

Housing data-based socioeconomic index and risk of invasive pneumococcal disease: an exploratory study

M. D. JOHNSON¹, S. H. URM^{1,2}, J. A. JUNG¹, H. D. YUN¹, G. E. MUNITZ³,
C. TSIGRELIS⁴, L. M. BADDOUR⁵ AND Y. J. JUHN^{1*}

¹ Department of Pediatrics, University of Utah, Salt Lake City, UT, USA

² Department of Preventive Medicine, College of Medicine, Inje University, Busan, South Korea

³ Mayo Medical School, Mayo Clinic, Rochester, MN, USA

⁴ Department of Medicine, Division of Infectious Diseases, Cooper Medical School of Rowan University, Camden, NJ, USA

⁵ College of Medicine, Department of Medicine, Mayo Clinic, Rochester, MN, USA

Received 15 November 2011; Final revision 11 April 2012; Accepted 25 May 2012;
first published online 3 July 2012

SUMMARY

We previously developed and validated an index of socioeconomic status (SES) termed HOUSES (housing-based index of socioeconomic status) based on real property data. In this study, we assessed whether HOUSES overcomes the absence of SES measures in medical records and is associated with risk of invasive pneumococcal disease (IPD) in children. We conducted a population-based case-control study of children in Olmsted County, MN, diagnosed with IPD (1995–2005). Each case was age- and gender-matched to two controls. HOUSES was derived using a previously reported algorithm from publicly available housing attributes (the higher HOUSES, the higher the SES). HOUSES was available for 92·3% ($n=97$) and maternal education level for 43% ($n=45$). HOUSES was inversely associated with risk of IPD in unmatched analysis [odds ratio (OR) 0·22, 95% confidence interval (CI) 0·05–0·89, $P=0·034$], whereas maternal education was not (OR 0·77, 95% CI 0·50–1·19, $P=0·24$). HOUSES may be useful for overcoming a paucity of conventional SES measures in commonly used datasets in epidemiological research.

Key words: Child, education, epidemiology, housing, immunization, invasive pneumococcal disease, social class, socioeconomic status, *Streptococcus pneumoniae*.

INTRODUCTION

Socioeconomic status (SES) or class is an important determinant of health outcomes. Its influences on health have been widely documented in the USA and

other countries [1–3]. There are a growing number of studies suggesting childhood SES influences health during adulthood [4–6]. However, accurate and reliable data on SES are difficult to obtain, often not available in commonly used data sources for epidemiological research such as medical records, and may be an important source of confounding [7–10]. Epidemiological associations which could have been either partially or fully accounted for by SES are often not properly examined and rarely reported [9, 10].

* Author for correspondence: Y. J. Juhn, MD, MPH, Associate Professor, Department of Pediatric and Adolescent Medicine, Mayo Clinic College of Medicine, Mayo Clinic, 200 First Street SW, Baldwin 3, Rochester, MN 55905, USA.
(Email: juhn.young@mayo.edu)

To overcome a lack of availability of SES measures, we recently developed and validated an index of SES based on real property data termed HOUSES (housing-based index of socioeconomic status) [7, 8]. Although validated as an index of SES, HOUSES had not yet been applied in a medical record-based retrospective study. It was unknown whether HOUSES overcomes the absence of SES measures in medical records or is associated with specific health outcomes. Recent studies show that housing attributes recorded in public databases of real property data (e.g. square footage, assessed housing value, and number of bedrooms and bathrooms, etc.) are correlated with health indicators [11, 12]. For example, children living with parents who do not smoke will have higher urine cotinine levels if they live in multiunit housing vs. detached housing [12]. Moreover, housing insecurity defined by crowdedness and frequent moving has been associated with poor health, lower weight and developmental risk in young children [13].

We postulate that as a composite index of housing attributes, HOUSES is associated with health outcomes. Thus, we assessed whether HOUSES is measurable when conventional SES measures are not available in medical records. Further, we assessed whether HOUSES is associated with risk of invasive pneumococcal disease (IPD), an infection which is associated with SES and is a significant health problem in children [14–18]. We therefore conducted a population-based case-control study to assess whether there was an association between HOUSES and risk of IPD. As a secondary aim, we examined the association between individual housing attributes and risk of IPD.

METHODS

The Institutional Review Boards of both Mayo Clinic and Olmsted Medical Center approved this study. It is a population-based, retrospective case-control study of 105 children living in Olmsted County, MN between 1996 and 2005.

Study population and setting

Olmsted County is the 8th most populous county in Minnesota with 141 360 residents. Rochester, MN is the seat of Olmsted county with 71% of Olmsted County's residents living within Rochester's city limits [19]. The majority (89.4%) of Olmsted County's residents are Caucasian [20] and the population is

similar to other US Caucasian populations, with the exception of a higher rate of employment in the healthcare industry [20, 21]. Olmsted County, is an ideal study setting to conduct a population-based epidemiological study because medical care is virtually self-contained within the community with a unified medical record system for research between two medical centres for the past 90 years. The medical record contains all inpatient and outpatient data of Olmsted County residents, indexed in an automated form since 1935 under the auspices of the Rochester Epidemiology Project [20, 21].

Case ascertainment of IPD cases

The details for identification of IPD cases in children have been reported previously [22]. Briefly, IPD was defined as the isolation of *Streptococcus pneumoniae* from a normally sterile site, from an Olmsted County resident aged < 18 years. Cases of IPD were identified from 1 January 1995 to 31 December 2005 from the only two microbiology laboratories in Olmsted County: Mayo Clinic and Olmsted Medical Center. Non-Olmsted County residents and subjects without research authorization for use of medical records for research review were excluded.

Selection of controls

Two age- and gender-matched controls were randomly selected for each case from children living in Olmsted County who never developed IPD during the study period. A database of potential controls was generated using the Rochester Epidemiology Project database and matched for birthday (within 6 months), gender and registration year (within 1 year of index date of case). Controls had a documented visit within 1 year of the index date of their corresponding IPD cases. This ensured that cases and controls had similar follow-up duration. The same exclusion criteria were applied to controls as cases. The broad matching criteria for age were used out of concern for potential difficulties identifying controls meeting all matching criteria.

Measurement of SES

Individual-level SES variables were derived from two sources: medical records and address-linked real property data. For each case and control, maternal educational levels were abstracted from the entire

medical record including social history questionnaires. Address-linked real property data were obtained from the government of Olmsted County and matched to the subject's address in the medical record (geocoding). These data are maintained by Olmsted County for the purposes of property taxation and are publicly available. Data points abstracted from the property data and used to formulate HOUSES include ownership status, lot size of the housing unit, square footage of the housing unit, residential status (whether a housing unit is in a residential zoning), number of bathrooms, number of bedrooms and building value (assessor's estimated building value). Higher HOUSES is associated with greater SES. Briefly, in formulating HOUSES, we conceptualized that a composite index derived from size, type, ownership status, and value of housing unit combined with neighbourhood socioeconomic characteristics (census tract level) would reflect an individual's SES. We conducted a principal-component factor analysis of seven individual housing characteristics including ownership and six neighborhood SES measures. The final factor analysis results showed square footage, assessed housing value, and number of bedrooms and bathrooms were single factors reflecting the underlying latent construct (i.e. SES). The HOUSES index was formulated by summing up *z*-score transformations of each of the four housing characteristics. This corrects skewness of the distribution by standardization, making comparisons easier among different communities. The detailed analytical procedures for formulation of the HOUSES index have been reported previously [8].

Other covariates

From social history questionnaires in the medical record, the number of children living in the house was collected. We calculated crowdedness based on the number of children at home and the number of bedrooms and square footage from real property data. From the medical record, we collected pneumococcal vaccination status and the presence of high-risk conditions for the development of IPD (i.e. the Advisory Committee on Immunization Practices' recommended pneumococcal vaccine-eligible conditions). Type of air circulation system for cooling and heating (i.e. central air circulation system *vs.* other such as electric or installable air conditioning unit) was collected from address-linked real property data.

Statistical analysis

We compared the availability of maternal educational level based on medical records to the availability of HOUSES based on address-linked real property data to determine whether HOUSES overcomes the absence of SES measures in medical records. Data were fitted to a conditional logistic regression model to determine the association of HOUSES and maternal education with risk of IPD. Odds ratios (ORs) by matched analysis and their corresponding 95% confidence intervals (CIs) were calculated. In this analysis, because of the small sample size, we conducted data analysis using HOUSES and maternal educational levels as a continuous variable in a multivariate model. To allow full interpretation, we report the results by both matched and unmatched analysis [23]. As a secondary analysis, we assessed the association between air circulation system (separately for air conditioning and heating) and risk of IPD. Further, Pearson's correlation coefficient was used to determine correlation of HOUSES with maternal education level. All statistical significance was tested at a two-sided alpha error of 0.05.

RESULTS

The characteristics of cases and controls are summarized in Table 1. Of 35 IPD cases identified during the study period, 20 (57%) were male and 26 (79%) were Caucasian. Of those who had IPD, nine (26%) had a high-risk condition for the development of IPD.

Comparison of availability of SES measures

Of 105 eligible cases and controls, HOUSES was determined in 92.3% ($n=97$). Eight subjects could not undergo geocoding. In contrast, maternal education levels were available for 43% of subjects ($n=45$) ($P<0.001$). The mean family income at a census-tract level (area-level SES measure) was available for 93% ($n=98$).

Association between SES measures and risk of IPD

In univariate matched analysis, HOUSES was associated with risk of IPD but in a multivariate matched analysis, the association between HOUSES and the risk of IPD only approached statistical significance (see Tables 1 and 2). For comparison, multivariate unmatched analysis results showed that

Table 1. Characteristics of study subjects and their risk of developing of IPD

Characteristics	Cases (<i>n</i> = 35)	Controls (<i>n</i> = 70)	OR (95% CI)	<i>P</i> value
Gender (<i>n</i> = 105), <i>n</i> (%), male	20 (57)	40 (57)	1.00 (0.44–2.27)	1.00
Age, years (<i>n</i> = 105)				
Mean (s.d.)	4.34 (4.98)	4.31 (5.00)	1.47 (0.35–6.16)	0.63
Median (IQR)	2.39 (1.16–5.10)	2.31 (0.91–5.23)		
Ethnicity (<i>n</i> = 91), <i>n</i> (%)				
Caucasians	26 (79)	51 (88)	Referent	0.57
Non-Caucasians	7 (21)	7 (12)	1.38 (0.46–4.17)	
Maternal education, years (<i>n</i> = 45)				
Mean (s.d.)	13.30 (1.89)	14.49 (2.06)	0.77 (0.50–1.19)	0.24
Median (IQR)	12 (12–14)	14 (12–16)		
High risk conditions prior to index date (<i>n</i> = 105), <i>n</i> (%)	9 (26)	2 (3)	9.00 (1.95–41.65)	0.005
Parental cigarette smoking at home (<i>n</i> = 52), <i>n</i> (%)	4 (29)	12 (32)	1.30 (0.32–5.28)	0.72
Pneumococcal vaccination prior to index date (<i>n</i> = 105), <i>n</i> (%)				
PCV7	7 (20)	13 (19)	—	—
PPV23	2 (6)	0		
Both	1 (3)	0		
Neither	25 (71)	57 (81)		
Pneumococcal vaccinations prior to index date (<i>n</i> = 105), <i>n</i> (%)*				
PCV7 or PPV23	10 (28.57)	13 (18.57)	Referent	0.094
Neither	25 (71.43)	57 (81.43)	6.43 (0.73–56.71)	
Housing characteristics				
HOUSES index in <i>z</i> score (<i>n</i> = 97)				
Mean (s.d.)	−0.20 (1.16)	0.26 (0.14)	0.22 (0.05–0.89)	0.034
Median (IQR)	0.30 (−0.04 to 0.33)	0.32 (0.24–0.34)		
Range	−5.25 to 0.39	−0.40 to 0.40		
HOUSES index in tertile, (<i>n</i> = 97), <i>n</i> (%)†				
HOUSES group1 (lowest SES)	15 (47)	18 (28)	Referent	
HOUSES group2 (middle SES)	9 (28)	23 (35)	0.47 (0.17–1.32)	0.151
HOUSES group3 (highest SES)	8 (25)	24 (37)	0.40 (0.14–1.14)	0.088
Household income in census tract in 1999 (<i>n</i> = 98) (US\$1000)				
Mean (s.d.)	53 523 (21 357)	55 550 (17 460)	0.99 (0.97–1.02)	0.52
Median (IQR)	51 094 (37 446–62 250)	55 483 (40 278–63 583)		
Air conditioning using central air circulation (<i>n</i> = 99), <i>n</i> (%)				
Yes	14 (42.42)	41 (62.12)	0.39 (0.15–0.97)	
No	19 (57.58)	25 (37.88)	Referent	0.042
Heating using central air circulation (<i>n</i> = 99), <i>n</i> (%)				
Yes	17 (51.52)	49 (74.24)	0.39 (0.16–0.95)	
No	16 (48.48)	17 (25.76)	Referent	0.038
Number of rooms per child (<i>n</i> = 72)				
Mean (s.d.)	2.35 (0.95)	2.43 (1.06)	0.78 (0.42–1.46)	0.44
Median (IQR)	2 (1.58–3.16)	2.5 (1.58–3.16)		
Amount of square feet per child (<i>n</i> = 72)				
Mean (s.d.)	12 866 (30 388)	4415 (8270)		
Median (IQR)	2724 (1542–5444)	2648 (1676–3629)	1.00 (0.99–1.00)	0.15

IPD, Invasive pneumococcal disease; OR, odds ratio; CI, confidence interval; IQR, interquartile range; SES, socioeconomic status. Parameter estimates were based on matched analysis.

* Parameter estimate for children who had received any pneumococcal vaccinations (PCV7, PPV23, or both PCV7 and PPV23) using children who had never received pneumococcal vaccinations.

† Parameter estimates based on unmatched logistic regression.

Table 2. *Multivariate model evaluating the role of the HOUSES index in defining risk of IPD, controlling for covariates*

	Adjusted OR* (95% CI)	P value
HOUSES in continuous variable (z score)	0.16 (0.02–1.11)	0.063
High-risk conditions for IPD		
Yes	15.71 (1.92–128.64)	0.010
No	Referent	
Pneumococcal vaccinations prior to index date		
Yes	0.15 (0.01–1.97)	0.147
No	Referent	

IPD, Invasive pneumococcal disease; OR, odds ratio; CI, confidence interval.

* Parameter estimate was based on matched analysis (a conditional logistic regression model).

HOUSES was inversely associated with risk of IPD adjusting for high-risk conditions associated with the development of IPD and use of pneumococcal vaccination (adjusted OR 0.12, 95% CI 0.02–0.76, $P=0.024$). Individual maternal education was not associated with risk of IPD ($n=45$, OR 0.97, 95% CI 0.74–1.28, $P=0.86$). HOUSES and individual maternal education were not correlated ($r=0.19$, $P=0.22$). Mean income at the census-tract level was not associated with risk of IPD ($n=98$, OR 0.99, 95% CI 0.97–1.02, $P=0.52$).

Association between other housing attributes and risk of IPD

Of the 33 eligible IPD cases, 14 children (42%) lived in a house with a central air circulation system for cooling *vs.* 41 (62%) of the 66 controls (unadjusted OR 0.39, 95% CI 0.15–0.97, $P=0.042$). This trend was true for heating system as well (52% *vs.* 74%, OR 0.39, 95% CI 0.16–0.95, $P=0.038$). The median number of bedrooms and the median square feet per child were not associated with the risk of IPD.

Other variables and risk of IPD

High-risk conditions for IPD were strongly associated with risk of IPD. Pneumococcal vaccinations were associated with a reduced risk of IPD but did not reach statistical significance ($P=0.094$). Race and the smoking of tobacco in the household were not associated with risk of IPD.

DISCUSSION

In this exploratory study, the HOUSES index based on real property data was more commonly available

than maternal education level within the medical record. The HOUSES index could be used as a proxy measure of individual SES.

Although SES is an important determinant of health, SES measures are often not available in commonly used datasets or sources such as medical records, limiting the understanding of the role of SES in health outcomes. To overcome this limitation, we recently developed and validated HOUSES as a measure of SES based on individual housing data, a source of information that is consistently kept by local governments in order to assess property taxes [7, 8].

The present study is our first application of HOUSES in epidemiological research. HOUSES appears to be more available in the medical record than maternal education (92% *vs.* 43%) ($P<0.001$) and has similar availability to mean family income at census-tract level (area-level SES measure) (92% *vs.* 93%). The role of the HOUSES index in matched analysis only approached statistical significance due to limited statistical power (beta error 0.38). However, considering the exploratory nature of this investigation, HOUSES could be used as a measure of SES when conventional measures (maternal education, in this instance) are not available in the dataset. In support of this association, previous studies have shown a significant association between SES measures and risk of IPD [14, 15, 17, 24]. For example, a previous study reported an association between lower maternal education and risk of IPD [18]. The lack of association of IPD with maternal education in our study appears primarily due to a significant number of missing values (*i.e.* absence of documented maternal education in the medical record).

The mechanisms that are operative in the impact of HOUSES on risk of IPD are unknown. We postulate that potential direct and indirect mechanisms are pivotal. One direct mechanism could be related to housing features. For example, crowdedness or closer proximity among family members increases the risk of diseases like pneumococcal infection [25, 26]. Those with smaller living spaces are likely to be at greater risk of transmission and disease. Although the size of the housing unit is likely to be correlated with HOUSES, we did not find a significant association between the number of rooms per child and risk of IPD. Infrequent documentation in the medical record of the number of children in the household illustrates the challenge of obtaining a measure of crowdedness through retrospective review of social histories in medical records.

We found a significant association between air circulation system in housing units and risk of IPD. Children who lived in housing units with central air circulation had a lower risk of IPD compared to those who lived in housing units with non-centrally distributed air circulation (see Table 1). The increased air filtration present in a central air system may filter droplets containing pneumococci and other airborne irritants, allergens, or infectious agents that would affect a child's risk of subsequent pneumococcal infection. A previous study reported that better filtration of household air reduced airborne particles and the need for urgent visits in childhood [27]. Another study by Hoge *et al.* demonstrated a direct association of air flow rates with risk of pneumococcal transmission [28]. We cannot account for the nature of air filtration (e.g. type of filtration, air turnover, etc.) in our subjects' homes. Given the association found here, we suspect that air circulation in homes may play a role in the risk of microbial infection but further study is necessary to determine this association. Alternatively, type of air circulation present in a home may simply be another proxy indicator of SES, as children who lived in housing units with a central air circulation system also had much higher HOUSES (0.31 ± 0.04) compared to those who lived in housing units with non-central air circulation system (-0.16 ± 1.01 , $P < 0.001$). Nonetheless, our study highlights the importance of broader identification of poorly studied risk factors for IPD, especially those related to housing features.

The potential indirect impact of HOUSES on risk of IPD can be conceptualized from the well-documented influence of SES on health outcomes

[29]. Housing is a pivotal place in the day-to-day life of most individuals that reflects socioeconomic position and income, control over life circumstances and access to elements of material, social and human capital. Housing has been shown to be clearly associated with health outcomes [11, 30]. For example, a recent study showed that housing insecurity defined by crowding (>2 people/bedroom or >1 family/residence) and multiple moves (≥ 2 moves within the previous year) is associated with poor health, lower weight, and developmental risk in young children [13]. Although the findings are not directly related to our study results, a recent study showed that offering housing and case management to a population of homeless adults with chronic medical illnesses resulted in fewer hospital days and emergency department visits, compared to usual care for homeless adults [31].

A noteworthy finding in our study is a lack of association between area-based SES measures such as mean family income at census-tract level and risk of IPD. Area-based SES or neighbourhood SES through census data has been commonly used in epidemiological research as a surrogate marker for an individual's SES when SES measures at an individual level are not available [32–34]. Area-based SES measures can be a surrogate measure for individuals' SES in a socioeconomically segregated setting. In a socioeconomically more homogeneous setting like our community [35], area-based measures might result in a significant misclassification bias [36].

We found no association between pneumococcal vaccination and risk of IPD. Since we only collected pneumococcal vaccination status as a binary variable (received *vs.* never received) prior to index date, our study was limited in assessing the effects of herd immunity for non-vaccinated children, vaccine schedule (timing and doses), and serotype replacement. Although we adjusted the results for pneumococcal vaccination status, the potential impact of these factors on our study findings need to be elucidated in future study.

The strengths of our study as an exploration of a novel proxy measure of SES include the epidemiological advantages of our study setting with a self-contained healthcare environment and unified medical records for research. Our study is population-based and enables us to overcome selection bias. HOUSES has been recently validated and has unique advantages as a SES measure, which has been described previously [7, 8].

Being retrospective and structured to explore the utility of HOUSES as a proxy for SES, our study has inherent limitations. We did not have all data related to known risks for IPD such as childcare programme attendance. Some pertinent variables had missing data points such as smoking exposure. HOUSES was formulated based on data that might not be updated when structural changes or other renovation of housing units takes place. Some real property data were missing or unavailable (e.g. newly built housing units not in the database or use of a P.O. Box number). However, this deficiency should be subject to a non-differential misclassification bias not affecting the interpretation of the main results. Our study setting has a predominant white population and thus, the utility of the HOUSES index needs to be assessed in other communities. To its advantage, the HOUSES index was originally validated in two communities including Jackson County, MO with an ethnically more diverse setting [7, 8]. The lack of availability of maternal education data (as well as income and occupation) in the medical records examined in this study may not be reflective of the presence or absence of similar measures in other medical record systems. Our study was limited by a small sample that hindered pertinent analyses (e.g. effect of secular trends of immunizations). This small sample size was also reflected in the lack of significance of maternal education (as a marker of SES) and vaccination status with risk of IPD, both associations which have been demonstrated elsewhere [18]. With further study with larger statistical power, we anticipate being able to obtain a more refined view of the association of housing characteristics with risk for disease.

In conclusion, housing data can be used as a measure of SES in epidemiological research examining the role of SES in health outcomes. Housing features may play an important role in the epidemiology of infectious disease such as IPD. Future study is needed to evaluate the association of HOUSES with other diseases and in other communities.

ACKNOWLEDGEMENTS

We thank the staff of the Pediatric Asthma Epidemiology Research Unit who made this study possible. We also thank Mr Philip Wheeler, the director of the Olmsted-Rochester Planning Department and his staff for supporting the study. This work was supported by the Clinician Scholarly Award from

the Mayo Foundation and was made possible by the Rochester Epidemiology Project (R01-AG34676) from the National Institute on Aging.

DECLARATION OF INTEREST

None.

REFERENCES

1. Berkman LF, Kawachi I. *Social Epidemiology*. New York: Oxford University Press, 2000, pp. xxii, 391.
2. Dowd JB, *et al.* Socioeconomic gradients in immune response to latent infection. *American Journal of Epidemiology* 2008; **167**: 112–120.
3. Mackenbach JP, *et al.* Socioeconomic inequalities in health in 22 European countries. *New England Journal of Medicine* 2008; **358**: 2468–2481.
4. Cohen S, *et al.* Childhood socioeconomic status and adult health. *Annals of the New York Academy of Sciences* 2010; **1186**: 37–55.
5. Maty SC, *et al.* Life-course socioeconomic position and incidence of diabetes mellitus among blacks and whites: the Alameda County Study, 1965–1999. *American Journal of Public Health* 2010; **100**: 137–145.
6. Ziol-Guest KM, *et al.* Early childhood poverty and adult body mass index. *American Journal of Public Health* 2009; **99**: 527–532.
7. Butterfield MC, *et al.* A two-county comparison of the HOUSES index on predicting self-rated health. *Journal of Epidemiology and Community Health* 2011; **65**: 254–259.
8. Juhn YJ, *et al.* Development and initial testing of a new socioeconomic status measure based on housing data. *Journal of Urban Health* 2011; **88**: 933–944.
9. Liberatos P, *et al.* The measurement of social class in epidemiology. *Epidemiologic Reviews* 1988; **10**: 87–121.
10. Van Sickle D, *et al.* Understanding socioeconomic and racial differences in adult lung function. *American Journal of Respiratory and Critical Care Medicine* 2011; **184**: 521–527.
11. Dunn JR, Hayes MV. Social inequality, population health, and housing: a study of two Vancouver neighborhoods. *Social Science and Medicine* 2000; **51**: 563–587.
12. Wilson KM, *et al.* Tobacco-smoke exposure in children who live in multiunit housing. *Pediatrics* 2011; **127**: 85–92.
13. Cutts DB, *et al.* US Housing insecurity and the health of very young children. *American Journal of Public Health* 2011; **101**: 1508–1514.
14. Chen FM, *et al.* Geocoding and linking data from population-based surveillance and the US Census to evaluate the impact of median household income on the epidemiology of invasive *Streptococcus pneumoniae* infections. *American Journal of Epidemiology* 1998; **148**: 1212–1218.
15. Huang SS, *et al.* Community-level predictors of pneumococcal carriage and resistance in young

- children. *American Journal of Epidemiology* 2004; **159**: 645–654.
16. **Nuorti JP, et al.** Cigarette smoking and invasive pneumococcal disease. *New England Journal of Medicine* 2000; **342**: 681–689.
 17. **Pastor P, et al.** Invasive pneumococcal disease in Dallas County, Texas: results from population-based surveillance in 1995. *Clinical Infectious Diseases* 1998; **26**: 590–595.
 18. **Pilishvili T, et al.** Risk factors for invasive pneumococcal disease in children in the era of conjugate vaccine use. *Pediatrics* 2010; **126**: e9–17.
 19. **U.S. Census Bureau.** 2008 Population estimates (http://factfinder.census.gov/home/en/official_estimates_2008.html). Population Estimates Program. Accessed 28 September 2009.
 20. **Bureau USC.** Rochester (city), Minnesota. State & County QuickFacts, 2009.
 21. **Katusic SK, et al.** Potential influence of migration bias in birth cohort studies. *Mayo Clinic Proceedings* 1998; **73**: 1053–1061.
 22. **Tsigrelis C, et al.** Incidence of invasive pneumococcal disease among children after introduction of a 7-valent pneumococcal conjugate vaccine: a population-based study in Olmsted County, Minnesota. *Mayo Clinic Proceedings* 2009; **84**: 871–875.
 23. **Alvan R. Feinstein M.** *Clinical Epidemiology: The Architecture of Clinical Research*. New Haven: W. B. Saunders Company, 1985, pp. 494.
 24. **Whitney CG, et al.** Rethinking recommendations for use of pneumococcal vaccines in adults. *Clinical Infectious Diseases* 2001; **33**: 662–675.
 25. **Bogaert D, et al.** Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infectious Diseases* 2004; **4**: 144–154.
 26. **House T, Keeling MJ.** Household structure and infectious disease transmission. *Epidemiology and Infection* 2009; **137**: 654–661.
 27. **Lanphear BP, et al.** Effects of HEPA air cleaners on unscheduled asthma visits and asthma symptoms for children exposed to secondhand tobacco smoke. *Pediatrics* 2011; **127**: 93–101.
 28. **Hoge CW, et al.** An epidemic of pneumococcal disease in an overcrowded, inadequately ventilated jail. *New England Journal of Medicine* 1994; **331**: 643–648.
 29. **Warnecke RB, et al.** Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. *American Journal of Public Health* 2008; **98**: 1608–1615.
 30. **Laaksonen M, et al.** Home ownership and mortality: a register-based follow-up study of 300,000 Finns. *Journal of Epidemiology and Community Health* 2008; **62**: 293–297.
 31. **Sadowski LS, et al.** Effect of a housing and case management program on emergency department visits and hospitalizations among chronically ill homeless adults: a randomized trial. *Journal of the American Medical Association* 2009; **301**: 1771–1778.
 32. **Goyal NK, et al.** Association of late-preterm birth with asthma in young children: practice-based study. *Pediatrics* 2011; **128**: e830–838.
 33. **Krieger N.** Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. *American Journal of Public Health* 1992; **82**: 703–710.
 34. **Subramanian SV, et al.** Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: a multilevel analysis of Massachusetts births, 1989–1991. *American Journal of Epidemiology* 2006; **164**: 823–834.
 35. **Metropolitan Racial and Ethnic Change – Census 2000.** American Communities Project (<http://mumford1.dyndns.org/cen2000/report.html>). University at Albany, SUNY, 2009. Accessed 28 September 2009.
 36. **Geronimus AT.** Invited commentary: using area-based socioeconomic measures – think conceptually, act cautiously. *American Journal of Epidemiology* 2006; **164**: 835–840; discussion 41–43.