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