Title

Bias with respect to socioeconomic status: A closer look at zip code matching in a pneumococcal vaccine effectiveness study

Authors

Link-Gelles, Ruth
Westreich, Daniel
Aiello, Allison E
et al.

Publication Date

2016-12-01

DOI

10.1016/j.ssmph.2016.08.005

Peer reviewed

SSM -Population Health 2 (2016) 587–594

Contents lists available at ScienceDirect

SSM -Population Health

journal homepage: www.elsevier.com/locate/ssmph

Article

Bias with respect to socioeconomic status: A closer look at zip code matching in a pneumococcal vaccine effectiveness study

Ruth Link-Gelles [a,b,*], Daniel Westreich [b], Allison E. Aiello [b], Nong Shang [b], David J. Weber [b], Corinne Holtzman [c], Karen Scherzinger [d], Arthur Reingold [e], William Schaffner [f], Lee H. Harrison [g], Jennifer B. Rosen [h], Susan Petit [i], Monica Farley [j,p], Ann Thomas [k], Jeffrey Eason [l], Christine Wigen [m], Meghan Barnes [n], Ola Thomas [m], Shelley Zansky [o], Bernard Beall [a], Cynthia G. Whitney [a], Matthew R. Moore [a]

[a] National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, MS C-25, Atlanta, GA 30329, USA
[b] Department of Epidemiology, University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, NC 27516, USA
[c] Minnesota Department of Health, P.O. Box 64975, St. Paul, MN 55164, USA
[d] Institute for Public Health, University of New Mexico, Emerging Infections Program, 1601 Randolph SE, Suite 1005, Albuquerque, NM 87106, USA
[e] California Emerging Infections Program, 360 22nd Street, Suite 750, Oakland, CA 94612, USA
[f] Department of Preventive Medicine, Vanderbilt University School of Medicine, 1500 21st Avenue South, Suite 2600, Nashville, TN 37212, USA
[g] Department of International Health, Johns Hopkins Bloomberg School of Public Health, 201 W. Preston Street, Baltimore, MD 21201, USA
[h] New York City Department of Health and Mental Hygiene, 125 Worth St, New York, NY 10013, USA
[i] Connecticut Department of Public Health, 410 Capitol Avenue, Hartford, CT 06134, USA
[j] Division of Infectious Diseases, Emory University School of Medicine, Atlanta, GA, USA
[k] Oregon Public Health Division and Oregon Emerging Infections Program, 800 NE Oregon Street, Suite 772, Portland, OR 97232, USA
[l] Bureau of Epidemiology, Utah Department of Health, 288 N 1460 W, Salt Lake City, UT 84116, USA
[m] County of Los Angeles Department of Public Health, 313 N Figueroa St, Los Angeles, CA 90012, USA
[n] Colorado Emerging Infections Program, 4300 Cherry Creek Drive South, Denver, CO 80246, USA
[o] New York State Department of Health, Corning Tower, Room 651, Empire State Plaza, Albany, NY 12237, USA
[p] Infection Disease Section, Medical Specialty Care Service Line, Atlanta VA Medical Center, 1670 Clairmont Road, Station 152, Decatur, GA 30033, USA

ARTICLE INFO

Article history:
Received 24 March 2016
Received in revised form
5 August 2016
Accepted 8 August 2016

Keywords:
Socioeconomic status
PCV13
Pneumococcus
Pneumococcal vaccine
Vaccine effectiveness
Matched case-control

ABSTRACT

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

Low socioeconomic status (SES) is frequently found to be associated with poor health outcomes, despite substantial advances in prevention and treatment of disease (Phelan & Link, 2005; Braveman, Cubbin, Marchi, Egerter, & Chavez, 2001; Janssen, Boyce, Simpson, & Pickett, 2006). 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 (Phelan & Link, 2005; Braveman et al., 2001; Janssen et al., 2006; Burton, Flannery, & Bennett, 2010; Cohen, Doyle, & Baum, 2006; Iwane,

* Corresponding author at: Centers for Disease Control and Prevention, 1600 Clifton Road, MS C-25, Atlanta, GA 30329, USA.
E-mail address: ruth.linkgelles@cdc.hhs.gov (R. Link-Gelles).

http://dx.doi.org/10.1016/j.ssmph.2016.08.005

Chaves, & Szilagyi, 2013; Lantz, House, Mero, & Williams, 2005; Spicer, Thomas, Holst, Baughman, & Farley, 2014). 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 (Phelan & Link, 2005; Janssen et al., 2006; Pardo-Crespo, Narla, & Williams, 2013; Krieger, Chen, Waterman, & Rehkopf, 2003; Krieger, Chen, Kosheleva, & Waterman, 2012; Krieger, Singh, & Chen, 2015; VanderWeele & Robinson, 2014). 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 (Braveman et al., 2001; Lantz et al., 2005; Krieger, Chen, Waterman, Rehkopf, & Subramanian, 2005; Braveman, Cubbin, & Egerter, 2005; Williams, 1999; Shavers, 2007).

When SES is 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. 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 (Diez-Roux, Kiefe, & Jacobs, 2001; Diez Roux, 2004).

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 (Taber et al., 2015; Feinglass, Rydzewski, & Yang, 2015; Agarwal, Menon, & Jaber, 2015). 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 (Krieger et al., 2003; Feinglass et al., 2015). 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 (Diez Roux, 2004; Diez Roux, Schwartz, & Susser, 2002).

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 (Iwane et al., 2013; Spicer et al., 2014; Cutts, Orenstein, & Bernier, 1992; Hutchins, Baughman, Orr, Haley, & Hadler, 2004; Hutchins, Jiles, & Bernier, 2004; Walker, Smith, & Kolasa, 2014; Smith & Stevenson, 2008; Boom, Tate, & Sahni, 2010; Whitney, Pilishvili, & Farley, 2006; Cochran et al., 2010; McTiernan, Thomas, Whitehead, & Noonan, 1986). 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 (Boom et al., 2010; Whitney et al., 2006; Cochran et al., 2010; Krieger et al., 2002).

We were concerned about confounding by individual SES in a zip code-matched case-control study of 13-valent pneumococcal conjugate vaccine (PCV13) effectiveness (Moore et al., 2016). PCV13 was licensed for use in children in the US in February 2010 and replaced the effective, but more limited, 7-valent vaccine (PCV7) (09PRT/8166, 2009; Centers for Disease Control and Prevention, 2010). 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 invasive pneumococcal disease (IPD) and is therefore of concern as a potential confounder (Cutts et al., 1992; Hutchins et al., 2004; Walker et al., 2014; Smith & Stevenson, 2008; Flannery, Schrag, & Bennett, 2004; Wortham, Zell, & Pondo, 2014; Smith, Nuorti, Singleton, Zhao, & Wolter, 2007). 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 (Moore et al., 2016). 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 (Active Bacterial Core surveillance (ABCs), 2014). 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 (09PRT/8166, 2009). 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 (Moore et al., 2016). 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 (Whitney et al., 2006; Pilishvili, Zell, & Farley, 2010). 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 within 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 our analytic model of vaccine effectiveness, we constructed a directed acyclic graph

(DAG) (VanderWeele & Robinson, 2014; Greenland, Pearl, & Robins, 1999; Hernan, Hernandez-Diaz, & Robins, 2004). DAGs, or causal diagrams, are an increasingly utilized tool in epidemiology for identifying variables that should be controlled for to obtain unbiased effect estimates (Greenland et al., 1999). Briefly, investigators begin by putting all variables potentially related to the cause and effect relationship under study on a graph, connected via unidirectional arrows showing causal relationships between the variables. The graph enables investigators to explicitly show assumptions about the underlying causal structure and to identify confounding pathways that should be controlled (the "minimally sufficient subset" of confounders) (Greenland et al., 1999; VanderWeele & Robins, 2007). We used DAGitty.net (version 2.2) software(Textor, Hardt, & Knuppel, 2011) 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 included both census tract and individual SES measures to determine if zip code was an adequate proxy (i.e., if controlling for zip code alone would block all confounding pathways between census tract, individual SES, and PCV13/IPD). 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 at the census tract-level from 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-child'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 (Centers for Disease Control & Prevention, 2010)), breastfeeding (ever vs. never), presence of other children in the household, and household crowding (>2 people per room) (Spicer et al., 2014; Smith & Stevenson, 2008; Wortham et al., 2014; Pilishvili et al., 2010; Zhao & Smith, 2013). Because only parents/guardians of enrolled case- and control-children were interviewed, these variables were not available for unenrolled control-children.

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 (Kotelchuck, 1994; Kotelchuck, 1994)), maternal race/ethnicity, maternal education, and insurance status at birth (U.S. Standard Birth Certificate, 2014). 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 (American Community Survey: Information Guide, 2014). 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 (rerun without Colorado and Maryland), 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 (Whitney et al., 2006). All models were conditional logistic regression to account for the matched design and to calculate the matched odds ratio (with each case and its matched controls composing a single strata). An absolute difference in the vaccine effectiveness of five percentage points 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 1040 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, 2774 (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 1). Of the 12,050 eligible controls, the majority, 8690 (72.1%), came from the same zip code as their matched case. However, only 1250 (10.4%) came from the same census tract as their matched case. A similar pattern was seen among enrolled controls, with, 1921 (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 eligible controls (Table 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 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 (Table 2). 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. Median income, crowding, and income inequality (as measured by Gini Index) were also similar between cases and eligible controls (Table 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 2). The birth

Table 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	2774
Unique zip codes	557	1209	577
Unique Census Tracts	632	4835	2126

Table 2
Characteristics of eligible cases and matched controls. Data come from (a) birth certificates, (b) American Community Survey, or (c) the parent interview/medical provider follow-up.

2 (a) Birth certificate Characteristic	**Enrolled cases (*n*=661)**	**Eligible controls (*n*=12,050)**	**Enrolled controls (*n*=2774)**
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
2 (b) American Community Survey			
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* (%)[a]			
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)

Table 2 (*continued*)

2 (a) Birth certificate Characteristic	Enrolled cases (*n*=661)	Eligible controls (*n*=12,050)	Enrolled controls (*n*=2774)
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)
2(c) Parent interview/medical provider follow-up			
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

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

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%) (Moore et al., 2016). 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 percentage points 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) (Wireless Substitution, 2014; Klosky et al., 2009; Kramer, Wilkins, & Goulet, 2009). 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.

Table 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.[a]

Model[b]	VE (95% CI)		PCV13-type discordant pairs[c]	Absolute % difference in VE vs. unadjusted for PCV13-types
	PCV13	NVT[a]		
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%

¥ 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.

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

[b] 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).

[c] 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.

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 (Whitney, Zhou, Singleton, & Schuchat, 2014).

Our study had limitations. We were not able to conduct interviews with unenrolled controls and had some missing data even for those children whose parents were interviewed (e.g., for household income) 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 (Vinikoor, Messer, Laraia, & Kaufman, 2010; Northam and Knapp, 2006; Zollinger, Przybylski, & Gamache, 2006). 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. Finally, control-children who moved from the ABCs catchment area between birth and their matched case-child's culture date were not included. It is not known what percentage of controls who could not be contacted had moved out of the catchment area.

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. Future research could focus on understanding the principal components underlying enrollment and improving ways to locate and contact eligible children. 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.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Funding

This work was supported by the Centers for Disease Control and Prevention.

Conflicts

Dr. Lynfield and Ms. Holtzman report grants from Centers for Disease Control and Prevention, during the conduct of the study. In addition, Dr. Lynfield is an editor for a book on Infectious Disease Surveillance published by Blackwell-Wiley (royalty money donated to Minnesota Department of Health) during the conduct of the study. Dr. Schaffner reports personal fees from Merck, Pfizer, the Cleveland Clinic, and Novavax outside the submitted work. For all other authors, we declare that we have no conflicts of interest.

Acknowledgments

We acknowledge the following individuals for their contributions to the establishment and maintenance of the ABCs system and expanded surveillance areas. California Emerging Infections Program: Susan Brooks, Hallie Randel. Colorado Emerging Infections Program: Benjamin White, Deborah Aragon, Meghan Barnes, Jennifer Sadlowski. Connecticut Emerging Infections Program: Matt Cartter, Carmen Marquez, Michelle Wilson. Georgia Emerging Infections Program: Sasha Harb, Nicole Romero, Stephanie Thomas, Amy Tunali, Wendy Baughman. Maryland Emerging Infections Program: Joanne Benton, Terresa Carter, Rosemary Hollick, Kim Holmes, Andrea Riner, Kathleen Shutt, Catherine Williams. Minnesota Emerging Infections Program: David Boxrud, Larry Carroll, Kathy Como-Sabetti, Richard Danila, Ginny Dobbins, Liz Horn, Catherine Lexau, Kerry MacInnes, Megan Sukalski, Billie Juni. New Mexico Emerging Infections Program: Kathy Angeles, Lisa Butler, Sarah Khanlian, Robert Mansmann, Megin Nichols, Sarah Shrum. New York Emerging Infections Program: Suzanne McGuire, Sal Currenti, Eva Pradhan, Jessica Nadeau, Rachel Wester, Kathryn Woodworth. Oregon Emerging Infections Program: Mark Schmidt, Jamie Thompson, Tasha Poissant, Keenan Williamson. Tennessee Emerging Infections Program: Brenda Barnes, Karen Leib, Katie Dyer, Lura McKnight, Tiffanie Markus. Los Angeles Epidemiology and Laboratory Capacity Site: Christine Benjamin, Nicole Green, Anali Gutierrez, David Jensen, Annelise Lupica, Laurene Mascola, Olamide Thomas, Christine Wigen. New York City Department of Health and Mental Hygiene: Sarah Borderud, Katherine Lawrence, Orin Forde, Andrea Farnham, Ifeoma Ezeoke. Utah Epidemiology and Laboratory Capacity Site: Jonathan Anderson, Susan Mottice, Kristina Russell, and Amanda Whipple. CDC: Tamara Pilishvili, Ryan Gierke, Karrie-Ann Toews, Emily Weston, Londell McGlone, Gayle Langley, Bernard Beall, Delois Jackson, Joy Rivers, Logan Sherwood, Hollis Walker.

References

Phelan, J. C., & Link, B. G. (2005). Controlling disease and creating disparities: a fundamental cause perspective. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences*, *60*, 27–33.

Braveman, P., Cubbin, C., Marchi, K., Egerter, S., & Chavez, G. (2001). Measuring socioeconomic status/position in studies of racial/ethnic disparities: maternal and infant health. *Public Health Reports*, *116*, 449–463.

Janssen, , I, Boyce, W. F., Simpson, K., & Pickett, W. (2006). Influence of individual- and area-level measures of socioeconomic status on obesity, unhealthy eating, and physical inactivity in Canadian adolescents. *American Journal of Clinical Nutrition*, *83*, 139–145.

Burton, D. C., Flannery, B., Bennett, N. M., et al. (2010). Socioeconomic and racial/ethnic disparities in the incidence of bacteremic pneumonia among US adults. *American Journal of Public Health*, *100*, 1904–1911.

Cohen, S., Doyle, W. J., & Baum, A. (2006). Socioeconomic status is associated with stress hormones. *Psychosom Med*, *68*, 414–420.

Iwane, M. K., Chaves, S. S., Szilagyi, P. G., et al. (2013). 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*, *177*, 656–665.

Lantz, P. M., House, J. S., Mero, R. P., & Williams, D. R. (2005). Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. *Journal of Health Social Behaviour*, *46*, 274–288.

Spicer, J. O., Thomas, S., Holst, A., Baughman, W., & Farley, M. M. (2014). Socioeconomic and racial disparities of pediatric invasive pneumococcal disease after the introduction of the 7-valent pneumococcal conjugate vaccine. *The PediatriC Infectious Disease Journal*, *33*, 158–164.

Pardo-Crespo, M. R., Narla, N. P., Williams, A. R., et al. (2013). 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*, *67*, 305–310.

Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2003). 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*, *93*, 1655–1671.

Krieger, N., Chen, J. T., Kosheleva, A., & Waterman, P. D. (2012). Not just smoking and high-tech medicine: socioeconomic inequities in U.S. mortality rates, overall and by race/ethnicity, 1960–2006. *International Journal of Health Service*, *42*, 293–322.

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

VanderWeele, T. J., & Robinson, W. R. (2014). On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*, *25*, 473–484.

Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2005). Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *American Journal of Public Health*, *95*, 312–323.

Braveman, P. A., Cubbin, C., Egerter, S., et al. (2005). Socioeconomic status in health research: one size does not fit all. *Jama*, *294*, 2879–2888.

Williams, D. R. (1999). Race, socioeconomic status, and health. The added effects of racism and discrimination. *Annals of the New York Academy of Sciences*, *896*, 173–188.

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, *99*, 1013–1023.

Diez-Roux, A. V., Kiefe, C. I., Jacobs, D. R., Jr., et al. (2001). Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Annals of Epidemiology*, *11*, 395–405.

Diez Roux, A. V. (2004). The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiologic Reviews*, *26*, 104–111.

Taber, D. R., Chriqui, J. F., Powell, L. M., Perna, F. M., Robinson, W. R., & Chaloupka, F. J. (2015). Socioeconomic differences in the association between competitive food laws and the school food environment. *Journal of School Health*, *85*, 578–586.

Feinglass, J., Rydzewski, N., & Yang, A. (2015). 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*, *25*, 549–555.

Agarwal, S., Menon, V., & Jaber, W. A. (2015). Residential zip code influences outcomes following hospitalization for acute pulmonary embolism in the United States. *Vascular Medicine*.

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

Cutts, F. T., Orenstein, W. A., & Bernier, R. H. (1992). Causes of low preschool immunization coverage in the United States. *Annual Review of Public Health*, *13*, 385–398.

Hutchins, S. S., Baughman, A. L., Orr, M., Haley, C., & Hadler, S. (2004). Vaccination levels associated with lack of measles transmission among preschool-aged populations in the United States, 1989-1991. *Journal of Infectious Diseases*, *189* (Suppl 1), S108–S115.

Hutchins, S. S., Jiles, R., & Bernier, R. (2004). Elimination of measles and of disparities in measles childhood vaccine coverage among racial and ethnic minority populations in the United States. *Journal of Infectious Diseases*, *189*(Suppl 1), S146–S152.

Walker, A. T., Smith, P. J., & Kolasa, M. (2014). Centers for disease C, prevention. Reduction of racial/ethnic disparities in vaccination coverage, 1995–2011. *MMWR Surveillance Summaries*, *63*(Suppl 1), 7–12.

Smith, P. J., & Stevenson, J. (2008). 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*, *27*, 4107–4118.
Boom, J. A., Tate, J. E., Sahni, L. C., et al. (2010). Effectiveness of pentavalent rotavirus vaccine in a large urban population in the United States. *Pediatrics*, *125*, e199–e207.
Whitney, C. G., Pilishvili, T., Farley, M. M., et al. (2006). Effectiveness of seven-valent pneumococcal conjugate vaccine against invasive pneumococcal disease: a matched case-control study. *Lancet*, *368*, 1495–1502.
Cochran, L. W., Black, S., Klein, N. P., Dekker, C. L., Lewis, E., & Reingold, A. L. (2010). Vaccine effectiveness against laboratory-confirmed influenza in infants: a matched case control study. *Human Vaccines* (p. 6), 6.
McTiernan, A., Thomas, D. B., Whitehead, A., & Noonan, E. (1986). Efficient selection of controls for multi-centered collaborative studies of rare diseases. *American Journal of Epidemiology*, *123*, 901–904.
Krieger, N., Waterman, P., Chen, J. T., Soobader, M. J., Subramanian, S. V., & Carson, R. (2002). 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*, *92*, 1100–1102.
Moore, M.R., Link-Gelles, R., Schaffner, W., et al., 2016. Effectiveness of 13-valent pneumococcal conjugate vaccine for prevention of invasive pneumococcal disease in children in the USA: a matched case-control study, Lancet Respir. Med. 4 (5), 399–406.
09PRT/8166 (2009). Evaluating the effectiveness of a 13-valent pneumococcal conjugate vaccine among children. *The Lancet*. Accessed 11 June, 2014, at ⟨http://wwwthelancet.com/protocol-reviews/09PRT-8166⟩.
Centers for Disease Control and Prevention (2010). Invasive pneumococcal disease in young children before licensure of 13-valent pneumococcal conjugate vaccine – United States, 2007. *MMWR*, *59*, 253–257.
Flannery, B., Schrag, S., Bennett, N. M., et al. (2004). Impact of childhood vaccination on racial disparities in invasive Streptococcus pneumoniae infections. *JAMA*, *291*, 2197–2203.
Wortham, J. M., Zell, E. R., Pondo, T., et al. (2014). Racial disparities in invasive Streptococcus pneumoniae infections, 1998-2009. *Clin Infect Dis*, *58*, 1250–1257.
Smith, P. J., Nuorti, J. P., Singleton, J. A., Zhao, Z., & Wolter, K. M. (2007). Effect of vaccine shortages on timeliness of pneumococcal conjugate vaccination: results from the 2001–2005 National Immunization Survey. *Pediatrics*, *120*, e1165–e1173.
Active Bacterial Core surveillance (ABCs). (Accessed 24 October 2014, at ⟨http://www.cdc.gov/abcs/index.html⟩.).
Pilishvili, T., Zell, E. R., Farley, M. M., et al. (2010). Risk factors for invasive pneumococcal disease in children in the era of conjugate vaccine use. *Pediatrics*, *126* e9-17.
Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*, 37–48.
Hernan, M. A., Hernandez-Diaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, *15*, 615–625.
VanderWeele, T. J., & Robins, J. M. (2007). Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *American Journal of Epidemiology*, *166*, 1096–1104.
Textor, J., Hardt, J., & Knuppel, S. (2011). DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology*, *22*, 745.
Centers for Disease Control & Prevention (2010). Licensure of a 13-valent pneumococcal conjugate vaccine (PCV13) and recommendations for use among children, 2010. *The Morbidity and Mortality Weekly Report*, *59*, 258–261.
Zhao, Z., & Smith, P. J. (2013). Trends in vaccination coverage disparities among children, United States, 2001-2010. *Vaccine*, *31*, 2324–2327.
Kotelchuck, M. (1994). The Adequacy of Prenatal Care Utilization Index: its US distribution and association with low birthweight. *American Journal of Public Health*, *84*, 1486–1489.
Kotelchuck, M. (1994). An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. *American Journal of Public Health*, *84*, 1414–1420.
U.S. Standard Birth Certificate. (Accessed 20 October 2014, at ⟨http://www.cdc.gov/nchs/data/dvs/birth11-03final-acc.pdf⟩.).
American Community Survey: Information Guide. (Accessed 20 October 2014, at ⟨http://www.census.gov/acs/www/Downloads/ACS_Information_Guide.pdf⟩.).
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⟩.).
Klosky, J. L., Tyc, V. L., Lawford, J., Ashford, J., Lensing, S., & Buscemi, J. (2009). *Predictors of non-participation in a randomized intervention trial to reduce environmental tobacco smoke (ETS) exposure in pediatric cancer patients*. *Pediatric Blood & Cancer*, 52.
Kramer, M. S., Wilkins, R., Goulet, L., et al. (2009). Investigating socio-economic disparities in preterm birth: evidence for selective study participation and selection bias. *Paediatric and Perinatal Epidemiology*, *23*, 301–309.
Whitney, C. G., Zhou, F., Singleton, J., & Schuchat, A. (2014). Centers for Disease Control and Prevention. Benefits from immunization during the vaccines for children program era - United States, 1994-2013. *The Morbidity and Mortality Weekly Report*, *63*, 352–355.
Vinikoor, L. C., Messer, L. C., Laraia, B. A., & Kaufman, J. S. (2010). Reliability of variables on the North Carolina birth certificate: a comparison with directly queried values from a cohort study. *Paediatric and Perinatal Epidemiology*, *24*, 102–112.
Northam, S., & Knapp, T. R. (2006). The reliability and validity of birth certificates. *Journal of Obstetric, Gynecologic, and Neonatal Nursing: JOGNN/NAACOG*, *35*, 3–12.
Zollinger, T. W., Przybylski, M. J., & Gamache, R. E. (2006). Reliability of Indiana birth certificate data compared to medical records. *Annals of Epidemiology*, *16*, 1–10.