐ None 1☐ Unknown

1☐ AIDS or CD4 count <200
1☐ Alcohol Abuse, Current
1☐ Alcohol Abuse, Past
1☐ Asthma
1☐ Atherosclerotic Cardiovascular Disease (ASCVD)/CAD
1☐ Bone Marrow Transplant (BMT)
1☐ Cerebral Vascular Accident (CVA)/Stroke
1☐ Chronic Kidney Disease
1☐ Current Chronic Dialysis
1☐ Chronic Skin Breakdown
1☐ Cirrhosis/Liver Failure
1☐ Cochlear Implant
1☐ Complement Deficiency
1☐ CSF Leak
1☐ Current Smoker
1☐ Deaf/Profound Hearing Loss
1☐ Dementia
1☐ Diabetes Mellitus
1☐ Emphysema/COPD
1☐ Heart Failure/CHF
1☐ HIV Infection
1☐ Hodgkin's Disease/Lymphoma
1☐ Immunoglobulin Deficiency
1☐ Immunosuppressive Therapy (Steroids, Chemotherapy, Radiation)
1☐ IVDU, Current
1☐ IVDU, Past
1☐ Leukemia
1☐ Multiple Myeloma
1☐ Multiple Sclerosis
1☐ Nephrotic Syndrome
1☐ Neuromuscular Disorder
1☐ Obesity
1☐ Parkinson's Disease
1☐ Other Drug Use, Current
1☐ Other Drug Use, Past
1☐ Peripheral Neuropathy
1☐ Plegias/Paralysis
1☐ Premature Birth (specify gestational age at birth) ☐☐ *(wks)*
1☐ Seizure/Seizure Disorder
1☐ Sickle Cell Anemia
1☐ Solid Organ Malignancy
1☐ Solid Organ Transplant
1☐ Splenectomy/Asplenia
1☐ Systemic Lupus Erythematosus (SLE)
1☐ Other prior illness *(specify)* _____

– IMPORTANT – PLEASE COMPLETE FOR THE RELEVANT ORGANISM –

HAEMOPHILUS INFLUENZAE

28a. What was the serotype? 1☐ b 2☐ Not Typeable 3☐ a 4☐ c 5☐ d 6☐ e 7☐ f 8☐ Other *(specify)* _____ 9☐ Not Tested or Unknown

28b. If <15 years of age and serotype 'b' or 'unknown' did patient receive *Haemophilus influenzae* b vaccine? 1☐ Yes 2☐ No 9☐ Unknown
If YES, please complete the list below.

DOSE	DATE GIVEN (Mo. Day Year)	VACCINE NAME	MANUFACTURER	LOT NUMBER
1	☐☐ ☐☐ ☐☐☐☐			
2	☐☐ ☐☐ ☐☐☐☐			
3	☐☐ ☐☐ ☐☐☐☐			
4	☐☐ ☐☐ ☐☐☐☐			

28c. Were records obtained to verify vaccination history? *(<5 years of age with Hib/unknown serotype, only)*
1☐ Yes 2☐ No

If YES, what was the source of the information? *(Check all that apply)*
1☐ Vaccine Registry
1☐ Healthcare Provider
1☐ Other*(specify)* _____

NEISSERIA MENINGITIDIS

29. What was the serogroup? 1☐ A 3☐ C 5☐ W135 9☐ Unknown 2☐ B 4☐ Y 6☐ Not groupable 8☐ Other _____

30. Is patient currently attending college? 1☐ Yes 2☐ No 9☐ Unknown

31. Did patient receive meningococcal vaccine? 1☐ Yes 2☐ No 9☐ Unknown
If YES, please complete the following information:

DOSE	DATE GIVEN (Mo. Day Year)	VACCINE NAME	MANUFACTURER	LOT NUMBER
1	☐☐ ☐☐ ☐☐☐☐			
2	☐☐ ☐☐ ☐☐☐☐			
3	☐☐ ☐☐ ☐☐☐☐			

STREPTOCOCCUS PNEUMONIAE

32. Did patient receive pneumococcal vaccine?
1☐ Yes 2☐ No 9☐ Unknown

If YES, please note which pneumococcal vaccine was received:
(Check all that apply)
1☐ Prevnar®, 7-valent Pneumococcal Conjugate Vaccine (PCV7)
1☐ Prevnar-13®, 13-valent Pneumococcal Conjugate Vaccine (PCV13)
1☐ Pneumovax®, 23-valent Pneumococcal Polysaccharide Vaccine (PPV23)
1☐ Vaccine type not specified

If between ≥3 months and<18 years of age and an isolate is available for serotyping, please complete the Invasive Pneumococcal Disease in Children expanded form.

GROUP A STREPTOCOCCUS *(#33–35 refer to the 14 days prior to first positive culture)*

33. Did the patient have surgery or any skin incision? 1☐ Yes 2☐ No 9☐ Unknown

If YES, date of surgery or skin incision: Mo. ☐☐ Day ☐☐ Year ☐☐☐☐

34. Did the patient deliver a baby *(vaginal or C-section)*?
1☐ Yes 2☐ No 9☐ Unknown

If YES, date of delivery: Mo. ☐☐ Day ☐☐ Year ☐☐☐☐

35. Did patient have:
1☐ Varicella
1☐ Penetrating trauma
1☐ Blunt trauma
1☐ Surgical wound *(post operative)*
1☐ Burns

If YES to any of the above, record the number of days prior to the first positive culture
(if > 1, use the most recent skin injury)
1☐ 0-7 days 2☐ 8-14 days

36. COMMENTS: _____

– SURVEILLANCE OFFICE USE ONLY –

37. Was case first identified through audit? 1☐ Yes 2☐ No 9☐ Unknown

38. Does this case have recurrent disease with the same pathogen? 1☐ Yes 2☐ No 9☐ Unknown
If YES, previous (1st) state I.D.: ☐☐☐☐☐☐☐

39. Initials of S.O.: _____

Submitted By:_____ Phone No. : ( ) _____ Date:____/____/____

Physician's Name:_____ Phone No. : ( ) _____

APPENDIX 3: MEDICAL AND VACCINE HISTORY FORM

Vaccine and medical history form

Child's name:	Pneumococcal vaccine effectiveness evaluation subject ID _____-_____
Health care provider's name:	Provider ID:
Clinic name:	Person completing this form:
Clinic phone number:	Title of person completing this form:
Clinic fax number:	Date form completed ___/ ___/ ___ (Month/Day/Year)

For the child listed at the top, please attach a list of vaccines that the child has received with the dates given and, for pneumococcal conjugate vaccine, the lot numbers. Alternatively, you may complete the following table. Take extra caution to distinguish between the 7-valent and 13-valent formulations. Enter "99" for lot number if unknown.

	Dates of indicated doses					
Vaccine	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5	Dose 6
Prevnar-13® (13-valent pneumococcal conjugate vaccine)	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____
Prevnar® (7-valent pneumococcal conjugate vaccine)	__/__/__ Lot _____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____
Pneumovax® (23-valent pneumococcal polysaccharide vaccine)	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____	__/__/__ Lot_____
Diptheria/pertussis/ tetanus vaccine (DTaP or DTP or DT) (e.g. Acel-Imune, Infanrix, Tripedia, Certiva, Tri-Immunol, Trivivac)	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__
Haemophilus influenzae type B (Hib vaccine) (e.g. PedvaxHIB)	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__
Combined HiB/DTaP or HiB/DTP or DTaP/IPV/Hib vaccine (e.g. TriHiBit® Pentace)	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__
Seasonal influenza vaccine (e.g., Fluzone ®, FLUVIRIN®, FluMist®)	2009-2010 __/__/__	2009-2010 __/__/__	2010-2011 __/__/__	2010-2011 __/__/__	2011-2012 __/__/__	2011-2012 __/__/__
2009 Novel Influenza A(H1N1) vaccine	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__	__/__/__

To the best of your knowledge, has this child ever had invasive pneumococcal disease, defined by isolation of *Streptococcus pneumoniae* from a normally sterile site (e.g., blood, cerebrospinal fluid)? No / Yes If yes, date of episode _____ (Month/Day/Year)

For the child listed at the top, please indicate the presence of illnesses that increase the risk of pneumococcal disease using the following table.

Condition	**To your knowledge, has the child EVER been diagnosed with this condition?**	**If yes, was the condition diagnosed/present before (the case's specimen collection date)?**
Sickle cell disease (SS or SC)	Yes No	Yes No
Kidney disease	Yes No **If yes, what type:**_____ **If yes, did the child require dialysis?** Yes No	Yes No
Heart disease	Yes No **If yes, what type:**	Yes No
Diabetes	Yes No	Yes No
Asplenia	Yes No	Yes No
Cancer	Yes No **If yes, what type**:	Yes No
Bone marrow or organ transplant	Yes No	Yes No
Immune system disorder	Yes No **If yes, what type**:	Yes No
Asthma, reactive airways disease, or ≥1 episode of wheezing	Yes No	Yes No
If answered "Yes" to asthma questions above answer following 2 questions: 1. In the 30 days before (the case's specimen collection date), was the child prescribed any medications for asthma, reactive airways disease, or wheezing? Yes No **2. In the 12 months before (the case's specimen collection date), how many times has the child sought care in your office or an emergency room for an acute exacerbation?** Number of visits _____		
Other chronic lung condition	Yes No If yes, what type:	Yes No
Cochlear implant	Yes No	Yes No
Cerebrospinal fluid leak	Yes No	Yes No
Any other chronic illness	Yes No If yes, what type:	Yes No
Acute otitis media	Yes No	Yes No
If answered "Yes" for acute otitis media: In the 12 months before (the case's specimen collection date), how many times was the child treated for acute otitis media by a member of your practice? (check) 0 times ____ 1-3 times ____ ≥4 times ____		
Was the child prescribed prophylactic antibiotics for recurrent otitis media?		Yes No
Did the child have PE tubes placed?		Yes No
In the 30 days before (the case's specimen collection date):		
Did the child take any prescription antibiotics? Yes No		
Was the child diagnosed with influenza? Yes No **If yes, was the child hospitalized?** Yes No **If yes, what type (circle)**: Seasonal influenza Novel/swine/pandemic H1N1		
Was the child prescribed systemic steroids? Yes No		
Was the child prescribed any other immunosuppressive agents (e.g., Cytoxan, Etanercept, etc.)? Yes No		

REFERENCES

1. Advisory Committee on Immunization Practices, Centers for Disease Control and Prevention. Pertussis vaccination: use of acellular pertussis vaccines among infants and young children. Recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report (MMWR) 1997;46:1-25.

2. Advisory Committee on Immunization Practices, Centers for Disease Control and Prevention. Preventing pneumococcal disease among infants and young children. Recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report (MMWR) 2000;49:1-35.

3. Nuorti JP, Whitney CG, Centers for Disease Control Prevention. Prevention of pneumococcal disease among infants and children - use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine - recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report (MMWR) 2010;59:1-18.

4. Bilukha OO, Rosenstein N, National Center for Infectious Diseases, Centers for Disease Control Prevention. Prevention and control of meningococcal disease. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2005;54:1-21.

5. Klosky JL, Tyc VL, Lawford J, Ashford J, Lensing S, Buscemi J. Predictors of non-participation in a randomized intervention trial to reduce environmental tobacco smoke (ETS) exposure in pediatric cancer patients. Pediatric blood & cancer 2009;52:644-9.

6. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. American journal of epidemiology 2010;172:107-15.

7. Janssen I, Boyce WF, Simpson K, Pickett W. Influence of individual- and area-level measures of socioeconomic status on obesity, unhealthy eating, and physical inactivity in Canadian adolescents. Am J Clin Nutr 2006;83:139-45.

8. Pardo-Crespo MR, Narla NP, Williams AR, et al. Comparison of individual-level versus area-level socioeconomic measures in assessing health outcomes of children in Olmsted County, Minnesota. Journal of epidemiology and community health 2013;67:305-10.

9. Kramer MS, Wilkins R, Goulet L, et al. Investigating socio-economic disparities in preterm birth: evidence for selective study participation and selection bias. Paediatric and perinatal epidemiology 2009;23:301-9.

10. Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology 2007;17:643-53.

11. Gerber Y, Jacobsen SJ, Killian JM, Weston SA, Roger VL. Participation bias assessment in a community-based study of myocardial infarction, 2002-2005. Mayo Clinic proceedings 2007;82:933-8.

12. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. American journal of public health 2003;93:1655-71.

13. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. American journal of public health 2005;95:312-23.

14. Pilishvili T, Zell ER, Farley MM, et al. Risk factors for invasive pneumococcal disease in children in the era of conjugate vaccine use. Pediatrics 2010;126:e9-17.

15. Burton DC, Flannery B, Bennett NM, et al. Socioeconomic and racial/ethnic disparities in the incidence of bacteremic pneumonia among US adults. American journal of public health 2010;100:1904-11.

16. Lantz PM, House JS, Mero RP, Williams DR. Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. J Health Soc Behav 2005;46:274-88.

17. Spicer JO, Thomas S, Holst A, Baughman W, Farley MM. Socioeconomic and racial disparities of pediatric invasive pneumococcal disease after the introduction of the 7-valent pneumococcal conjugate vaccine. The Pediatric infectious disease journal 2014;33:158-64.

18. Cutts FT, Orenstein WA, Bernier RH. Causes of low preschool immunization coverage in the United States. Annu Rev Public Health 1992;13:385-98.

19. Walker AT, Smith PJ, Kolasa M, Centers for Disease C, Prevention. Reduction of racial/ethnic disparities in vaccination coverage, 1995-2011. MMWR Surveill Summ 2014;63 Suppl 1:7-12.

20. Hutchins SS, Baughman AL, Orr M, Haley C, Hadler S. Vaccination levels associated with lack of measles transmission among preschool-aged populations in the United States, 1989-1991. J Infect Dis 2004;189 Suppl 1:S108-15.

21. Hutchins SS, Jiles R, Bernier R. Elimination of measles and of disparities in measles childhood vaccine coverage among racial and ethnic minority populations in the United States. J Infect Dis 2004;189 Suppl 1:S146-52.

22. Smith PJ, Stevenson J. Racial/ethnic disparities in vaccination coverage by 19 months of age: an evaluation of the impact of missing data resulting from record scattering. Statistics in medicine 2008;27:4107-18.

23. Krieger N, Waterman P, Chen JT, Soobader MJ, Subramanian SV, Carson R. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas--the Public Health Disparities Geocoding Project. American journal of public health 2002;92:1100-2.

24. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. American journal of epidemiology 2002;156:471-82.

25. Centers for Disease C, Prevention. Licensure of a 13-valent pneumococcal conjugate vaccine (PCV13) and recommendations for use among children - Advisory Committee on Immunization Practices (ACIP), 2010. MMWR Morbidity and mortality weekly report 2010;59:258-61.

26. Cohen S, Doyle WJ, Baum A. Socioeconomic status is associated with stress hormones. Psychosom Med 2006;68:414-20.

27. Centers for Disease Control and Prevention. Ten great public health achievements--United States, 1900-1999. MMWR Morbidity and mortality weekly report 1999;48:241-3.

28. Centers for Disease C, Prevention. Impact of vaccines universally recommended for children--United States, 1990-1998. MMWR Morbidity and mortality weekly report 1999;48:243-8.

29. Centers for Disease Control and Prevention. Ten great public health achievements--United States, 2001-2010. MMWR Morbidity and mortality weekly report 2011;60:619-23.

30. Centers for Disease Control and Prevention. Ten great public health achievements--worldwide, 2001-2010. MMWR Morbidity and mortality weekly report 2011;60:814-8.

31. Pickering LK, Walton LR. Vaccines in the pipeline: the path from development to use in the United States. Pediatr Ann 2013;42:146-52.

32. Adivsory Committee on Immunization Practices (ACIP). (Accessed 28 October 2014, at http://www.cdc.gov/vaccines/acip/about.html.)

33. Vaccine Effectiveness - How Well Does the Flu Vaccine Work? (Accessed 12 October 2014, at http://www.cdc.gov/flu/about/qa/vaccineeffect.htm.)

34. Park SY, Van Beneden CA, Pilishvili T, et al. Invasive pneumococcal infections among vaccinated children in the United States. The Journal of pediatrics 2010;156:478-83 e2.

35. Centers for Disease Control and Prevention. Invasive pneumococcal disease and 13-valent pneumococcal conjugate vaccine (PCV13) coverage among children aged </=59 months---selected US regions, 2010--2011. MMWR Morbidity and mortality weekly report 2011;60:1477-81.

36. Halloran ME, Haber M, Longini IM, Jr., Struchiner CJ. Direct and indirect effects in vaccine efficacy and effectiveness. American journal of epidemiology 1991;133:323-31.

37. Patel MM, Tate J, Cortese M, et al. The impact of indirect benefits of vaccination on postlicensure vaccine effectiveness estimates: A scenario analysis. Vaccine 2010;28:7987-92.

38. Jackson ML, Nelson JC, Weiss NS, Neuzil KM, Barlow W, Jackson LA. Influenza vaccination and risk of community-acquired pneumonia in immunocompetent elderly people: a population-based, nested case-control study. Lancet 2008;372:398-405.

39. Szilagyi PG, Fairbrother G, Griffin MR, et al. Influenza vaccine effectiveness among children 6 to 59 months of age during 2 influenza seasons: a case-cohort study. Arch Pediatr Adolesc Med 2008;162:943-51.

40. Eisenberg KW, Szilagyi PG, Fairbrother G, et al. Vaccine effectiveness against laboratory-confirmed influenza in children 6 to 59 months of age during the 2003-2004 and 2004-2005 influenza seasons. Pediatrics 2008;122:911-9.

41. Flannery B, Thaker SN, Clippard J, et al. Interim estimates of 2013-14 seasonal influenza vaccine effectiveness - United States, February 2014. MMWR Morbidity and mortality weekly report 2014;63:137-42.

42. Whitney CG, Pilishvili T, Farley MM, et al. Effectiveness of seven-valent pneumococcal conjugate vaccine against invasive pneumococcal disease: a matched case-control study. Lancet 2006;368:1495-502.

43. Cohn AC, MacNeil JR, Clark TA, et al. Prevention and control of meningococcal disease: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2013;62:1-28.

44. Cochran LW, Black S, Klein NP, Dekker CL, Lewis E, Reingold AL. Vaccine effectiveness against laboratory-confirmed influenza in infants: A matched case control study. Human vaccines 2010;6.

45. Domingues CM, Verani JR, Montenegro Renoiner EI, et al. Effectiveness of ten-valent pneumococcal conjugate vaccine against invasive pneumococcal disease in Brazil: a matched case-control study. The Lancet Respiratory medicine 2014;2:464-71.

46. De Wals P, Deceuninck G, De Serres G, et al. Effectiveness of serogroup C meningococcal polysaccharide vaccine: results from a case-control study in Quebec. Clin Infect Dis 2005;40:1116-22.

47. Bose A, Coen P, Tully J, Viner R, Booy R. Effectiveness of meningococcal C conjugate vaccine in teenagers in England. Lancet 2003;361:675-6.

48. Vazquez M, LaRussa PS, Gershon AA, Steinberg SP, Freudigman K, Shapiro ED. The effectiveness of the varicella vaccine in clinical practice. N Engl J Med 2001;344:955-60.

49. Cohen C, von Mollendorf C, de Gouveia L, et al. Effectiveness of 7-valent pneumococcal conjugate vaccine against invasive pneumococcal disease in HIV-infected and -uninfected children in south africa: a matched case-control study. Clin Infect Dis 2014;59:808-18.

50. Estimated Hib and pneumococcal deaths for children under 5 years of age, 2000. (Accessed 17 June, 2014, at http://www.who.int/immunization/monitoring_surveillance/burden/estimates/Pneumo_hib_2000/en/.)

51. ABCs Report: Streptococcus pneumoniae, 1999. (Accessed 17 June, 2014, at http://www.cdc.gov/abcs/reports-findings/survreports/spneu99.html.)

52. Black S, Shinefield H, Fireman B, et al. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. The Pediatric infectious disease journal 2000;19:187-95.

53. Dagan R, Givon-Lavi N, Zamir O, et al. Reduction of nasopharyngeal carriage of Streptococcus pneumoniae after administration of a 9-valent pneumococcal conjugate vaccine to toddlers attending day care centers. J Infect Dis 2002;185:927-36.

54. Dagan R, Givon N, Yagupsky P, Porat N, Janco J, I C. Effect of a 9-valent pneumococcal vaccine conjugated to CRM197 on nasopharyngeal (NP) carriage of vaccine type and non-vaccine type S. pneumoniae strains among day care center (DCC) attendees. 38th Interscience Conference on Antimicrobial Agents and Chemotherapy; 1998; San Diego, CA.

55. Mbelle N, Huebner RE, Wasas AD, Kimura A, Chang I, Klugman KP. Immunogenicity and impact on nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. J Infect Dis 1999;180:1171-6.

56. Rennels MB, Edwards KM, Keyserling HL, et al. Safety and immunogenicity of heptavalent pneumococcal vaccine conjugated to CRM197 in United States infants. Pediatrics 1998;101:604-11.

57. O'Brien KL, Steinhoff MC, Edwards K, Keyserling H, Thoms ML, Madore D. Immunologic priming of young children by pneumococcal glycoprotein conjugate, but not polysaccharide, vaccines. The Pediatric infectious disease journal 1996;15:425-30.

58. Dagan R, Muallem M, Melamed R, Leroy O, Yagupsky P. Reduction of pneumococcal nasopharyngeal carriage in early infancy after immunization with tetravalent pneumococcal vaccines conjugated to either tetanus toxoid or diphtheria toxoid. The Pediatric infectious disease journal 1997;16:1060-4.

59. Obaro SK, Huo Z, Banya WA, et al. A glycoprotein pneumococcal conjugate vaccine primes for antibody responses to a pneumococcal polysaccharide vaccine in Gambian children. The Pediatric infectious disease journal 1997;16:1135-40.

60. Vandenheede H, Vanroelen C, Gadeyne S, De Grande H, Deboosere P. Household-based socioeconomic position and diabetes-related mortality among married and cohabiting persons: findings from a census-based cohort study (Flanders, 2001-2010). Journal of epidemiology and community health 2013;67:765-71.

61. King JC, Jr., Vink PE, Farley JJ, et al. Comparison of the safety and immunogenicity of a pneumococcal conjugate with a licensed polysaccharide vaccine in human immunodeficiency virus and non-human immunodeficiency virus-infected children. The Pediatric infectious disease journal 1996;15:192-6.

62. Centers for Disease C, Prevention. Direct and indirect effects of routine vaccination of children with 7-valent pneumococcal conjugate vaccine on incidence of invasive pneumococcal disease--United States, 1998-2003. MMWR Morbidity and mortality weekly report 2005;54:893-7.

63. Lexau CA, Lynfield R, Danila R, et al. Changing epidemiology of invasive pneumococcal disease among older adults in the era of pediatric pneumococcal conjugate vaccine. Jama 2005;294:2043-51.

64. Kyaw MH, Lynfield R, Schaffner W, et al. Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae. N Engl J Med 2006;354:1455-63.

65. Poehling KA, Talbot TR, Griffin MR, et al. Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. Jama 2006;295:1668-74.

66. Pilishvili T, Lexau C, Farley MM, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis 2010;201:32-41.

67. Moore MR, Gertz RE, Jr., Woodbury RL, et al. Population snapshot of emergent Streptococcus pneumoniae serotype 19A in the United States, 2005. J Infect Dis 2008;197:1016-27.

68. Centers for Disease C, Prevention. Invasive pneumococcal disease in young children before licensure of 13-valent pneumococcal conjugate vaccine - United States, 2007. MMWR Morbidity and mortality weekly report 2010;59:253-7.

69. Review of the Post Marketing Plans (PMP) for Prevnar13. Food and Drug Administration, 2010. (Accessed 27 October 2014, at http://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM206264.pdf.)

70. Vaccines, Blood & Biologics: Prevnar13. (Accessed 27 October 2014, at http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm201667.htm.)

71. Schuck-Paim C, Taylor R, Lindley D, Klugman KP, Simonsen L. Use of near-real-time medical claims data to generate timely vaccine coverage estimates in the US: the dynamics of PCV13 vaccine uptake. Vaccine 2013;31:5983-8.

72. NIS Table Data for 2012. (Accessed 12 October 2014, at http://www.cdc.gov/vaccines/imz-managers/coverage/nis/child/data/tables-2012.html.)

73. 09PRT/8166: Evaluating the effectiveness of a 13-valent pneumococcal conjugate vaccine among children. The Lancet, 2009. (Accessed 11 June, 2014, at http://www.thelancet.com/protocol-reviews/09PRT-8166.)

74. Foundational Health Measures: Disparities. (Accessed 28 October 2014, at http://www.healthypeople.gov/2020/about/foundation-health-measures/Disparities.)

75. Iwane MK, Chaves SS, Szilagyi PG, et al. Disparities between black and white children in hospitalizations associated with acute respiratory illness and laboratory-confirmed influenza and

respiratory syncytial virus in 3 US counties--2002-2009. American journal of epidemiology 2013;177:656-65.

76. Williams DR, Yan Y, Jackson JS, Anderson NB. Racial Differences in Physical and Mental Health: Socio-economic Status, Stress and Discrimination. J Health Psychol 1997;2:335-51.

77. Andrulis DP. Access to care is the centerpiece in the elimination of socioeconomic disparities in health. Annals of internal medicine 1998;129:412-6.

78. Flannery B, Schrag S, Bennett NM, et al. Impact of childhood vaccination on racial disparities in invasive Streptococcus pneumoniae infections. Jama 2004;291:2197-203.

79. Wortham JM, Zell ER, Pondo T, et al. Racial disparities in invasive Streptococcus pneumoniae infections, 1998-2009. Clin Infect Dis 2014;58:1250-7.

80. Muhammad RD, Oza-Frank R, Zell E, et al. Epidemiology of invasive pneumococcal disease among high-risk adults since the introduction of pneumococcal conjugate vaccine for children. Clin Infect Dis 2013;56:e59-67.

81. Payne AB, Link-Gelles R, Azonobi I, et al. Invasive pneumococcal disease among children with and without sickle cell disease in the United States, 1998 to 2009. The Pediatric infectious disease journal 2013;32:1308-12.

82. Smith PJ, Nuorti JP, Singleton JA, Zhao Z, Wolter KM. Effect of vaccine shortages on timeliness of pneumococcal conjugate vaccination: results from the 2001-2005 National Immunization Survey. Pediatrics 2007;120:e1165-73.

83. National Center for Education Statistics. Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress. In: National Center for Education Statistics, ed.2012.

84. Williams DR. Race, socioeconomic status, and health. The added effects of racism and discrimination. Ann N Y Acad Sci 1999;896:173-88.

85. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: one size does not fit all. Jama 2005;294:2879-88.

86. Braveman P, Cubbin C, Marchi K, Egerter S, Chavez G. Measuring socioeconomic status/position in studies of racial/ethnic disparities: maternal and infant health. Public health reports 2001;116:449-63.

87. Shavers VL. Measurement of socioeconomic status in health disparities research. Journal of the National Medical Association 2007;99:1013-23.

88. Geographic Terms and Concepts - Census Tract. (Accessed 17 October 2014, at https://www.census.gov/geo/reference/gtc/gtc_ct.html.)

89. Silverman SL. From randomized controlled trials to observational studies. The American journal of medicine 2009;122:114-20.

90. Black N. Why we need observational studies to evaluate the effectiveness of health care. Bmj 1996;312:1215-8.

91. Howard SC, Algra A, Warlow CP, Rothwell PM. Potential consequences for recruitment, power, and external validity of requirements for additional risk factors for eligibility in randomized controlled trials in secondary prevention of stroke. Stroke; a journal of cerebral circulation 2006;37:209-15.

92. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365:82-93.

93. Baxter R, Lee J, Fireman B. Evidence of bias in studies of influenza vaccine effectiveness in elderly patients. J Infect Dis 2010;201:186-9.

94. De Serres G. Selection bias in the comparison between adjuvanted and unadjuvanted influenza vaccine effectiveness. Vaccine 2013.

95. Fukushima W, Hayashi Y, Mizuno Y, et al. Selection bias in evaluating of influenza vaccine effectiveness: a lesson from an observational study of elderly nursing home residents. Vaccine 2008;26:6466-9.

96. Connally NP, Yousey-Hindes K, Meek J. Selection of neighborhood controls for a population-based Lyme disease case-control study by using a commercial marketing database. American journal of epidemiology 2013;178:276-9.

97. Olson SH, Mignone L, Harlap S. Selection of control groups by using a commercial database and random digit dialing. American journal of epidemiology 2000;152:585-92.

98. Ross JA, Spector LG, Olshan AF, Bunin GR. Invited commentary: Birth certificates--a best control scenario? American journal of epidemiology 2004;159:922-4; discussion 5.

99. Hatch EE, Kleinerman RA, Linet MS, et al. Do confounding or selection factors of residential wiring codes and magnetic fields distort findings of electromagnetic fields studies? Epidemiology 2000;11:189-98.

100. Mezei G, Kheifets L. Selection bias and its implications for case-control studies: a case study of magnetic field exposure and childhood leukaemia. International journal of epidemiology 2006;35:397-406.

101. Exposure to power-frequency magnetic fields and the risk of childhood cancer. UK Childhood Cancer Study Investigators. Lancet 1999;354:1925-31.

102. Schuz J, Grigat JP, Brinkmann K, Michaelis J. Residential magnetic fields as a risk factor for childhood acute leukaemia: results from a German population-based case-control study. International journal of cancer Journal international du cancer 2001;91:728-35.

103. Ma X, Buffler PA, Layefsky M, Does MB, Reynolds P. Control selection strategies in case-control studies of childhood diseases. American journal of epidemiology 2004;159:915-21.

104. Schuchat A, Hilger T, Zell E, et al. Active bacterial core surveillance of the emerging infections program network. Emerg Infect Dis 2001;7:92-9.

105. Active Bacterial Core surveillance (ABCs). (Accessed 24 October 2014, at http://www.cdc.gov/abcs/index.html.)

106. American Community Survey: Information Guide. (Accessed 20 October 2014, at http://www.census.gov/acs/www/Downloads/ACS_Information_Guide.pdf.)

107. Rothman K, Greenland S, Lash T. Modern Epidemiology. 3 ed. Philadelphia, PA: Lippincott, Williams & Wilkins; 2008.

108. 2013 Active Bacterial Core Surveillance (ABCs) Case Report. 2013. (Accessed 16 June, 2014, at http://www.cdc.gov/abcs/downloads/ABCs-case-rpt-form.pdf.)

109. US Standard Birth Certificate. (Accessed 20 October 2014, at http://www.cdc.gov/nchs/data/dvs/birth11-03final-acc.pdf.)

110. Kotelchuck M. The Adequacy of Prenatal Care Utilization Index: its US distribution and association with low birthweight. American journal of public health 1994;84:1486-9.

111. Kotelchuck M. An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. American journal of public health 1994;84:1414-20.

112. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15:615-25.

113. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10:37-48.

114. Textor J, Hardt J, Knuppel S. DAGitty: a graphical tool for analyzing causal diagrams. Epidemiology 2011;22:745.

115. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology 2014;25:473-84.

116. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. Journal of the Royal Statistical Society Series A 2011;174:369-86.

117. Frangakis C. The calibration of treatment effects from clinical trials to target populations. Clinical trials 2009;6:136-40.

118. Weisberg HI, Hayden VC, Pontes VP. Selection criteria and generalizability within the counterfactual framework: explaining the paradox of antidepressant-induced suicidality? Clinical trials 2009;6:109-18.

119. Moore MR, Link-Gelles R, Schaffner W, et al. Effectiveness of 13-valent pneumococcal conjugate vaccine for prevention of invasive pneumococcal disease in children in the USA: a matched case-control study. Lancet Respiratory Medicine (submitted).

120. Phelan JC, Link BG. Controlling disease and creating disparities: a fundamental cause perspective. The journals of gerontology Series B, Psychological sciences and social sciences 2005;60 Spec No 2:27-33.

121. Krieger N, Chen JT, Kosheleva A, Waterman PD. Not just smoking and high-tech medicine: socioeconomic inequities in US mortality rates, overall and by race/ethnicity, 1960-2006. Int J Health Serv 2012;42:293-322.

122. Krieger N, Singh N, Chen JT, et al. Why history matters for quantitative target setting: Long-term trends in socioeconomic and racial/ethnic inequities in US infant death rates (1960-2010). J Public Health Policy 2015;36:287-303.

123. Diez-Roux AV, Kiefe CI, Jacobs DR, Jr., et al. Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. Annals of epidemiology 2001;11:395-405.

124. Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. Epidemiologic reviews 2004;26:104-11.

125. Taber DR, Chriqui JF, Powell LM, Perna FM, Robinson WR, Chaloupka FJ. Socioeconomic Differences in the Association Between Competitive Food Laws and the School Food Environment. J Sch Health 2015;85:578-86.

126. Feinglass J, Rydzewski N, Yang A. The socioeconomic gradient in all-cause mortality for women with breast cancer: findings from the 1998 to 2006 National Cancer Data Base with follow-up through 2011. Annals of epidemiology 2015;25:549-55.

127. Agarwal S, Menon V, Jaber WA. Residential zip code influences outcomes following hospitalization for acute pulmonary embolism in the United States. Vasc Med 2015.

128. Diez Roux A, Schwartz S, Susser E. Ecological variables, ecological studies, and multilevel studies in public health research. 4 ed. Ann Arbor, MI: The Oxford Textbook of Public Health; 2002.

129. Boom JA, Tate JE, Sahni LC, et al. Effectiveness of pentavalent rotavirus vaccine in a large urban population in the United States. Pediatrics 2010;125:e199-207.

130. McTiernan A, Thomas DB, Whitehead A, Noonan E. Efficient selection of controls for multi-centered collaborative studies of rare diseases. American journal of epidemiology 1986;123:901-4.

131. Zhao Z, Smith PJ. Trends in vaccination coverage disparities among children, United States, 2001-2010. Vaccine 2013;31:2324-7.

132. Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January–June 2014. National Center for Health Statistics, 2014. at http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201412.pdf.)

133. Whitney CG, Zhou F, Singleton J, Schuchat A, Centers for Disease C, Prevention. Benefits from immunization during the vaccines for children program era - United States, 1994-2013. MMWR Morbidity and mortality weekly report 2014;63:352-5.

134. Vinikoor LC, Messer LC, Laraia BA, Kaufman JS. Reliability of variables on the North Carolina birth certificate: a comparison with directly queried values from a cohort study. Paediatric and perinatal epidemiology 2010;24:102-12.

135. Northam S, Knapp TR. The reliability and validity of birth certificates. Journal of obstetric, gynecologic, and neonatal nursing : JOGNN / NAACOG 2006;35:3-12.

136. Zollinger TW, Przybylski MJ, Gamache RE. Reliability of Indiana birth certificate data compared to medical records. Annals of epidemiology 2006;16:1-10.

137. Hannan EL. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. JACC Cardiovascular interventions 2008;1:211-7.

138. Adams-Campbell LL, Ahaghotu C, Gaskins M, et al. Enrollment of African Americans onto clinical treatment trials: study design barriers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2004;22:730-4.

139. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. British journal of cancer 2014;110:551-5.

140. Landrine H, Corral I. Advancing research on racial-ethnic health disparities: improving measurement equivalence in studies with diverse samples. Frontiers in public health 2014;2:282.

141. Kangovi S, Barg FK, Carter T, Long JA, Shannon R, Grande D. Understanding why patients of low socioeconomic status prefer hospitals over ambulatory care. Health affairs 2013;32:1196-203.

142. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology 2007;18:805-35.

143. Link-Gelles R, Westreich D, Aiello AE, et al. Bias with respect to socioeconomic status: a closer look at zip code matching in a pneumococcal vaccine effectiveness study. American Journal of Epidemiology (submitted) 2016.

144. US Department of Health and Human Services. US Standard Certificate of Live Birth. 2003.

145. The Public Health Disparities Geocoding Project. (Accessed 21 October 2014, at https://www.hsph.harvard.edu/thegeocodingproject/webpage/pdfs/absms_table_1.pdf.)

146. Westreich D, Cole SR, Funk MJ, Brookhart MA, Sturmer T. The role of the c-statistic in variable selection for propensity score models. Pharmacoepidemiology and drug safety 2011;20:317-20.

147. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. Journal of health services research & policy 1999;4:112-21.

148. Carvalho PC, Fischer JS, Perales J, Yates JR, Barbosa VC, Bareinboim E. Analyzing marginal cases in differential shotgun proteomics. Bioinformatics 2011;27:275-6.

149. Huang SS, Johnson KM, Ray GT, et al. Healthcare utilization and cost of pneumococcal disease in the United States. Vaccine 2011;29:3398-412.