

An Ecological Study of Tuberculosis Transmission in California

Ward P. Myers, MD, MPH, Janice L. Westenhouse, MPH, Jennifer Flood, MD, MPH, and Lee W. Riley, MD

Tuberculosis is a social disease caused by an airborne pathogen with low infectivity. The transmission of tuberculosis depends on human interaction within communities. However, some communities provide a better environment for disease transmission than others. Previous surveillance has documented great disparities in rates of tuberculosis among neighborhoods.¹ These differences depended in part on community-level, ecological risk factors that facilitate transmission—poverty, crowding, and other markers of deprivation have long been associated with increased rates of tuberculosis.^{2,3}

Because of its airborne transmission and societal impact, tuberculosis is closely monitored by local, state, and federal health departments. Cases of tuberculosis are subject to mandatory reporting in all 50 states, the District of Columbia, US dependencies and possessions, and independent nations within the United States (Native American lands).⁴ In addition to ensuring treatment, health departments collect case-specific demographic information (e.g., age, race, foreign-born status) and disease information (e.g., site of infection, drug resistance).⁵ The focus on individual cases, however, neglects the ecological context of this disease. Information about community-level, ecological risk factors for contracting tuberculosis is important for structuring a public health response to this illness.

Ecological data can be obtained by geocoding addresses from reported cases, and then linking these cases to geographic locations such as the census tract. The US Census defines a census tract as a “small, relatively permanent statistical subdivision of a county . . . designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time of establishment. Census tracts average about 4000 inhabitants.”⁶ Every 10 years the US Census collects detailed demographic and socioeconomic information about the US population. When linked to reported tuberculosis cases, this information permits the examination of ecological factors

Objectives. We evaluated the association between ecological factors and rates of tuberculosis within California, using pediatric tuberculosis as an indicator of new transmission.

Methods. Ecological variables such as racial/ethnic distribution, immigration level, education, employment, poverty, and crowding were obtained from the United States Census for each census tract in California. These data were incorporated into a negative binomial regression model with the rate of pediatric tuberculosis disease in each census tract as an outcome variable. Disease rates were obtained by geocoding reported cases. Subsections of the state (San Francisco and Los Angeles) were examined independently.

Results. Census tracts with lower median incomes, more racial/ethnic minorities, and more immigrants had higher rates of pediatric tuberculosis. Other frequently cited risk factors such as overcrowding and unemployment were not associated with increased disease after adjusting for other measures. Risks were comparable across regions, but subtle differences were noted.

Conclusions. The techniques used in this work provide a way to examine a disease within its social context. The results confirmed that tuberculosis in California continues to be a disease of poverty and racial/ethnic minorities. (*Am J Public Health*. 2006;96:685–690. doi:10.2105/AJPH.2004.048132)

that are associated with disease. Use of the census tract has many advantages over the use of other geographic units such as zip codes. Previous work has shown that populations defined by zip codes, being larger and more heterogeneous, give more variable results than census tracts in ecological analysis.⁷

Ecological analysis of tuberculosis is complicated by the disease’s long incubation period. A delay of 30 years or more between infection and clinical disease has been documented,⁸ bringing into question the validity of studies comparing current ecological data to case reports from adults. Cases of tuberculosis in children, compared with cases in adults, have a short delay between infection and onset of clinical disease. The incubation period is limited by the child’s lifespan and, thus, a greater proportion of cases are likely to be primary disease. Cases occurring in children represent recently acquired infection and serve as a surrogate marker for ongoing transmission. For this reason, tuberculosis cases in children are used by state and local health departments to monitor the success of tuberculosis-control activities.

Recent studies have supported the role of ecological risk factors, such as poverty, lack of social capital, and overcrowding, in tuberculosis disease.^{1,7,9–15} Although these studies have used a variety of techniques, there are limited data using exclusively pediatric cases to look at ecological risks for tuberculosis.¹⁶ In this work, we developed a multivariate model for prediction of tuberculosis transmission on the basis of ecological measures and pediatric cases from census tracts in the state of California. Data from California are particularly useful for understanding tuberculosis in the United States. In 2002, California reported 3159 cases of tuberculosis, or 21% of the national total.⁴ Furthermore, much of the United States is now beginning demographic and ethnic shifts that mirror the changes that have occurred in California over the past 10 years.

METHODS

Data Collection: Tuberculosis Cases

Case information was obtained from the California Department of Health Services,

TABLE 1—Ecological Measures Derived From Year 2000 US Census Tract Data

Measure	Operational Definition	Summary File	Census Variable
Demographic			
Asian race	Percentage of population in census tract that self-reports Asian race (1 race only, non-Hispanic)	1	P4
Black race	Percentage of population in census tract that self-reports black race (1 race only, non-Hispanic)	1	P4
Hispanic ethnicity	Percentage of population in census tract that self-reports Hispanic ethnicity	1	P4
Immigration	Percentage of population that was born outside the United States	3	P21
Education: Low attainment	Percentage of persons 25 years and older with less than a high-school diploma	3	P37
Occupation: Unemployment	Percentage of persons aged 16 and older in the labor force who are unemployed	3	P43
Economy: Median income	Median household income for census tract in 1999	3	P53
Housing			
Crowded households	Percentage of households with > 1 person per room	3	H20
Population density	Number of people per square mile	1	P1

Note. P = population subjects; H = housing subjects.

Tuberculosis Control Branch. We analyzed all 3208 cases of tuberculosis in children aged 0 to 14 years that were reported in the 10 years between January 1, 1993, and December 31, 2002. The cases were geocoded, and each case was linked to a census tract from the 2000 US Census. A census tract number was available for 3164 cases (98.6% of total). Use

of nonidentifying case information was approved by the California Department of Health Services, Tuberculosis Control Branch. Tuberculosis case rates per 100 000 person-years were calculated on the basis of populations from the 2000 Census.

The analysis was repeated, limiting tuberculosis cases to children aged 0 to 4 years. As this

approach yielded similar results, the final analysis used cases in patients aged 0 to 14 years.

Data Collection: Ecological Measures

Ecological measures were obtained from the 2000 US Census Web site.^{17,18} Individual variables were selected from summary files 1 and 3 (Table 1). Prior to analysis, variables were chosen that characterized traditional risk factors for transmission of tuberculosis.

Means and standard distributions for predictor variables were calculated for all included census tracts and are reported in Table 2. Variables were standardized to a z scale on the basis of their mean and standard deviation $([X - \text{mean}]/\text{SD})$. This standardization of variability permitted the generation of tuberculosis incidence rate ratios that could be compared among ecological measures (e.g., how does the incidence rate change for a 1-standard-deviation increase in population density, compared with a 1-standard-deviation increase in percentage of residents in poverty?).

Statistical Analysis

The number of pediatric cases for each census tract was modeled as a negative binomial distribution. In contrast to the Poisson distribution, a negative binomial distribution does not assume that the variance equals the mean and allows for more zero counts and overdispersion.¹⁹ Therefore, it is a useful model when the variance of a population exceeds the mean. In this analysis, the model took the form of

$$\log \lambda_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \sigma_\epsilon + \log(\text{pop}_i)$$

for each census tract $[i = 1, \dots, 7018]$, where λ is the expected cases in each census tract, x_j is each standardized ecological measure (with its associated β_j regression coefficient), σ_ϵ is the disturbance or error term, and pop is the 2000 population (age 0–14) in the census tract times the years exposed (times 10, for time exposed). The $\log(\text{pop}_i)$ term has no regression coefficient because it serves as an offset ($\log \lambda_i - \log(\text{pop}_i) = \log[\text{case rate}_i]$). The σ_ϵ term represents error and dispersion in the form of a negative binomial distribution. The exponent of each β_j regression coefficient provides the incidence rate ratio for a

TABLE 2—Descriptive Characteristics of 7018 Census Tracts in California^a

Variable	Mean	SD	Range
Total population per census tract	4819.7	2129.8	3–36 146
Pediatric (0–14 years) population	1109.1	662.8	1–7962
Cases of TB aged 0–14 years from 1993–2002	0.5	1.0	0–15
Pediatric case rate (per 100 000 person-years)	3.8	9.0	0–230
Asian race, %	10.6	12.9	0–95
Black race, %	6.4	11.4	0–91
Hispanic ethnicity, %	31.0	25.5	0–98
Foreign born, %	25.5	16.1	0–100
Lower educated, %	24.4	19.3	0–100
Unemployed, %	7.4	5.6	0–100
Median household income, \$	51 615.7	24 685.4	0–200 001
Living in crowded housing, %	16.9	16.5	0–100
Population density (people/square mile)	8064.3	9205.1	0–156 015

Note. TB = tuberculosis; SD = standard deviation.

^aCalifornia has 7049 census tracts. Prior to analysis, 31 tracts were excluded because their pediatric population was 0. No TB cases were present in the excluded census tracts.

TABLE 3—Univariate and Multivariate Incidence Rate Ratios for Pediatric Tuberculosis and Selected Ecological Measures in the State of California^a

Area-based measure	Univariate Analysis		Intermediate Model		Full Multivariate Analysis		US-Born Stratum Only	
	IRR	95% CI	IRR	95% CI	IRR	95% CI	IRR	95% CI
Asian race	1.08	(1.04, 1.13)			1.22	(1.14, 1.30)	1.18	(1.08, 1.28)
Black race	1.21	(1.17, 1.24)			1.19	(1.14, 1.23)	1.27	(1.22, 1.33) ^d
Hispanic ethnicity	1.56	(1.51, 1.62)			1.25	(1.12, 1.40)	1.38	(1.2, 1.58)
Foreign born	1.65	(1.59, 1.71)			1.26	(1.14, 1.40)	1.26	(1.11, 1.44)
Lower educated	1.67	(1.62, 1.73)	1.13	(1.01, 1.27)	1.12	(0.99, 1.27) ^c	1.13	(0.96, 1.32)
Unemployed	1.44	(1.40, 1.48)	1.04	(0.99, 1.10)	1.02	(0.97, 1.08) ^c	0.97	(0.9, 1.04)
Median income ^b	2.25	(2.11, 2.40)	1.55	(1.42, 1.70)	1.62	(1.48, 1.78)	1.75	(1.55, 1.97)
Crowded housing	1.59	(1.54, 1.64)	1.16	(1.05, 1.28)	0.87	(0.77, 0.98) ^c	0.81	(0.7, 0.93)
Population density	1.32	(1.28, 1.35)	1.07	(1.03, 1.12)	1	(0.95, 1.04) ^c	1	(0.95, 1.06)

Note. IRR = incidence rate ratio; CI = confidence interval.

^aIRRs reflect the change in the incidence rate that occurs when the area-based measure increases by 1 standard deviation. The multivariate analysis holds all other variables constant.

^bStandardized values for median income are inverted. IRR shows change for a 1-standard-deviation decrease in median income.

^cFour variables showed a loss of significance as a risk factor or changed to a mildly protective factor in the model that included all variables.

^dThe IRR for 1 variable in the US-born stratum was outside the 95% confidence intervals for the full multivariate analysis model.

1-standard-deviation change in the corresponding ecological measure.

Each ecological measure was initially examined alone and then as a part of a multivariate model with the other measures. To better understand the loss of significance for many socioeconomic variables in the full model, we analyzed an intermediate multivariate model (without race, ethnicity, or immigration). Incidence rate ratios with 95% confidence intervals for each measure are reported in Table 3. The multivariate model is reported in full. All variables were selected prior to analysis, and none were eliminated.

To assess goodness of fit, deviance residuals were calculated for the multivariate negative binomial model with constant dispersion. Greater than 99% of predicted standardized deviances fell within 2 standard deviations, signifying a very good fit.²⁰ We also modeled the data using a Poisson distribution. Goodness of fit for the Poisson model, however, was poor ($P < .01$). Because additional evidence that the negative binomial model was more appropriate than the Poisson, the likelihood ratio test for dispersion parameter being equal to 0 (in the Poisson model, dispersion parameter equals zero) was $P < .001$. To assess the extent to which the population adjustment factor ($\log[\text{pop}_i]$) might explain the goodness of fit, a correlation coefficient with the number of tuberculosis cases was

calculated ($r^2 = 0.1$). This value was significant (in part because of the larger number of census tracts), but was also too close to the null to solely explain the model's goodness of fit.

To reduce error from the inclusion of tuberculosis cases representing transmission that occurred outside the United States, a stratified analysis was also performed on the basis of country of origin. Analysis was repeated as in the full multivariate model, but the dependent variable included only cases in children born in the United States from each census tract. Incidence rate ratios and 95% confidence intervals for the stratum of cases in children born in the United States are reported in Table 3.

To allow the greater San Francisco and Los Angeles areas to vary independently from each other and the rest of the state, indicator variables were created for corresponding metropolitan statistical areas. The US Census defines a Metropolitan Statistical Area (MSA) as "a core area with a large population nucleus, plus adjacent communities having a high degree of economic and social integration with that core."²¹ Lists of counties and census tracts included in the Los Angeles and San Francisco MSAs are available from the US Census Web site.²¹

To compare differences in the predictive powers of ecological measures between the San Francisco and Los Angeles MSAs, an additional model was generated. This model included

cross-products that allowed coefficients for ecological measures from the 2 MSAs to vary independently. For clarity, cross-products that were less significant than $P = .05$ were removed by backward elimination. The results are depicted in Figure 1.

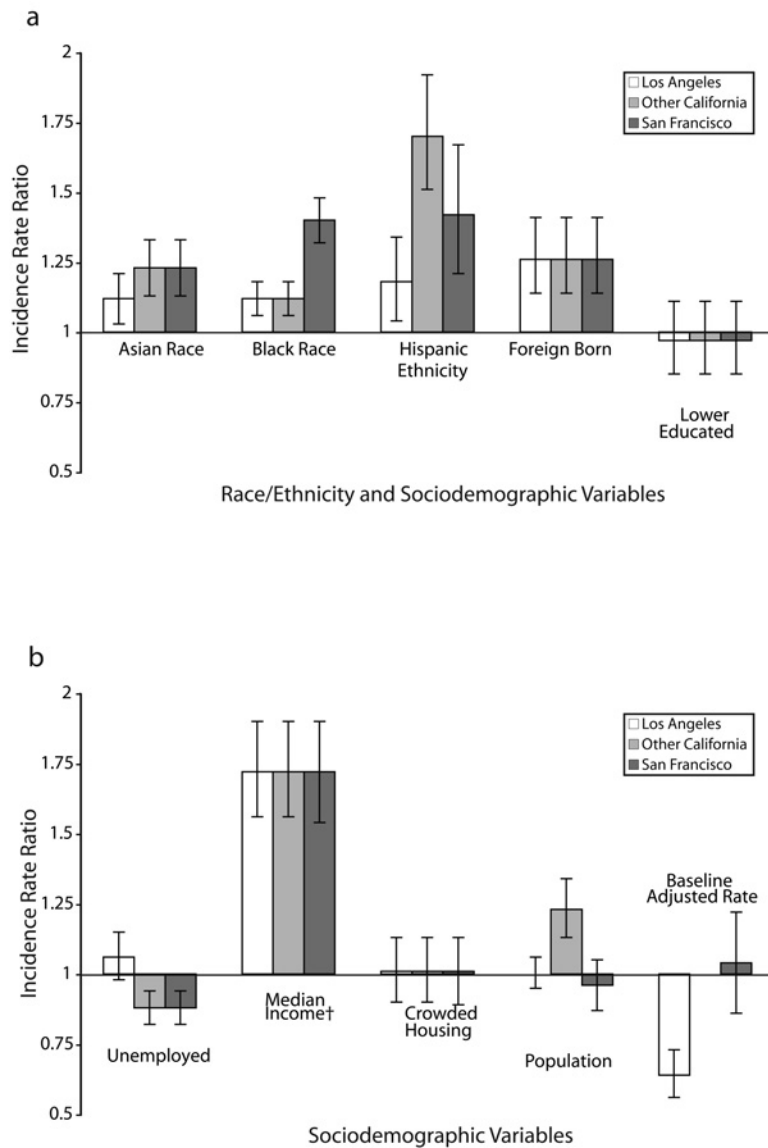
All analyses were conducted using Stata, Version 7.0 (Stata Corp, College Station, Tex).

RESULTS

Over the 10 years included in this study, California had 3208 cases of tuberculosis in its pediatric population. On the basis of the 2000 census, there were 7.78 million individuals aged 0 to 14 years, yielding a crude incidence rate of 4.1 cases per 100 000 person-years. Individual census tracts, however, showed very divergent rates. Incidence rates ranged from 0 to 230 per 100 000 person-years.

Results of univariate, intermediate, multivariate, and stratified models are depicted in Table 3. In the univariate models, the traditional ecological measures were all strongly associated with pediatric tuberculosis. However, when the variables were combined into a single multivariate model, measures such as lower education, unemployment, crowding, and population density became less predictive. Census tracts with lower median incomes and more racial/ethnic minorities and foreign-born individuals were particularly likely to have increased rates of disease when the other variables were held constant. Notably, Asian race appeared to be a greater risk factor in the multivariate model than in the univariate model, and crowded housing became a mildly protective factor in the multivariate model.

The intermediate model suggested that much of the loss of significance for lower education, unemployment, crowding, and population density was attributable to each factor's collinearity with income. The incidence rate ratios in Table 3 are best conceptualized as changes to a hypothetical "average census tract." This average census tract is characterized by the ecological measures shown in Table 2. As the percentage of foreign-born residents increases to 1 standard deviation above the average census tract (approximately from 26% to 42%) the incidence of pediatric tuberculosis would be expected to increase 1.3-fold (assuming all other variables were held constant).



Note. Incidence rate ratios reflect the change in the incidence rate that occurs when the area-based measure increases by 1 standard deviation. Standardized values for median income are inverted. Incidence rate ratio shows change for a 1-standard-deviation decrease in median income.

FIGURE 1—Regional differences in incidence rate ratios for pediatric tuberculosis and ecologic variables, by race/ethnicity (a) and sociodemographic variables (b).

Differences between the US-born stratum and the full multivariate analysis were small but informative. Compared with the full model, census tracts with more Blacks showed an increased risk of disease. Additionally, Asian race seemed less strongly correlated (but still significant), and income became a slightly stronger risk factor.

Figure 1 depicts incidence rates for pediatric tuberculosis that were allowed to vary

independently across regions (i.e., other California [i.e., San Diego, Sacramento, Arcata, and so on], Los Angeles, San Francisco). For many ecological measures, the effects on incidence rates in the different regions were the same. Notable exceptions included differences in the effect of race/ethnicity, unemployment, and population density. In adjusted analysis, San Francisco–area census tracts with more Black residents had higher rates of tuberculosis

than equivalent census tracts in the rest of California. This trend reversed itself for measures of the Hispanic population; increasing Hispanic population was less of a risk factor for disease in Los Angeles and San Francisco than in the rest of California. Population density was an important risk factor for disease in areas other than Los Angeles and San Francisco.

DISCUSSION

General Findings

Using a multivariate model and ecological data from census tract–level geography, we have shown that minority race/ethnicity, immigration, and low income are strong risk factors for new tuberculosis transmission.

This analysis is further support for earlier studies showing that minority race/ethnicity is a risk factor for disease. However, whereas previous research¹¹ has suggested that the risk of race/ethnicity is largely secondary to its correlation with socioeconomic risk factors such as low education, high unemployment, crowding, and high population density, our data did not support this conclusion. In our multivariate analysis, the variability in cases of tuberculosis was better explained by immigration, racial/ethnic minority groupings, and median income than by other variables such as low education, high unemployment, crowded housing, and high population density. The risk of race for disease could be caused by a combination of factors. Although genetic differences have been linked to increased mycobacterial susceptibility,^{22–25} it seems more likely that minority populations are surrogates for larger reservoirs of latent tuberculosis infection. Many minorities have emigrated from regions with higher baseline rates of latent tuberculosis infection, and African Americans have for the past few generations lived disproportionately in urban centers with higher rates of tuberculosis disease. In California, these groups are known to have high rates of active disease.²⁶ Additionally, race and ethnicity are complex social constructs that may be markers for other socioeconomic factors that are difficult to capture in such a model.

Like previous studies, our initial univariate analysis demonstrated that crowding is a risk factor for tubercular disease. However, after adjusting for other factors in the multivariate

model, crowding was noted as developing a protective effect. Part of this change was likely because of its correlation with other variables that better explained the variability in tuberculosis cases (most significantly, low education [$r^2=0.8$], foreign birth [$r^2=0.8$], and Hispanic ethnicity [$r^2=0.6$]). Nevertheless, its reemergence as a significant protective factor suggests some benefit may remain after the negative effects are removed by adjusting for other variables. These results could be explained within the context of recent research on “social capital” as a protective factor for tuberculosis.¹⁵ Crowding may be associated with a more tightly woven social network (i.e., increased social capital) that could protect against disease. Although this research has shown potential, much controversy still exists on the precise measurement of social capital. Further research in this area is clearly warranted.

Our study also supports the association between family income and tuberculosis disease. This finding is consistent with previous research showing a close link between tuberculosis and poverty. Although many racial or ethnic minorities may have higher rates of disease because of historical reservoirs of tuberculosis infection, current levels of economic deprivation are of critical importance.

Regional Differences

The effects for various ecological risk factors were generally consistent across the 3 regions studied. Differences were noted in the risk of population density and in the risk of high racial/ethnic minority populations. The lack of effect for population density in San Francisco and Los Angeles was not unexpected because these 2 regions have uniformly high population densities in comparison to the rest of the state.

Conversely, the regional differences in the risk factors for Black and Hispanic populations were somewhat surprising. These risk differences were not explained by differences in income or recent immigration. The increased rate of tuberculosis noted in predominantly Black census tracts near San Francisco may be at least partially attributable to a known persistent cluster of cases in a Black community in Contra Costa County (part of the San Francisco MSA).²⁷ To assess the impact of this cluster on the general finding, the analysis was repeated, excluding census tracts that corresponded to the

geographic location of the previously mentioned cluster. In the new analysis, the incidence rate ratio decreased slightly, but not completely (1.4 to 1.34), suggesting that the known cluster may reflect a larger trend in the San Francisco area.

Also worthy of additional investigation is the lower baseline rate of tuberculosis in the Los Angeles MSA. After adjusting for variables in the model, the disease rate in Los Angeles was one third lower than expected. This finding is reflected by the crude rate of disease in Los Angeles. Despite Los Angeles having a higher level of diversity and immigration than the rest of the state, the crude rate of pediatric tuberculosis there is roughly the same as that for the state as a whole.

Strengths and Limitations

This analysis, in comparison to other studies of ecological risk factors for tuberculosis, has the advantage of a focus on pediatric cases. This focus permits the results to more directly reflect risk factors for disease transmission. Previous studies of molecular epidemiology have shown that between 4% and 31% of all cases are the result of recent transmission.^{28,29} This means that for a vast majority of all cases, ecological data obtained at the time of disease onset may not represent factors relevant to transmission.

Insufficient data exist for similar estimations for pediatric cases, but it is generally assumed that pediatric cases represent recent transmission. Therefore, analyses using exclusively pediatric cases would be expected to provide results with less misclassification and greater precision. Stratification by country of birth could also theoretically reduce misclassification. Foreign-born children, compared with US-born children, may have been more likely to have acquired their infection overseas. Because the incidence rate ratios from the US-born-only stratum in our analysis are remarkably similar to the results from the full multivariate model, the degree of misclassification may be small.

Research that makes comparisons among different measures of social inequalities is challenging; social measures of income, education, and ethnic heritage all use different units and scales. Furthermore, the shape of each distribution differs, and threshold effects are often unknown. To address these challenges, we standardized variables to a scale on the basis of

mean and variance. Because each independent variable is transformed through addition and multiplication of constants, the magnitude of the resulting incidence rate ratio changes, but its direction and significance do not.

Alternative methods of standardization for predictor variables have been used elsewhere. These include use of raw variables,^{13,15} comparison by quartiles,⁷ use of the relative index of inequality,^{7,30} use of a multiple variable index score,^{7,9} and numerous others.³⁰ Each of these techniques has advantages and disadvantages (the full discussion of which is beyond the scope of this paper). Broadly speaking, these techniques tend to sacrifice either ease of comparison to other variables (in the case of raw scores and log transformations) or clarity of technique (in the case of indices). We propose that although the technique of standardization by mean and variance is by no means perfect, it is an acceptable compromise that permits the clear comparison between ecological measures by nonstatisticians.

This analysis, however, is not without limitations. Collinearity, which occurs when independent variables are identical or very similar to each other, can be problematic in ecological studies. This occurs because aggregated socioeconomic variables tend to be more highly correlated with each other than individual socioeconomic variables.³¹ This effect is magnified in studies with a small number of large heterogeneous regions. Generally speaking, collinearity reduces the significance of a study's findings by increasing the variance of its regression coefficients. This effect may have resulted in the underestimation of the incidence rate ratios reported in this article. We attempted to minimize this effect by analyzing 7018 census tracts and by selecting a variety of differing socioeconomic variables. Additionally, we confirmed that the potential collinearity because of crowding did not destabilize the full model, because the remaining statistics changed only minimally (0.5% to 5%) when crowding was removed from the analysis.

Some misclassification may have occurred through the use of cases reported between January 1993 and December 2002 and ecological measures taken from the 2000 US Census. Although ecological measures for each census tract do shift over time, data from the national census is only collected

every 10 years. Because there are insufficient cases of pediatric tuberculosis each year to analyze individually, this study combined cases over 10 years and used census data that were obtained during that time period.

Aggregated ecological measures, such as those used in this study, are distinct from their analogous individual-level characteristics.³² For example, having a low income affects an individual differently than living in a poor neighborhood. Because the California Department of Health does not currently collect data on income, education, or household crowding from individual tuberculosis cases, we were unable to directly compare ecological and individual-level factors. However, such a multilevel analysis would be informative and should be pursued in future research.

Finally, tuberculosis transmission is a complex process that depends on many factors. The models developed in this investigation include several variables, but other important variables may be missing.

CONCLUSIONS

Ecological studies such as this provide valuable information. Disease transmission within a population depends both on individual host risk factors and community-level risk factors that govern the individual's exposure to disease. This research suggests specific ecological factors that are associated with increased rates of tuberculosis disease. State and local tuberculosis control programs may use this information to identify "at risk" geographic areas that merit increased disease surveillance. These techniques underscore both the importance of geographic information in case reporting and its contribution to the better understanding of disease. ■

About the Authors

Ward P. Myers is with the Children's Hospital, Boston, and Boston Medical Center, Boston, Mass. Janice L. Westenhouse and Jennifer Flood are with the Tuberculosis Control Branch, California Department of Health Services, Berkeley, Calif. Lee W. Riley is with the University of California, Berkeley, School of Public Health.

Request for reprints should be sent to Lee Riley, 140 Warren Hall, Berkeley, CA 94720 (e-mail: lwriley@berkeley.edu). This article was accepted May 30, 2005.

Contributors

W.P. Myers originated the study and led the analysis and writing. J.L. Westenhouse assisted with the data collection

and analysis. J. Flood supervised data collection and analysis. L.W. Riley supervised the analysis and writing.

Acknowledgments

Material and financial support were provided by the California Department of Health Services and the University of California, Berkeley, School of Public Health.

We would like to thank Arthur Reingold for his guidance and assistance in reviewing the article.

Human Participant Protection

No institutional review board protocol approval was needed for this study.

References

- Barr RG, Diez-Roux AV, Knirsch CA, Pablos-Mendez A. Neighborhood poverty and the resurgence of tuberculosis in New York City, 1984–1992. *Am J Public Health*. 2001;91:1487–1493.
- Hetherington HW, Landis M, Opie A. survey to determine the prevalence of tuberculosis infection in school children. *Am Rev Tuberc*. 1929:421.
- Puccini G. *La Bohème*. Milan, Italy: G Ricordi & C; 1896.
- Reported Tuberculosis in the United States, 2002. Atlanta, Ga: US Dept of Health and Human Services, Centers for Disease Control and Prevention; September 2003.
- Krieger N, Chen JT, Ebel G. Can we monitor socioeconomic inequalities in health? A survey of U.S. health departments' data collection and reporting practices. *Public Health Rep*. 1997;112:481–491.
- United States Census Bureau. United States Census 2000 glossary of terms. Available at: <http://www.census.gov/dmd/www/glossary>. Accessed February 20, 2004.
- Krieger N, Waterman PD, Chen JT, Soobader MJ, Subramanian SV. Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based socioeconomic measures—the public health disparities geocoding project (US). *Public Health Rep*. 2003;118:240–260.
- Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen VO, Andersen AB. Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection. *J Infect Dis*. 2002;185:401–404.
- Spence DP, Hotchkiss J, Williams CS, Davies PD. Tuberculosis and poverty. *BMJ*. 1993;307:759–761.
- Doherty MJ, Davies PD, Bellis MA, Tocque K. Tuberculosis in England and Wales. Ethnic origin is more important than social deprivation. *BMJ*. 1995;311:187.
- Cantwell MF, McKenna MT, McCray E, Onorato IM. Tuberculosis and race/ethnicity in the United States: impact of socioeconomic status. *Am J Respir Crit Care Med*. 1998;157(4 pt 1):1016–1020.
- Tocque K, Doherty MJ, Bellis MA, Spence DP, Williams CS, Davies PD. Tuberculosis notifications in England: the relative effects of deprivation and immigration. *Int J Tuberc Lung Dis*. 1998;2:213–218.
- Hawker JI, Bakhshi SS, Ali S, Farrington CP. Ecological analysis of ethnic differences in relation between tuberculosis and poverty. *BMJ*. 1999;319:1031–1034.
- Bennett J, Pitman R, Jarman B, et al. A study of the variation in tuberculosis incidence and possible influential variables in Manchester, Liverpool, Birmingham and

Cardiff in 1991–1995. *Int J Tuberc Lung Dis*. 2001;5:158–163.

- Holtgrave DR, Crosby RA. Social determinants of tuberculosis case rates in the United States. *Am J Prev Med*. 2004;26:159–162.
- Drucker E, Alcabes P, Bosworth W, Sckell B. Childhood tuberculosis in the Bronx, New York. *Lancet*. 1994;343:1482–1485.
- United States Census Bureau. Census 2000 summary file 1: California. Available at: http://factfinder.census.gov/home/saff/main.html?_lang=en. Accessed February 20, 2004.
- United States Census Bureau. Census 2000 summary file 3: California. Available at: http://factfinder.census.gov/home/saff/main.html?_lang=en. Accessed February 20, 2004.
- Byers AL, Allore H, Gill TM, Peduzzi PN. Application of negative binomial modeling for discrete outcomes: a case study in aging research. *J Clin Epidemiol*. 2003;56:559–564.
- Dupont W. *Statistical Modeling for Biomedical Researchers*. Cambridge, England: Cambridge University Press; 2002.
- United States Census Bureau. Geography Division: Census 2000 products. Available at: <http://www.census.gov/geo/www/census2k.html>. Accessed May 15, 2004.
- Ma X, Reich RA, Wright JA, et al. Association between interleukin-8 gene alleles and human susceptibility to tuberculosis disease. *J Infect Dis*. 2003;188:349–355.
- Stead WW, Senner JW, Reddick WT, Lofgren JP. Racial differences in susceptibility to infection by Mycobacterium tuberculosis. *N Engl J Med*. 1990;322:422–427.
- Bellamy R. Genetics and pulmonary medicine. 3. Genetic susceptibility to tuberculosis in human populations. *Thorax*. 1998;53:588–593.
- Abel L, Sanchez FO, Oberti J, et al. Susceptibility to leprosy is linked to the human NRAMP1 gene. *J Infect Dis*. 1998;177:133–145.
- California Department of Health Services. Communicable disease control in California, 2000. Available at: http://www.dhs.ca.gov/ps/dcdc/pdf/CDC2000_Document.pdf. Accessed June 17, 2004.
- Chin DP, Crane CM, Diul MY, et al. Spread of Mycobacterium tuberculosis in a community implementing recommended elements of tuberculosis control. *JAMA*. 2000;283:2968–2974.
- Heldal E, Docker H, Caugant DA, Tverdal A. Pulmonary tuberculosis in Norwegian patients. The role of reactivation, re-infection and primary infection assessed by previous mass screening data and restriction fragment length polymorphism analysis. *Int J Tuberc Lung Dis*. 2000;4:300–307.
- Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med*. 1994;330:1703–1709.
- Wagstaff A, Paci P, van Doorslaer E. On the measurement of inequalities in health. *Soc Sci Med*. 1991;33:545–557.
- Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health*. 1995;16:61–81.
- Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health*. 1994;84:819–824.