Tuberculosis Infection in Oakland, CA 1997 – 2001: At-Risk Populations and Predicting “Hot Spots”

An Analysis Method Combining Kernel Density Estimation Mapping with Regression

Prepared by Urban Strategies Council
Oakland, California
November 2002

The work presented herein was produced pursuant to Health and Human Services Contract to the Urban Institute, (# HHS-100-99-0003) and a Subcontract to Urban Strategies Council (# 06928-019-00-USC-01)

TABLE OF CONTENTS

Executive Summary

Tuberculosis Infection in Oakland, CA 1997 – 2001:
Predicting Vulnerable Populations and "Hot Spots"

Urban Strategies Council (the Council) was one of five National Neighborhood Indicators Partnership sites selected by Urban Institute and the Department of Health and Human Services to participate in the Project to Develop Neighborhood Health Analysis. The project is an effort to expand knowledge of the relationships between characteristics of neighborhoods and health outcomes.

The Alameda County Public Health Department (PHD) is designing community based interventions to reduce the risk of certain diseases, among them tuberculosis (TB), which occur disproportionately in low-income neighborhoods. The Council collaborated with the PHD to investigate the specific social, physical and spatial characteristics that may be associated with TB infection in neighborhoods in Oakland, California. The results of these analyses will help inform both the community and the PHD as they undertake intervention and prevention programs.

Methods:

Using administrative data from the Alameda County Public Health Department's Tuberculosis Control Program and 2000 U.S. Census block-group level data, the Council conducted multivariate and small-area spatial analyses to investigate the correlation between incidences of TB and socio-economic conditions at the neighborhood level in Oakland, CA. We used regression analysis to test the following hypotheses:

- Incidence rates of TB infection are positively correlated with the percent non-white, percent foreign-born population and percent in poverty in neighborhoods.
- When the variables above are held constant, incidence rates of TB infection are positively correlated with the number of occupants per room in housing units.

We also experimented with the kernel density estimation method, which is a relatively new spatial analysis technique, to estimate incidents of TB cases for all locations in Oakland and calculate infection rates. This technique results in density estimates for all parts of a region at any location, and the resulting data were used in the regression analyses. In turn, the results of the regression analyses were used to produce calculations of infection risk, and these new data also were mapped. The results were displayed on contour maps that show the intensity of the occurrence of TB cases and the intensity of risk for TB infection at very specific geographic levels.

Results:

Our findings from the multivariate analysis confirmed our initial hypotheses. We found a positive association between the incidence of TB and high neighborhood concentrations foreign-born and non-white residents, residents in poverty and over-crowded housing. We found a small, but statistically significant, association between over-crowded housing rates and the incidence of TB when all other variables were held constant.

The spatial analysis found highest concentrations of TB cases in two neighborhoods, one in Downtown and the other in the Lower San Antonio area immediately south of Lake Merritt. The estimated infection rate is highest in the Downtown area immediately west of Lake Merritt. The areas of highest risk for TB infection include Downtown, Chinatown, and neighboring areas reaching into West Oakland, as well as the Lower San Antonio and Fruitvale.

This study also found kernel density estimation to be useful method to identify the location, spatial extent, and intensity of disease “hot spots” and risk of infection at the neighborhood level. We were able to identify specific neighborhoods, as opposed to entire Census Tracts, where the risk of TB is greatest. The resulting contour maps may be particularly useful for the design of targeted community intervention and prevention programs.

I. INTRODUCTION

Purposes and Theories for Analyses

Disproportionately high incidence of disease is one of a number of significant health problems facing residents in low-income neighborhoods. Factors such as a lack of preventive health care, blighted and crowded physical environments, environmental contaminants and social norms and networks that do not promote healthy behaviors have all been identified as contributors to poor health outcomes among residents of neighborhoods where poverty is concentrated.

A variety of health outcomes examined in the literature are associated with neighborhoods with high concentrations of people in poverty. (cf. Kawachi, Kennedy and Wilkinson, eds., 1999) Some outcomes may be particularly correlated with specific neighborhood conditions. For example, there is evidence that the incidence of AIDS cases in high poverty neighborhoods is correlated with STD cases (Ruiz, J. D. et al, 2000), and that STD cases are correlated with distance from clinics which provide STD testing. (Sumartojo, E. 2000) Also, there is evidence that tuberculosis infection is associated with crowded living conditions. (Barnes, P.F., et al, 1997) This relationship is explained by the airborne transmission route of tuberculosis.

The problem of geographically specific health conditions and impacts is a continuing concern to the Alameda County Public Health Department (PHD). Not only do they recognize that some diseases and health conditions tend to cluster in geographic areas, they also recognize that a geographic focus for working with residents and other stakeholders to address these conditions is essential to effective intervention. Consequently, the PHD has organized its service delivery system into geographically assigned health teams in order to build ongoing partnerships with residents and community-based organizations for assessment, action planning and implementation of strategies to improve health outcomes in neighborhoods.

A similar rationale undergirds the Partnership for the Public's Health (PPH), a multi-year health improvement initiative funded by The California Endowment. The PPH has funded 13 partnerships of public health departments and community-based organizations throughout California to engage in collaborative assessment, planning, and implementation of health improvement activities. In addition to its neighborhood focus, the PPH has also made the use of "data and information for action" a central element of the local partnerships' work, reasoning that residents and community-based organizations need to participate in the collection, analysis and interpretation of data if they are to be effective in improving health outcomes in their neighborhoods. The Alameda County PHD and three community-based organizations in Oakland form one of the 13 partnerships statewide.

As a participant in the National Neighborhood Indicators Partnership's Project to Develop Neighborhood Health Indicators, and in collaboration with the Alameda County PHD and the PPH, Urban Strategies Council (USC) initially attempted to accomplish the following:

1. Compile three administrative databases containing information on incidences of tuberculosis, sexually transmitted diseases (STDs) and AIDS cases and standardize their geographies to census tract and neighborhood levels.

2. Use multivariate analysis and spatial analysis methods appropriate for small areas to investigate the correlation between incidences of these diseases, socio-economic conditions of the census tract, and the test variables from our hypotheses. It was hoped that this analysis would yield data on all of Oakland's neighborhoods.

Obstacles in Acquiring Data

The Urban Strategies Council (USC) had an agreement with the Alameda County Public Health Department's Community Assessment, Planning and Education (PHD-CAPE) unit that CAPE would acquire the data sets on STDs, AIDS, and tuberculosis cases and provide USC access to them for purposes of this study. The PHD-CAPE unit displayed interested in the project, in particular the use of new methods of spatial analysis, and made efforts to obtain the data for the Urban Strategies Council. However, the unit encountered resistance from the data administrators in charge of the respective data sets. Of concern to these administrators were the issues of confidentiality and the efficacy of these new methods of spatial analysis. Through the PHD-CAPE unit, the Council attempted to convince these data administrators that the new spatial analysis methods not only protected the confidentiality of the data, but were also valid and powerful analysis methods. However, these efforts only proved to be partially successful; the tuberculosis data set was the only one that USC was able to obtain. Thus, the analysis here is limited to tuberculosis cases and does not include STDs and AIDS cases.

Hypotheses

With the tuberculosis incidence data, we set about to test the following hypotheses:

1. Incidence rates of tuberculosis infection are positively correlated with the percent non-white population and percent foreign-born population.
2. Incidence rates of tuberculosis infection are positively correlated with the percent of people in poverty.
3. Controlling for the demographic and economic relationships above, incidence rates of tuberculosis infection are positively correlated with the number of occupants per room in housing units.

These hypotheses were based on established research concerning demographic factors associated with high tuberculosis infection.

II ANALYSIS METHODOLGY

GIS Analyses

The calculation of disease incidence rates in small areas presents statistical and confidentiality problems. Both the numbers of incidents and the population – the numerator and the denominator – may be smaller than statistically valid thresholds. The analyses are further complicated when data are disaggregated by gender, ethnicity, age, or other demographic characteristics. Furthermore, when the analyses focus on small areas, there are issues of ensuring the non-identifiability of cases to maintain confidentiality. To address the problems of analyzing small areas and low-frequency events, we used both traditional data aggregation methods and kernel density estimation methods. We first aggregated five (5) years worth of tuberculosis case data from 1997 to 2001. The resulting data set, which included 456 cases of tuberculosis, was used for all analyses.

We then aggregated these data to several political and administrative geographies in the traditional manner. The geographies to which the data were aggregated include ZIP codes, City Council Districts, Alameda County PHD's Community Health Team Areas and Census Tracts.

Kernel Density Estimation

However, when detailed location information is available for individual cases, it is not necessary to aggregate data into administrative areas or even smaller geographies like census tracts. It is possible to use models that include spatial correlation components and do so without compromising the confidentiality of individuals by using varying levels of resolution to display patterns. Such methods enable a much more precise analysis of the distribution patterns of a disease and the demographic data pertaining to the underlying population.

The Council used one such method, the kernel density estimation method, to estimate incidents of tuberculosis infection through interpolation based on the distances between the cases. Interpolation is a technique for generalizing incident locations to an entire area. It results in density estimates for all parts of a region at any location. Consequently, the data can be displayed on contour maps that show the intensity at all locations. Problems of low incidence rates are also addressed by such interpolations of the data. The tuberculosis infection data provided by the Alameda County Public Health Department were already geocoded to geographic coordinates when the department provided it to the Council. The Council interpolated the data for the entire city of Oakland using the kernel density estimation method. We also included explanatory and contextual variables known to be associated with incidence of tuberculosis infection to produce a risk assessment map for the City of Oakland. These variables, described below, were aggregated by the Census Bureau to the Census Block Group level.

All kernel estimations were produced using CrimeStat II, a spatial statistics package that can analyze incident location data. It provides a variety of tools for the spatial analysis of crime incidents or other point locations. Developed by Ned Levine & Associates, Houston, Texas, with a grant from the National Institute of Justice (NIJ), it is a stand-alone Windows program that can interface with most desktop geographic information systems (GIS).

The data used in creating the contour maps as well as the statistical analyses were produced by CrimeStat II using the single kernel density routine. The method of interpolation used was the normal distribution, which produces an estimate over the entire region. The choice of bandwidth was that of the fixed interval, and our fixed interval was 0.253 miles, the mean nearest neighbor distance between the block group centroids of Oakland. The output units were square miles and absolute densities were calculated. An external file with 69,181 cells for the city of Oakland was created and used as the reference file. Each cell thus represents 22,546.8 square feet since Oakland's area is about 56 square miles. There were 456 cases of tuberculosis infection in the data set and 337 block groups in the city of Oakland. These numbers of points were interpolated to 69,181 records, one for each cell in the reference file.

Similarly, kernel estimations were produced for selected census data variables on a block group level. Statistical analysis of variables within the Oakland tuberculosis data set have revealed the following demographic factors to be associated with much of the population infected with tuberculosis: 1) age of 65 years old and over, 2) foreign born, 3) male, and 4) non-white race. In addition, prior research on tuberculosis infection from literature reviews indicates that: 5) poverty; and 6) high number of persons per room (overcrowding) are also important indicators of high tuberculosis infection. Therefore, data on a block group level pertinent to the above six factors were extracted from the 2000 Census and mapped using the kernel density routine, using the relevant variable as the intensity variable and the block group centroids as the location points. In addition, kernel estimates were produced for the total population of the block groups, so that the tuberculosis infection rate could be calculated. This was accomplished by dividing the kernel density estimation of tuberculosis infection by the kernel density estimation of the block group population and multiplying by 100,000 to obtain a rate of infection per 100,000 persons. All interpolations were mapped using ArcMap 8.2, a standard desktop geographic information system (GIS) produced by ESRI, Inc.

Regression Analysis

To test the hypothesis that tuberculosis infection incidence is correlated with the above six factors, the Council constructed a regression model with tuberculosis infection rate as the dependent variable and the above-mentioned six demographic characteristics (or transformations of them) as the independent variables. All data used in the regressions were derived from the kernel density estimations produced by CrimeStat II. Simple regression correlations as well as a multivariable one were conducted using SPSS. With the data produced by the kernel density estimations, SPSS was also used to derive an equation that attempts to predict the likelihood of tuberculosis infection based on the data for the above six demographic factors. This "likelihood of infection" was calculated for the entire city of Oakland and mapped using ArcMap, producing a “risk” map showing where prevention efforts should be targeted based on demographic factors.

III. LITERATURE REVIEW

Tuberculosis is an infectious disease that usually attacks the lungs, but can attack almost any part of the body. Tuberculosis is a contagious disease transmitted through the air when a person with active tuberculosis coughs, laughs, sneezes, sings, or even talks. Anyone inhaling air containing the tuberculosis bacteria may become infected. Repeated contact is usually required for infection. The immune system usually prevents the development of the active disease. However, if untreated, about 5 percent of those who are infected will develop active tuberculosis disease at some point in their lives. The risk is much higher for persons who have suppressed immune systems, including those with HIV infection. If active tuberculosis disease develops, it usually takes six or more months of drug therapy to cure it.

It is important to understand that there is a difference between being infected with tuberculosis and having active tuberculosis disease. Someone who is infected with tuberculosis has the tuberculosis bacteria, in their body. The body's defenses are protecting them from the germs and they are not sick. Someone with active tuberculosis disease is sick and can spread the disease to other people.

Historical Sketch

Until about 50 years ago, tuberculosis was considered nearly incurable. The discovery of several active anti-tuberculosis agents in 1940s brought in a new cure for the disease. Unfortunately, in only a few years, it became apparent that the use of these miracle drugs as single agents led to rapid drug resistance and treatment failures among a substantial number of patients. It was quickly realized, however, that the development of resistance could be prevented through treatment with several active agents at once in a combination regimen. From this point tuberculosis was considered a curable disease. In fact, until only recently, public health officials in the United States predicted that tuberculosis would be totally eradicated in this country by the year 2000.

From 1953 through 1984, the number of cases of tuberculosis infection declined. In 1985 however, in conjunction with a rise in the number of AIDS cases, we witnessed a striking reversal of that trend. Areas with the highest rates of HIV infected people (e.g. New York, New Jersey, Florida and California) showed the largest number of new tuberculosis cases. Some of the factors that have led to the tuberculosis resurgence are:

- AIDS epidemic
- Immigration from countries with tuberculosis epidemics
- Increased poverty, injection drug use, homelessness
- Increased number of residents in long-term care facilities

California's Challenges

The number of tuberculosis cases reported in California rose slightly in 2001 with 3,332 new cases reported, up from 3,297 in 2000. But the tuberculosis rate -- the number of cases per 100,000 residents -- remained the same at 9.5 cases for every 100,000 residents both years. The number of tuberculosis cases has declined since 1992 when 5,382 cases were reported.

Despite the stable rate of new tuberculosis cases recently, California still has the highest number of cases in the United States, and there are wide differences in tuberculosis rates among the state's ethnic groups. The rate of cases per 100,000 residents is 33.8 for Asians and Pacific Islanders; 12.4 for African Americans; 11.4 for Hispanics; 7.7 for Native Americans; and 2.1 for white non-Hispanic residents.

While the decline in tuberculosis cases since 1992 has been significant, major disparities in the risk of tuberculosis persist among racial and ethnic groups. For example, compared to Whites, case rates were 4.7 times higher for Latinos; 5.5 times higher for African-Americans; and 14 times higher for Asians in 1998.

California still faces a formidable challenge in fighting tuberculosis. More than two-thirds of California's tuberculosis cases are foreign-born. Birth in a country with high tuberculosis rates is an important reason for high case rates among Asians and Latinos in California: 95 percent of Asian cases and 75 percent of Latino cases reported in the state in 1998 were foreign-born.

California's U.S.-born cases (while only one-third of the total) are relatively difficult to treat. Among U.S.-born cases, for example, California has a disproportionately high number of persons who are homeless, move through the correctional system, or use drugs. It is relatively difficult to track and case-manage patients in these groups.

Currently, the California Department of Health Service indicates that it is focusing on better identification of persons infected with tuberculosis and on increasing the percentage of patients who complete drug therapy. Specifically, the department plans to: (1) increase targeted tuberculosis skin testing of high-risk groups including persons in contact with infectious tuberculosis, immigrants, correctional inmates, and those with HIV infection; (2) develop activities to reduce the incidence of tuberculosis among racial and ethnic groups; and (3) expand data collection and analysis in order to better evaluate the programs' impacts and cost-effectiveness.

Similarly, in Alameda County, despite increased tuberculosis reporting and surveillance, the number of tuberculosis cases rose from 224 to 242 during between 1999 and 2000. In Alameda County, 71% of persons with tuberculosis infection are foreign-born. Increasingly the elderly are being impacted: 30% of those with tuberculosis infections are 65 and older. Roughly 41% of those in the county with tuberculosis are Oakland residents. It is estimated that 10% -- or 144,000 – of Alameda County residents have tuberculosis infection. Persons infected with tuberculosis need medical evaluation and treatment to prevent the infection from progressing to tuberculosis.

Related Literature

Research shows that anyone can contract tuberculosis-people of all races and nationalities, the rich and poor, old and young. But for many reasons, some groups of people are at higher risk to get active tuberculosis disease. According to the American Lung Association, the groups that are at high risk include: people with HIV infection (the AIDS virus); people in close contact with those known to be infected with tuberculosis; people with medical conditions that make the body less able to protect itself from disease (for example: diabetes, the dust disease silicosis, or people undergoing treatment with drugs that can suppress the immune system, such as long-term use of corticosteroids); foreign-born people from countries with high tuberculosis rates; some racial or ethnic minorities; people who work in or are residents of long-term care facilities (nursing homes, prisons, some hospitals); those working with populations with tuberculosis, including health care workers and others such as prison guards; people who are mal-nourished; alcoholics and IV drug users.

In a study by Talbot, Moore, McCray and Binkin, *"**Tuberculosis Among Foreign-born Persons in the United States, 1993-1998**,"* the authors concluded that immigration is a major force sustaining the incidence of tuberculosis in the United States. In an examination of U.S. tuberculosis surveillance data from case reports submitted from 1993 to 1998, the authors documented that the rates of tuberculosis cases among foreign-born persons were higher than those for the native born. During the period 1993-1998, the tuberculosis case rate was 32.9 per 100,000 foreign-born persons compared with 5.8 per 100,000 for U.S.-born persons. Six states reported 73.4% of foreign-born cases (California, New York, Texas, Florida, New Jersey and Illinois). Approximately two-thirds of these cases were originally from Mexico, the Philippines, Vietnam, India, China, Haiti and South Korea. Among those for whom the date of U.S. entry was known, 51.5% arrived 5 years or less prior to the diagnosis of tuberculosis. Most were male and between 25 to 44 years of age. The authors concluded that as the United States moves toward the goal of tuberculosis elimination, success will depend increasingly on reducing the impact of tuberculosis in foreign-born persons. Continued efforts to tailor local tuberculosis control strategies to the foreign-born community and commitment to the global tuberculosis battle are essential.

In identifying risk factors for tuberculosis, Barr et al. in their paper "***Neighborhood Poverty and the Resurgence of Tuberculosis in New York City 1984-1992"*** cite a group of factors which propelled the resurgence tuberculosis in New York. These factors included AIDS, immigration, injection drug use, multi-drug resistance and homelessness, as well as a breakdown in public health measures. The study asserts that poverty has traditionally been a risk factor for tuberculosis. However, in an area with high rates of AIDS and immigration, the contribution of poverty to tuberculosis risk may be overlooked. The authors investigated whether poverty remained a major risk factor for tuberculosis at the peak of the NYC tuberculosis epidemic.

The results of the study indicated that the incidence of tuberculosis in 1992 in the 5,482 neighborhoods examined was 46.5 per 100,000 persons and the mean rate of poverty across these neighborhoods was 19.3 percent. In general, the authors found that from

1980 to 1990 the rate of new tuberculosis cases grew as the average neighborhood income dropped. The authors concluded that neighborhood income was the factor most strongly associated with tuberculosis risk and poverty was more closely linked to tuberculosis risk than HIV status, race or ethnicity.

Research by Troels Lillebaek et al. in Denmark shows that immigrants from countries with high incidence of tuberculosis are thought to have fueled the resurgence of tuberculosis in areas of low incidence such as Denmark. According to Lillebaek, in most industrialized countries, the annual numbers of cases and deaths caused by tuberculosis have steadily declined over the past century up to the mid-1980s. Since then, however, an increasing number of tuberculosis cases in immigrants have reversed this downward trend in countries that have had substantial levels of immigration from areas with high prevalence of the disease. The study further found that the annual incidence of tuberculosis declined only gradually during the first 7 years of residence, from an initial 2,000 per 100,000 to 700 per 100,000. This finding seriously challenges the adequacy of the customary practice of screening solely upon arrival.

In a similar study by Cowie and Sharpe in Canada, the authors observed that about 71% of the 351 cases of tuberculosis in Alberta during a five-year period were immigrants from Asia. The authors concluded that given the low annual incidence of tuberculosis in Canada (7.1 per 100 000), it is probable that tuberculosis occurring among immigrants reflects infection acquired before arrival in Canada. Supporting the findings in Lillebaek et al. regarding the length of stay and reduction in tuberculosis cases, Cowie and Sharpe found that the mean period between immigration and diagnosis was 11.2 years.

The Center for Diseases Control (CDC) reports that tuberculosis disproportionately affects minorities. In 2000**,** 77% of all reported tuberculosis cases occurred among racial and ethnic minorities. In 2000, rates of tuberculosis were dramatically higher for Asians/Pacific Islanders (32.9 per 100,000), African Americans (15.2 per 100,000), American Indians/Alaskan Natives (11.4 per 100,000), Hispanics (10.8 per 100,000), than for whites (1.9 per 100,000). The CDC concluded that several factors likely contribute to the disproportionate burden on minorities. Unequal distribution of tuberculosis risk factors, such as HIV infection, as well as the effects of lower socioeconomic status, particularly overcrowding, contribute to increased risk of tuberculosis.

According to CDC, tuberculosis also has had a disproportionate impact on men. In 2000, 62% (10,225) of tuberculosis cases occurred among men, and 38% (6,148) occurred among women. The rate of tuberculosis for men (7.4 per 100,000) was almost double that of women (4.3 per 100,000).

The CDC also reports that the greatest proportion of tuberculosis cases occur among people in the prime of their life, persons aged 25 to 44.

IV. SOURCES OF DATA

Two primary sources of data were utilized in this study including information on tuberculosis cases provided by the Alameda County Public Health Department and the 2000 U.S. Census data.

Tuberculosis Case Data

The tuberculosis case data are from the Alameda County Public Health Department's Tuberculosis Control Program. They are comprised of all reported cases of tuberculosis in Alameda County, except Berkeley, which maintains separate records through a City Department of Health. Health providers report to their local health jurisdiction all cases of tuberculosis. Cases are counted by city and county of residence, regardless of where the case was diagnosed. For the purpose of this research only reported cases of tuberculosis with Oakland indicated as place of residence are used for the analysis.

2000 Census Data

For the various analyses utilized in the study, we extracted data from the 2000 Census for the City of Oakland. The following data for the 337 census block groups of Oakland were extracted from Summary File 3 of the 2000 Census conducted by the U.S. Bureau of the Census. Kernel density estimations were conducted on the data and the results were used in the regressions and the production of maps.

CENSUS CATEGORY TITLE	Field Numbers
TOTAL POPULATION	P001001
RACE	P006001 - P006008
SEX BY AGE	P008001 - P008079
PLACE OF BIRTH BY CITIZENSHIP STATUS	P021001 - P021015
POVERTY STATUS IN 1999 BY AGE	P087001 - P087017
TENURE BY OCCUPANTS PER ROOM	H020001 - H020017

RACE provides data on the non-white population. SEX BY AGE provides data for the male population as well as the population 65 years and older. PLACE OF BIRTH BY CITIZENSHIP STATUS provides data on the foreign-born population. POVERTY STATUS IN 1999 BY AGE provides poverty data, while TENURE BY OCCUPANTS PER ROOM provides data on persons per room. The data on TOTAL POPULATION was used to normalize much of the data to percentages and was essential in calculations of rates of infection.

V. COMPARISON OF DEMOGRAPHIC CHARACTERISTICS OF OAKLAND POPULATION AND TUBERCULOSIS CASE POPULATION

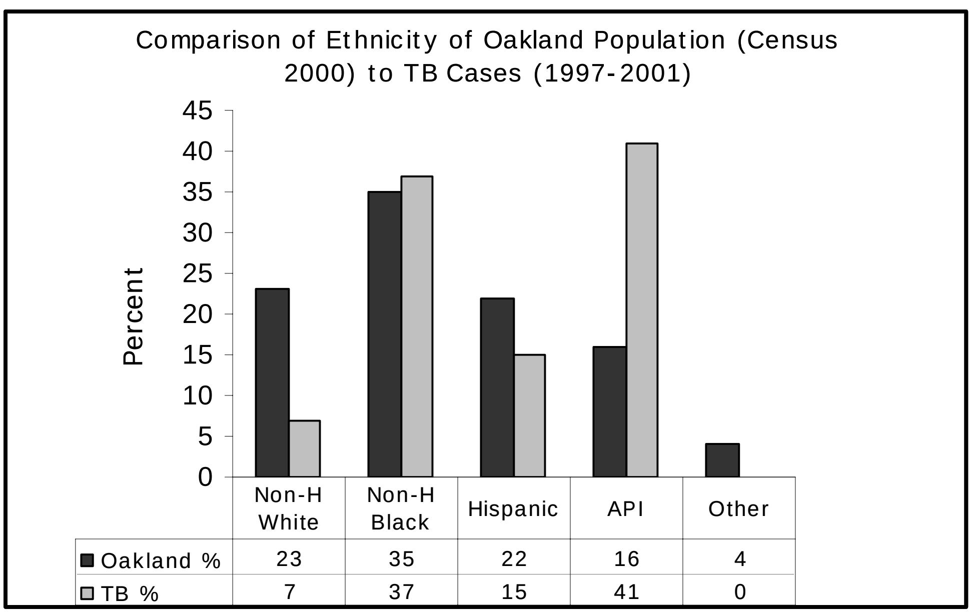
This section describes several basic demographic characteristics of the Oakland population and compares the general population characteristics to those of the tuberculosis case population. The variables reviewed here include race-ethnicity, gender, age and U.S. citizenship status.

The 2000 Census reports that 399,477 people live in the city of Oakland. The race-ethnic break down for Oakland's population is 41% Asian/Pacific Islander (API), 37% Non-Hispanic Black, 15% Hispanic, 7% Non-Hispanic White, and 7% American Indian and Alaskan Native. There are more females (52%) than males (48%) residing in Oakland.

Between 1997 and 2001, there were 456 reported cases of tuberculosis. In general, the tuberculosis case population differs substantially from the general population in that a higher percentage of tuberculosis cases are male, Asian/Pacific Islander, in the 65+ age category and immigrants than in the general population. Chart 1 is a summary of the ethnicity of the two populations.

The race and ethnicity of Oakland's population differed from the tuberculosis case population in several respects. While Asian-Pacific Islanders were 16% of the city's population and 41% of the tuberculosis cases. Non-Hispanic Whites were 23% of the city's population and 7% of the tuberculosis cases; Non-Hispanic Blacks were 35% of Oakland's population but 37% of the tuberculosis cases; Hispanics were 22% of Oakland's population and 15% of tuberculosis cases..

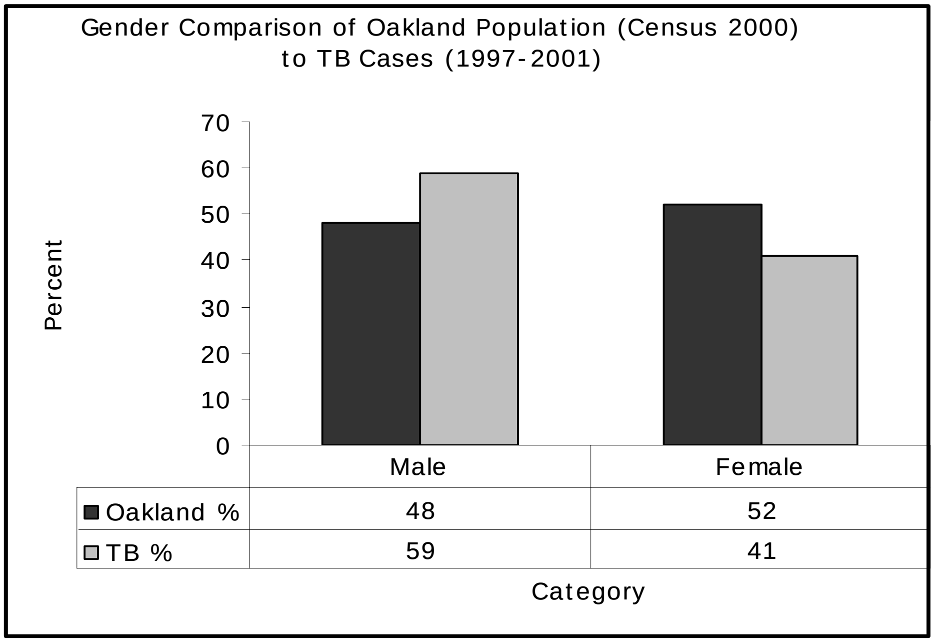
Chart 1 *Oakland's race and ethnicity compared to tuberculosis cases*



S*ource: U.S Census Bureau, Census 2000 and Alameda County Public Health Dept. 1997-2001*

Chart 2 presents the gender breakdown of the Oakland and tuberculosis populations. Oakland's population consists of 48% males and 52% females while the tuberculosis case population consists of 59% males and 41% females, meaning males were overrepresented in the tuberculosis cases by 11%.

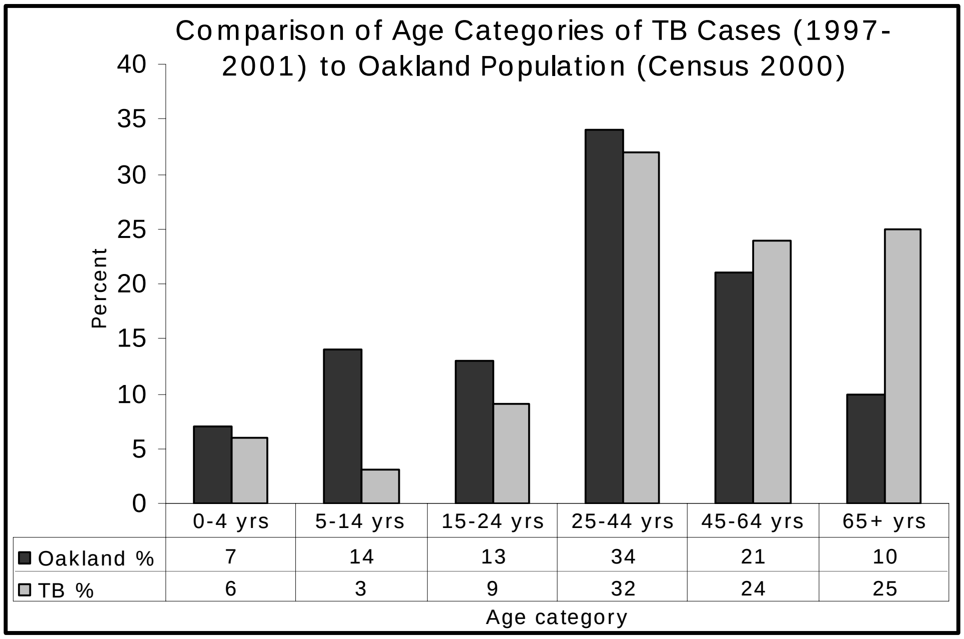
Chart 2: Gender of Oakland population compared to tuberculosis population



S*ource: U.S Census Bureau, Census 2000 and Alameda County Public Health Dept. 1997-2001*

Chart 3 below presents the age comparison for the two populations. While those 0-4 years of age were 7% in Oakland's population they were 6% in the tuberculosis cases; 5-14 year olds were 14% in Oakland's population but 3% in the tuberculosis cases; 15-24 year olds were 13% of Oakland's population and 9% of tuberculosis cases; 25-44 years old were 34% of Oakland's population and 32% of tuberculosis cases; 45-64 years were 21% of Oakland's population and 24% of tuberculosis cases; and 65+ were 10% of Oakland's population, and 25% of tuberculosis cases. Those in the O-44 age groups were all underrepresented in tuberculosis cases and while the two age categories 45 years and older were overrepresented with the 65+ category was overrepresented by 15% among tuberculosis cases.

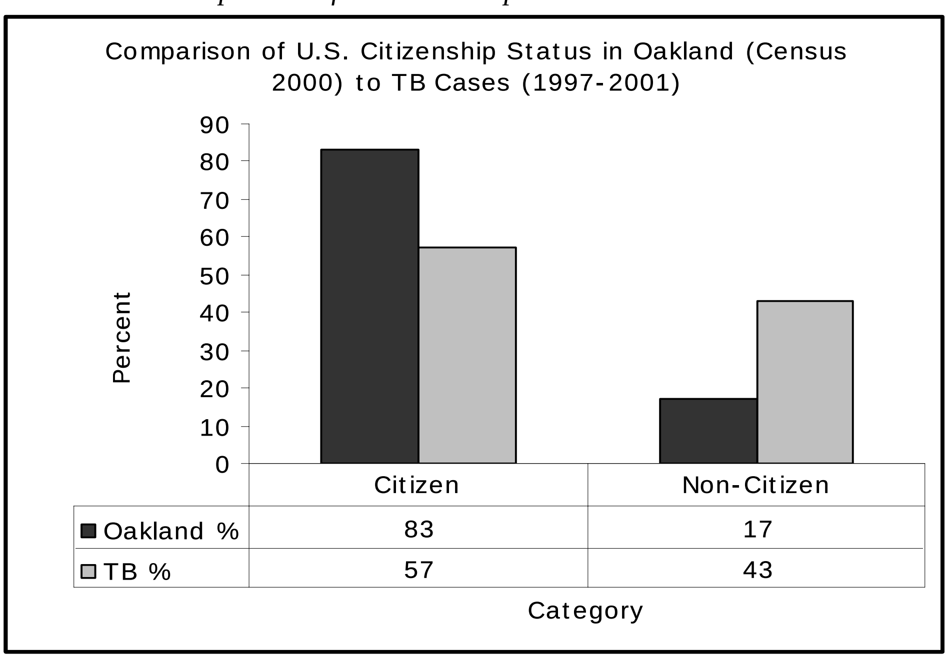
Chart 3: Age categories of Oakland compared to tuberculosis case



S*ource: U.S Census Bureau, Census 2000 and Alameda County Public Health Dept. 1997-2001*

The 2000 Census reported that the citizenship status of Oakland's population was 83% U.S. citizens and 17% non-citizens. For the tuberculosis cases, 43% were U.S. citizens, compared to 57% non-citizens. This shows the substantial over- representation of non - citizens in the tuberculosis cases (See chart 4 below).

Chart 4: Citizenship status of Oakland compared to tuberculosis cases



S*ource: U.S Census Bureau, Census 2000 and Alameda County Public Health Dept. 1997-2001*

Summary of Comparisons between Oakland's Population and tuberculosis cases

- While males comprise 48% of the population, they account for 59% of the tuberculosis cases.
- Asian Pacific Islanders are substantially overrepresented among reported tuberculosis cases, accounting for 41% of the cases but only 16% of the population. African Americans account for approximately 35% of Oakland population, and 37% of reported tuberculosis cases. Hispanics, on the other hand, are slightly underrepresented comprising 22% of the population but only 15% of the reported tuberculosis cases. Whites are substantially underrepresented accounting for 23% of the population but only 7% of reported cases.
- Those 65 years of age and older account for a disproportionately high number of the tuberculosis cases. While they comprise 10% of the population, they account for 25% of reported cases
- While non-citizens account for 17% of the population of the city, they comprise 57% of the reported cases of tuberculosis.

VI. CORRELATION AND REGRESSION ANALYSES

In this section, we discuss the correlation and regression analyses of the dependent and independent variables (predictors) used in this study. The purpose of these analyses is to determine whether the incidence rates of tuberculosis infections are positively correlated with:

- the percent non-white and percent foreign-born population;
- the percent of persons in poverty; and
- the number of occupants per room in housing units.

Also in this section, we present the results of the regression analysis model we developed to explain tuberculosis infections based on the relationship with the predictor variables.

Variables used in the analysis

The following predictor variables are used in the analysis:

- percent 65+ (PER_65),
- occupants per room (C_CROWD),
- percent male (PER_MALE),
- percent non-white (PER_NONW),
- percent foreign born (PER_IMM) and
- percent in poverty (PER_POV)

The dependent variable is the incidence of tuberculosis multiplied by 100,000 (TB100K) to give us the rate per 100,000.

Regression analysis operates under the assumption that the predictor variables are normally distributed. In order to conform to that assumption, the variables PER_POV, PER_NONW and PER_IMM were transformed from nonlinear to linear variables using the logit model and C_CROWD and PER_65 were squared. To determine whether the variables were linear or not, simple scatterplots were used. The only variable that was not transformed for this analysis was percent male, because the original data was linear.

It is important to note that when examining more than one predictor variable in any analysis, the influence of a specific predictor cannot be solely determined by the correlation between the two variables. It is likely that the predictor variables influence each other and, in so doing, influence the relationship between the predictor and the dependent variable. This is because when two predictor variables correlate strongly with each other, their ability to predict unique portions of the dependent variable decreases. To account for this, we examine partial correlations under our regression analysis to examine the effectives of the predictors on each other. This will allow us to determine the actual variations between the predictors and the dependent variable.

Software: Statistical Package for the Social Sciences (SPSS)

SPSS is the statistical package commonly used by social scientists. Many of the widely used social science data sets come with an easy method to translate them into SPSS; this significantly reduces the preliminary work needed to explore new data. As a result, SPSS was used for this analysis. It has features which allow users to perform tasks such as frequencies, cross tabulation, correlation and regression analysis. SPSS also allows users to transform data into different variables in order to perform appropriate significance tests and other analyses. For the current study, we analyzed to what extent the variables are related (correlation) and to what extent the dependent variable can be explained by the predictor variables (regression).

Correlation Analyses

Correlation is a measure of the strength of the association between variables. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship). A correlation coefficient of -1.0 indicates a perfect negative correlation, while +1.0 describes a perfect positive relationship. The closer the correlation coefficient is to either -1.0 or +1.0, the stronger the relationship. To observe how the predictor variables associate with the dependent variable, a correlation analysis was performed. This analysis also helps one observe the direction of the relationship between the predictor variables and the dependent variable. The analysis helped draw conclusions about our hypotheses, i.e., how the incidence rates of tuberculosis infections are positively correlated with the percent non-white and percent foreign-born population, and how the incidence rates of tuberculosis infection are positively correlated with the percent of poverty. This analysis is presented in Table 1 below with a summary of the table following it.

Table 1: Correlation of the predictor variables and the dependent variable.

Correlations

		TB100K	GOOD-Logit of per_pov	GOOD-Logit of per_imm	GOOD-Square of c_crowd	GOOD-Square of per_65	GOOD-Logit of per_nonw	GOOD-per_male
TB100K	Pearson Correlation	1	.542**	.644**	.536**	-.120**	.449**	.164**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000
	Sum of Squares and Cross-products	5.9E+10	37786894.289	30300704.406	9276489.331	-77255.128	43392508.482	270896.6
	Covariance	996343.9	636.936	510.749	156.365	-1.302	731.425	4.566
	N	59327	59327	59327	59327	59327	59327	59327
GOOD-Logit of per_pov	Pearson Correlation	.542**	1	.601**	.703**	-.426**	.774**	.087**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000
	Sum of Squares and Cross-products	3.8E+07	82144.069	33316.023	14353.510	-323.440	88188.527	169.614
	Covariance	636.936	1.385	.562	.242	-.005	1.487	.003
	N	59327	59327	59327	59327	59327	59327	59327
GOOD-Logit of per_imm	Pearson Correlation	.644**	.601**	1	.749**	-.304**	.489**	.332**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000
	Sum of Squares and Cross-products	3.0E+07	33316.023	37416.823	10319.893	-155.439	37551.823	436.287
	Covariance	510.749	.562	.631	.174	-.003	.633	.007
	N	59327	59327	59327	59327	59327	59327	59327
GOOD-Square of c_crowd	Pearson Correlation	.536**	.703**	.749**	1	-.448**	.733**	.479**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000
	Sum of Squares and Cross-products	9276489	14353.510	10319.893	5070.504	-84.419	20742.348	231.971
	Covariance	156.365	.242	.174	.085	-.001	.350	.004
	N	59327	59327	59327	59327	59327	59327	59327
GOOD-Square of per_65	Pearson Correlation	-.120**	-.426**	-.304**	-.448**	1	-.393**	.047**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000
	Sum of Squares and Cross-products	-77255.1	-323.440	-155.439	-84.419	7.010	-413.538	.851
	Covariance	-1.302	-.005	-.003	-.001	.000	-.007	.000
	N	59327	59327	59327	59327	59327	59327	59327
GOOD-Logit of per_nonw	Pearson Correlation	.449**	.774**	.489**	.733**	-.393**	1	.192**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000
	Sum of Squares and Cross-products	4.3E+07	88188.527	37551.823	20742.348	-413.538	157929.569	519.707
	Covariance	731.425	1.487	.633	.350	-.007	2.662	.009
	N	59327	59327	59327	59327	59327	59327	59327
GOOD-per_male	Pearson Correlation	.164**	.087**	.332**	.479**	.047**	.192**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.
	Sum of Squares and Cross-products	270896.6	169.614	436.287	231.971	.851	519.707	46.291
	Covariance	4.566	.003	.007	.004	.000	.009	.001
	N	59327	59327	59327	59327	59327	59327	59327

**. Correlation is significant at the 0.01 level (2-tailed).

- All of the correlations between the predictors and tuberculosis are positive except age.
- The strongest of these is percent foreign born and tuberculosis, reported as .644 (closer to +1.0 than the other variables). This correlation indicates that there is a positive relationship between tuberculosis cases and non-citizenship status, meaning the incidence of tuberculosis rises as the percent of persons with non-citizenship status rises.
- The second strongest correlation is with percent poverty at .542. This correlation indicates that blocks which have high poverty concentration tend to have high reported cases of tuberculosis.
- The correlation between tuberculosis and percent crowding was also high at .536. This correlation indicates incidence of tuberculosis cases tend to be associated with higher rates of over-crowded housing conditions.
- The correlation between percent non-white and tuberculosis was also high at .449. This show that incidence of tuberculosis cases is associated with blocks with a higher percentage of non-white residents.
- The correlation between tuberculosis and percent male is .164. This correlation shows that tuberculosis infection tend to be associated with male, although the association is not as strong as for the above variables.

- Finally, being 65+ negatively correlated with tuberculosis, meaning the 65+ and over population are less likely to be among the reported tuberculosis cases. This was unexpected, given that 65+ was highly represented in the tuberculosis cases used for this analysis. The possible explanation of this negative relation is the influence of other variables on the relationship.

Regression Analyses

Based on our hypotheses, we constructed a regression model to explain and make predictions about tuberculosis infections, which included the six predictor variables described above as independent variables and the rate of tuberculosis infections as the dependent variable. The results of this model's regression equation were used to develop maps of high tuberculosis areas. (See Maps 11 and 12).

See appendix I for the derivation of the regression equation.

Model Summary

The model summary indicates that all six predictors were used in the analysis and that the rate of tuberculosis infection was the dependent variable. The overall relationship between the six predictors and tuberculosis infection is reported as .694. When this multiple correlation (R) is squared, we find that 48.1% of the variance in the tuberculosis infection can be explained using these variables. An R squared of this magnitude indicates that this is a powerful predictive model.

Table 2: Summary of the regression model

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.694[a]	.481	.481	719.0810	.481	9165.634	6	59320	.000	1.625

a. Predictors: (Constant), GOOD-Square of per_65, GOOD-per_male, GOOD-Logit of per_pov, GOOD-Logit of per_imm, GO per_nonw, GOOD-Square of c_crowd

b. Dependent Variable: TB100K

Regression Coefficient

The regression coefficient provides information about the relationship between the dependent variable and a specific independent variable when the effects of the other independent variables are held constant (Table 3). The standardized coefficient allows us to compare the independent effects of each predictor variable on the dependent variable. Notice that the variable representing percent foreign born had the biggest impact on the rate of tuberculosis infection.

Table 3: Regression coefficient of the six predictors and the dependent variable.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval - B		Correlations		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part
1	(Constant)	3127.742	60.173		51.979	.000	3009.802	3245.683			
	Per_male	2966.191	137.673	-.083	-21.545	.000	-3236.030	-2696.353	.164	-.088	-.064
	Logit of per_pov	189.351	4.572	.223	41.414	.000	180.390	198.313	.542	.168	.122
	Logit of per_nonw	37.795	3.271	.062	11.555	.000	31.384	44.206	.449	.047	.034
	Logit of per_imm	635.067	5.912	.505	107.425	.000	623.480	646.654	.644	.404	.318
	Square of c_crowd	277.149	23.864	.081	11.614	.000	230.375	323.923	.536	.048	.034
	Square of per_65	7725.612	323.760	.193	54.749	.000	17091.041	18360.184	-.120	.219	.162

a. Dependent Variable: TB100K

Findings and observations:

The findings from the analyses confirm our three hypotheses:

1. tuberculosis infection is positively associated with percent non-white and percent foreign born.
2. tuberculosis infection is positively associated with percent in poverty.
3. tuberculosis infection is positively associated with crowding when all the other variables are controlled.

Overall observations of the analyses:

- 48.1% of the variance in tuberculosis can be explained by the six variables, indicating that this is a strong model.
- Immigration is the single most important variable that predicts tuberculosis. This is followed by poverty.
- The partial correlation for percent poverty, occupants per room and percent non-white drops off substantially when all the other predictors are held constant.

VII. SPATIAL ANALYSES RESULTS

This section presents maps produced by our spatial analyses. Chloropleth maps are maps where the boundaries of the zones are established independently from the data, and so data from different sources can be aggregated to these units to generate maps. The maps presented here allow a comparison between the results of a traditional chloropleth analysis and the kernel density estimation method.

Kernel density estimation is based on the actual locations of cases and the distances between them. Maps produced by this technique allow for a more detailed identification of tuberculosis "hot spots" than the chloropleth maps. The chloropleth analysis only allows for the display of the incidence of tuberculosis cases within particular political or administrative units. Therefore, the kernel density maps may prove more useful for community prevention and intervention strategies.

First we present traditional chloropleth maps, which display the incidence of tuberculosis cases by several political and administrative units. The units are shaded in relationship to the number of cases occurring within them, with darker shading corresponding to a higher number of cases.

The maps following these chloropleth maps are contour maps which display geographically the data produced by the kernel density estimations interpolated for the entire city. First are the tuberculosis infection cases, followed by the total population of the city by block groups, which lead to the maps of tuberculosis infection rates per 100,000 persons. Finally, we display maps of tuberculosis risk where the data were calculated by using an equation derived from the multivariate regression of the six demographic variables associated with the population infected. An appendix contains maps of these six demographic variables, presented for the purpose of visual comparison. The maps were all produced by ArcMap, which automatically generated the ranges of data displayed using a method that searched for natural breaks in the data and broke the data up accordingly. The section concludes with a discussion on the efficacy of the new methods of spatial analysis.

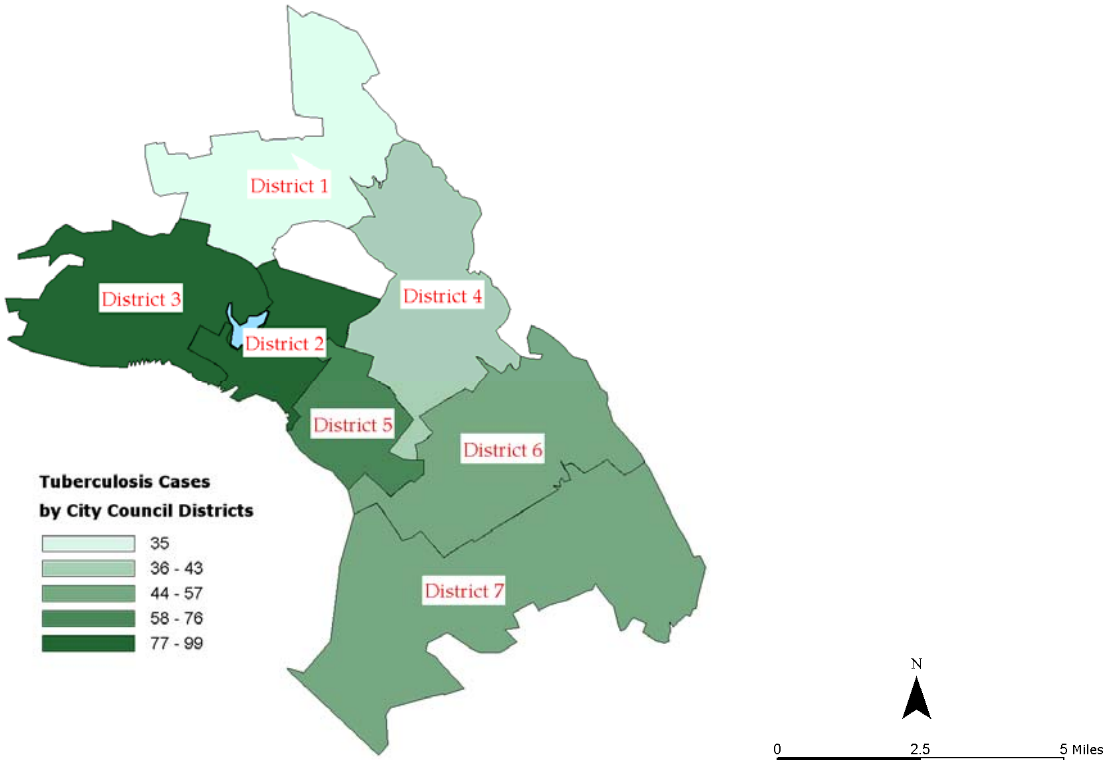
Chloropleth Maps

Maps 1 to 3 are chloropleth maps of the number of tuberculosis infection cases in Oakland by City Council Districts, PHD Health Team Areas and ZIP codes.

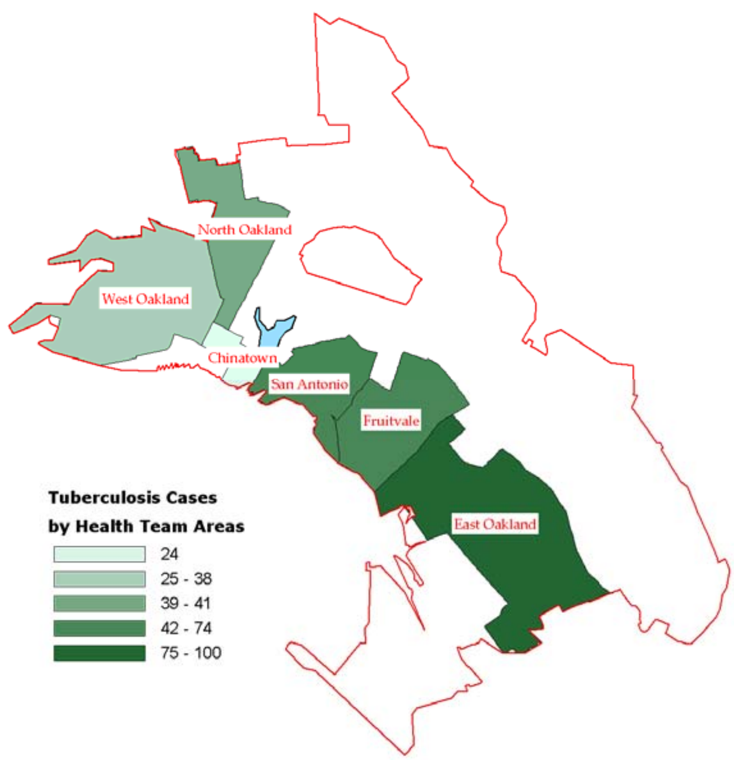
Findings

The size of the geographic units and the inconsistency of the boundaries between them lead to inconsistent results. Each of the first three maps suggests that different areas of the city are experiencing higher incidences of tuberculosis cases, and the geographies are too large to provide useful information to guide intervention and prevention efforts. Analyzing the data at the census tract level (Map 4) provides more geographically specific results but still doesn't provide enough information about the location and context of the tuberculosis cases.

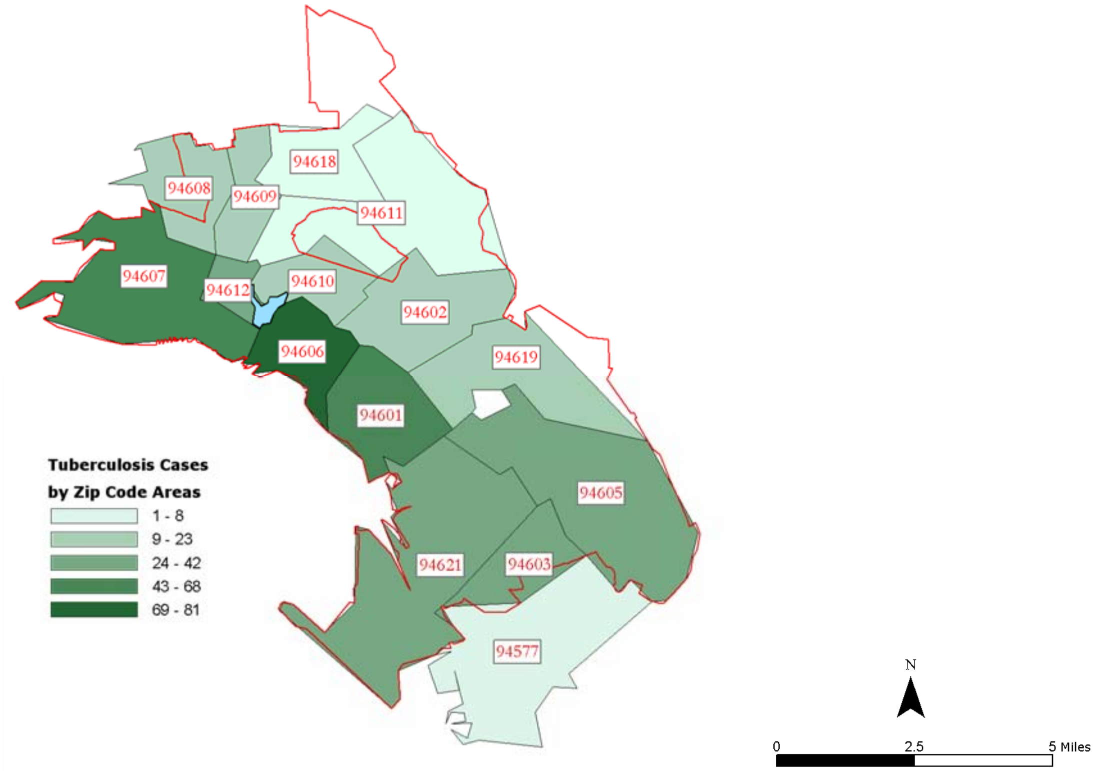
Map 1: Tuberculosis Cases by City Council Districts



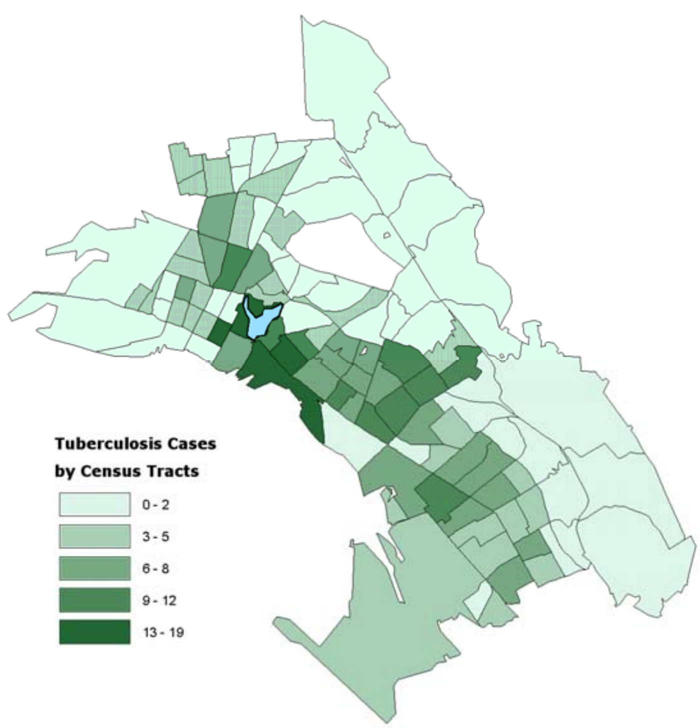
Map 2: Tuberculosis Cases by County Community Health Team Areas



Map 3: Tuberculosis Cases by Zip Code Areas



Map 4: Tuberculosis Cases by Census Tracts



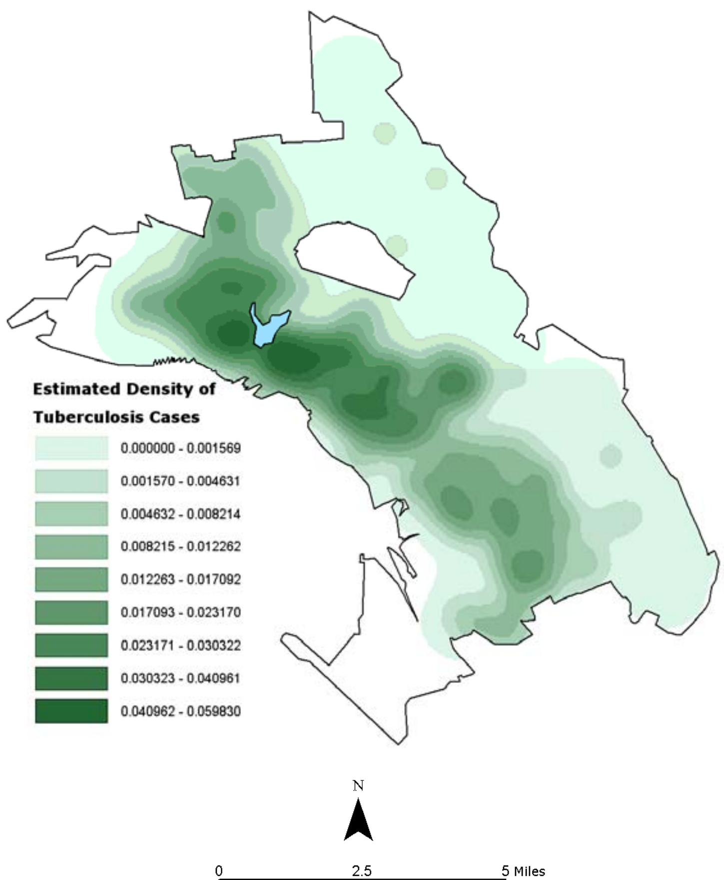
Contour Maps

A contour map is a type of isopleth map, which shows an imaginary surface by means of lines joining points of equal value so that continuous data can be represented. Map 5 is a contour map of a single kernel density of tuberculosis infection incidences in Oakland. Map 6 is the same as Map 5 but with the actual locations of tuberculosis infection cases, represented by dots, overlaid over the contours.

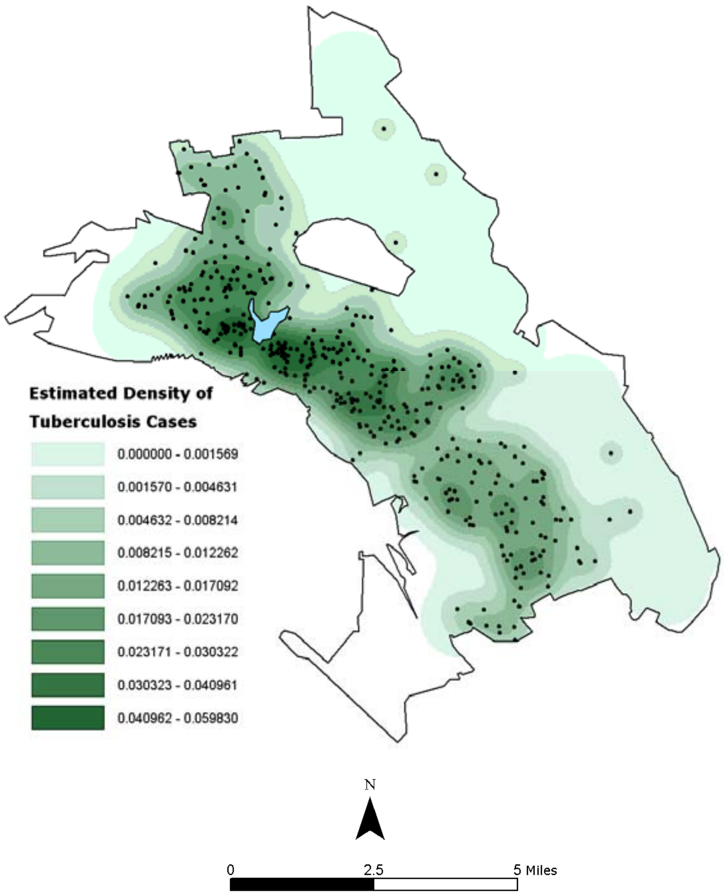
Findings

These maps show Downtown and the Lower San Antonio area immediately south of Lake Merritt, among other areas in Oakland, to be "hot spots" of tuberculosis infection. A layer of streets or neighborhoods can be overlaid on the contour map to obtain more exact locations. A comparison of this map to the four chloropleth maps demonstrates the advantage of this technique. Here the more darkly shaded sections of the map have a direct correspondence to the locations of actual tuberculosis cases. Prevention and intervention efforts guided by this analysis would lead one to concentrate on specific neighborhoods and sub-areas where tuberculosis is most prevalent, rather than on pre-defined geographies such as census tracts.

Map 5: Estimated Density of Tuberculosis Cases

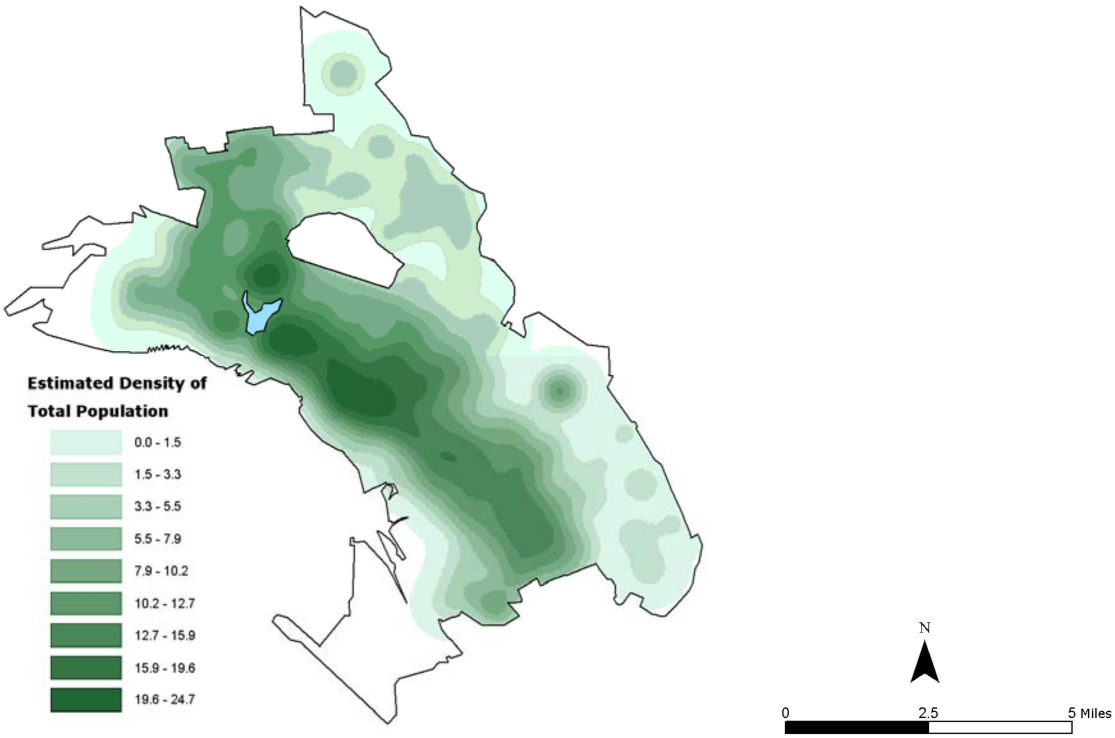


Map 6: Estimated Density of Tuberculosis Cases Overlaid with Locations of Infection

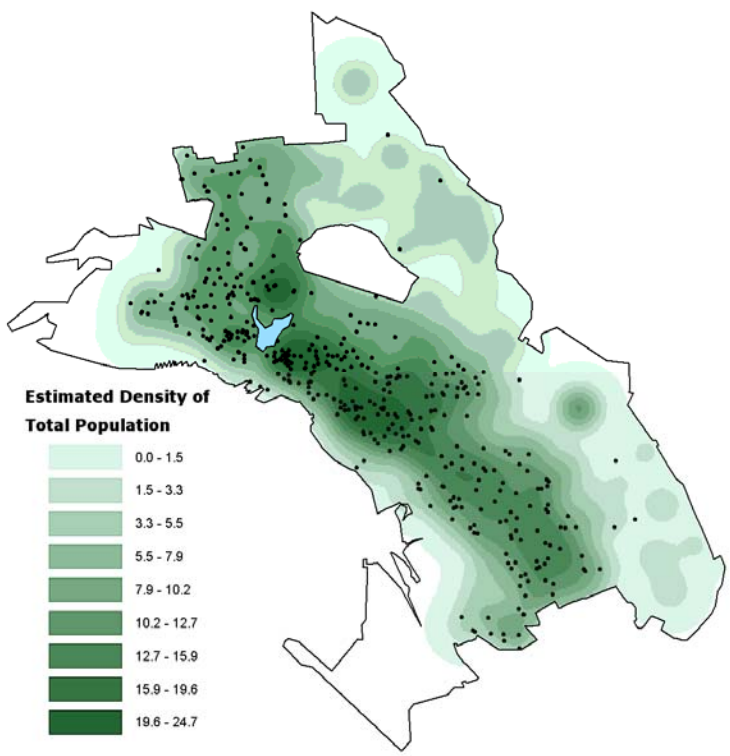


Map 7 uses kernel density estimation to display the population density of Oakland. Census block groups are analyzed with the number of persons as the intensity variable. Map 8 is the same, with the locations of tuberculosis cases represented by dots overlaid on top of the population contours. When estimating the density of disease infections within a city by interpolating the disease incidents, the result can be a high concentration in the areas with a high concentration of people. By comparing Map 5, the density of tuberculosis infections, with Map 7, the density of total population, this would appear to be the case. The question emerges of how likely is a disease infection incident when taking into account the number of people living in these areas.

Map 7: Estimated Density of Total Population



Map 8: Estimated Density of Total Population Overlaid with Locations of Infection

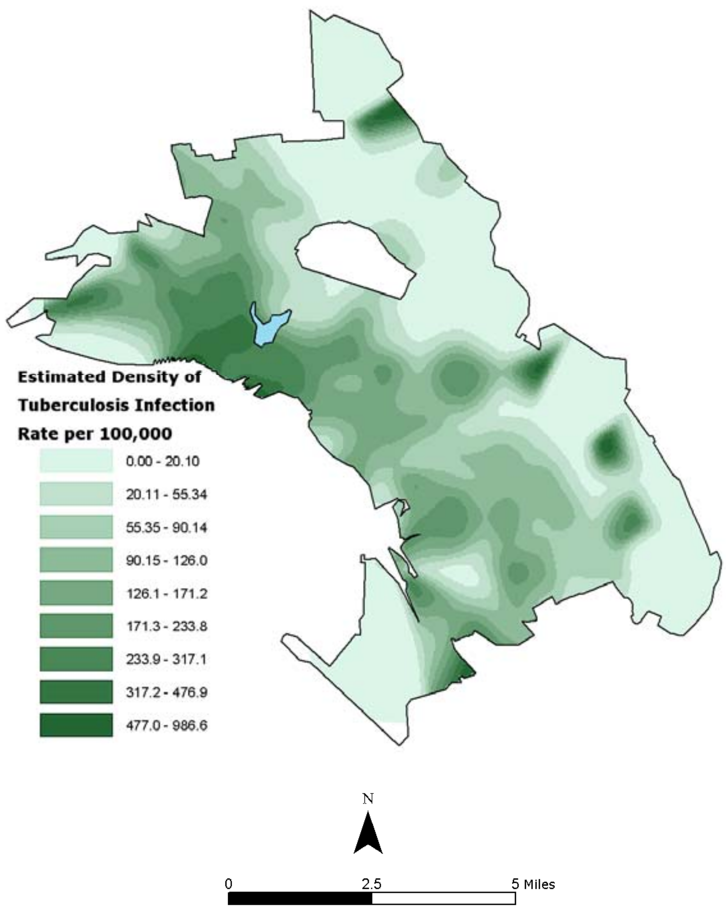


Thus, an estimate of a ratio density surface of disease relative to the population is necessary, and that is what Map 9 provides. The rate of tuberculosis infection is derived by dividing the estimated density of the number of tuberculosis cases by the estimated density of the total population (essentially dividing the data in Map 5 by the data in Map 7) and multiplying the data by 100,000 to derive a rate per 100,000 persons. Map 10 is the same as Map 9, again with the locations of tuberculosis cases represented as dots overlaid on the contours of infection rate.

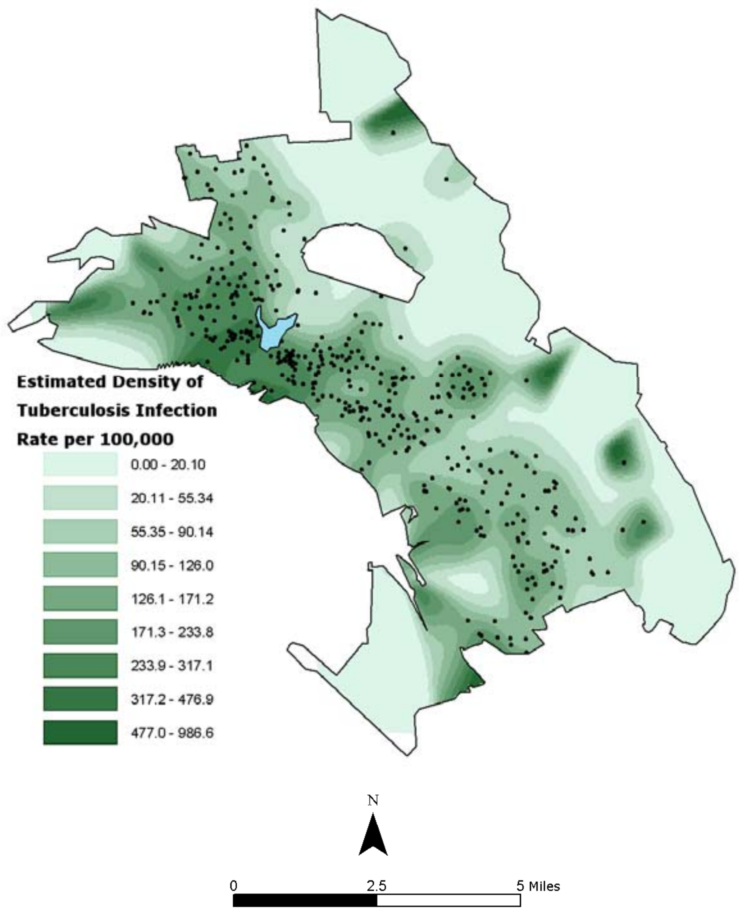
Findings

The effect of adjusting the disease distribution for the underlying population becomes quite visible. Whereas the concentration of tuberculosis cases is highest in several different areas of the city, the infection rate is highest in the area immediately west of Lake Merritt (which includes Downtown).

Map 9: Estimated Density of Tuberculosis Infection Rate per 100,000



Map 10: Estimated Density of Tuberculosis Infection Rate per 100,000 Overlaid with Locations of Infection

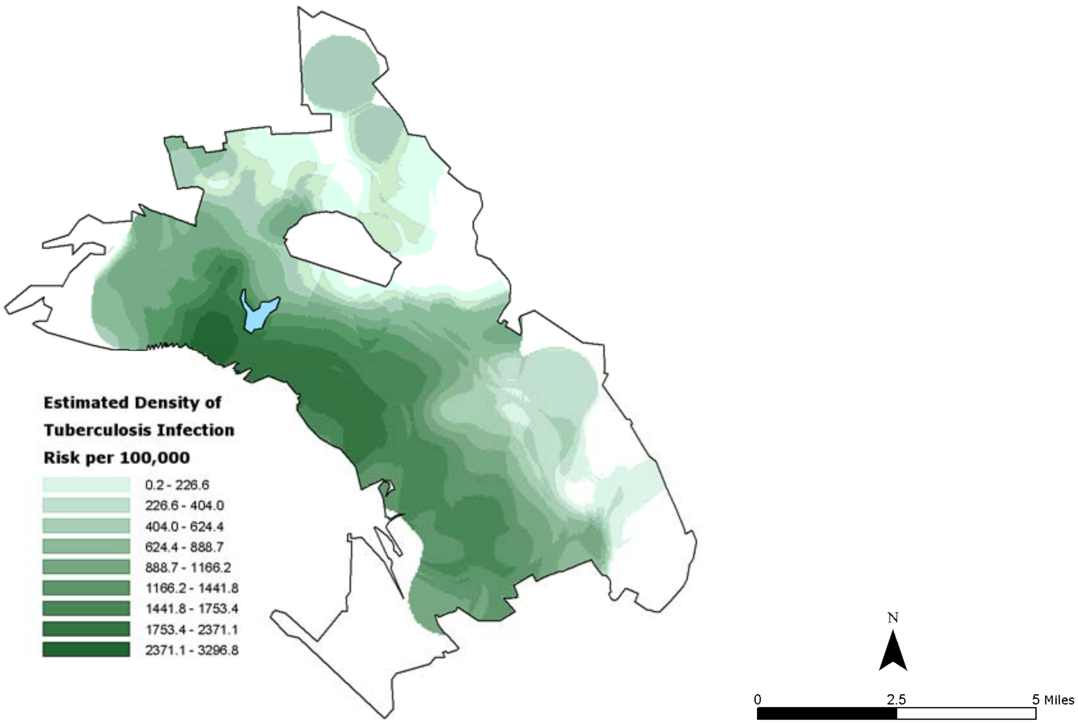


Risk Map

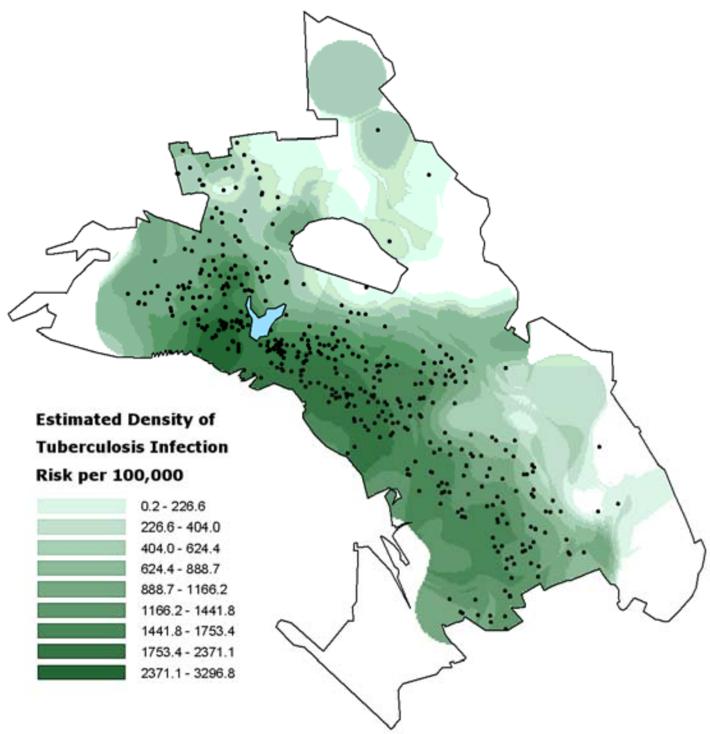
Map 11 is a "risk" map that shows where there is a more or less likelihood of tuberculosis infection based on the six demographic factors; Map 12 is the same with the locations of tuberculosis infection represented by dots overlaid on the risk contours. An equation that attempts to predict this likelihood of infection was derived from a multivariate regression, and values were calculated for the entire city of Oakland using the six variables or transformations of them. The numbers represent how many persons are more likely to be infected with tuberculosis per 100,000 persons.

Findings
This map shows where prevention efforts should be more targeted based on demographic factors. Specifically, the Downtown and Chinatown areas and neighboring areas reaching into West Oakland, as well as the Lower San Antonio and Fruitvale are all targets for prevention activities. Again, a street layer or neighborhood layer can be overlaid to identify more specific places if necessary.

Map 11: Estimated Density of Tuberculosis Infection Risk per 100,000



Map 12: Estimated Density of Tuberculosis Infection Risk per 100,000 Overlaid with Locations of Infection



Summary of spatial analysis findings

The estimated density map of tuberculosis cases shows Downtown and the Lower San Antonio area immediately south of Lake Merritt, among other areas in Oakland, to be "hot spots" of tuberculosis infection. Adjusting the disease distribution for the underlying population, the estimated density map of infection rates reveals that the infection rate is highest in the area immediately west of Lake Merritt (which includes Downtown). However, the risk map shows that Downtown and Chinatown and neighboring areas reaching into West Oakland, as well as the Lower San Antonio and Fruitvale, are areas where prevention efforts should be more targeted based on demographic factors. A layer of streets or neighborhoods can be overlaid on the contour maps to obtain more exact locations.

The contour maps derived from kernel density estimations provide much more information about the location and context of tuberculosis infection in Oakland, than the chloropleth maps. The differently shaded contours of the maps have direct correspondences to the locations of tuberculosis infection, areas of different rates of infection, and areas of different risks of infection. Prevention and intervention efforts guided by this analysis would lead one to concentrate on specific neighborhoods and sub-areas where tuberculosis is most prevalent, rather than on pre-defined geographies such as census tracts

Efficacy of New Methods

This study has shown that combining regression analysis with kernel density estimation can be a powerful way to analyze disease infection geographically, as well as associate infection rates with demographic variables representing socio-economic conditions on a local level.

For this study, the kernel density estimation method was used to interpolate the tuberculosis location data for the entire city. Estimations were also calculated for the various demographic variables for the entire city. The resulting data were mapped and used to run the multivariate regression analysis to test specific hypotheses. The results from the multivariate regression were then also used to derive an equation to calculate the risk of infection based on the demographic variables. This risk of infection was calculated for the entire city of Oakland, and the resulting data were mapped.

The analysis utilized 69,181 geographic areas, allowing for much greater statistical power than would have been possible hade we used Census Tracts (n=106) or Census Block Groups (n=337). Consequently we were able to very specifically identify areas where the risk of TB is greatest. The resulting contour maps may be particularly useful for the design of targeted community intervention and prevention programs.

Although more research and development of the kernel density estimation technique is needed and will certainly occur in the next few years, it has proven to be an effective tool for spatial statistical analysis. Combined with simple and multivariate regressions, this technique has the potential to be a powerful tool in analyzing tuberculosis infection in Oakland, particularly in helping to identify areas of high infection rates where intervention should be directed and areas of high risk where prevention should be targeted. For analyzing disease infection, the technique is an effective way of conducting both "hot spot" analysis as well as being able to link the "hot spots" to an underlying population-at-risk.

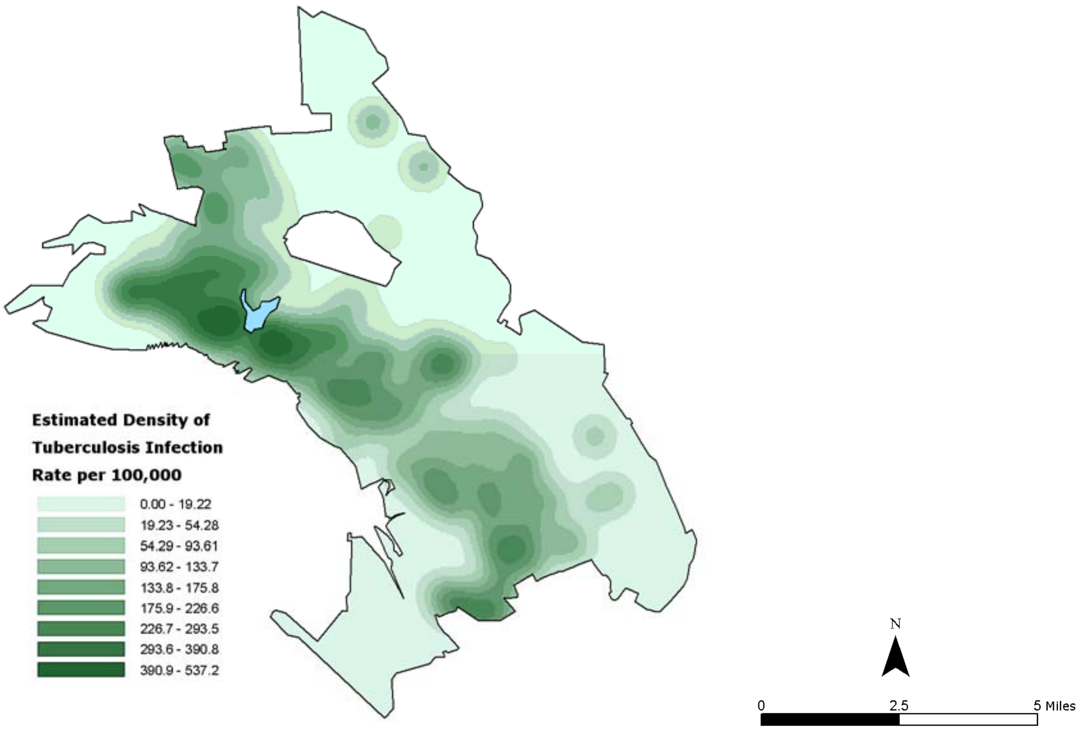
Research on the use of the kernel density estimation technique in both statistical theory and in developing applications is currently being conducted. For example, significance testing of the density estimates needs to be developed. Such techniques tend to focus on simulating surfaces under spatially random assumptions. Because of the still experimental nature of the testing, CrimeStat II does not include any testing of density estimates, although the GeoStatistical Analyst extension for ArcMap produced by ESRI has some rudimentary testing methods.

One disadvantage in our use of the kernel density estimation method is revealed by the "spikes" of data in the tuberculosis infection rate maps (Maps 9 and 10), that is, the areas where there are seemingly high infection rates even though the number of infection incidents are low. These are particularly noticeable on the eastern edge of the city, the "hills" of Oakland. This occurs because the ratio of tuberculosis incidence and total population blew up and became very large numbers when the density estimate for a particularly cell in the denominator (that of total population) approaches zero. In future studies, the denominator may be smoothed more than the numerator in order to reduce these spikes. Experimentation with using different intervals for the numerator and the denominator will be necessary for this. Experience has shown that these spikes occur when either there are two few cases or there is an irregular boundary to the region with a number of incidents grouped at one of the edges.

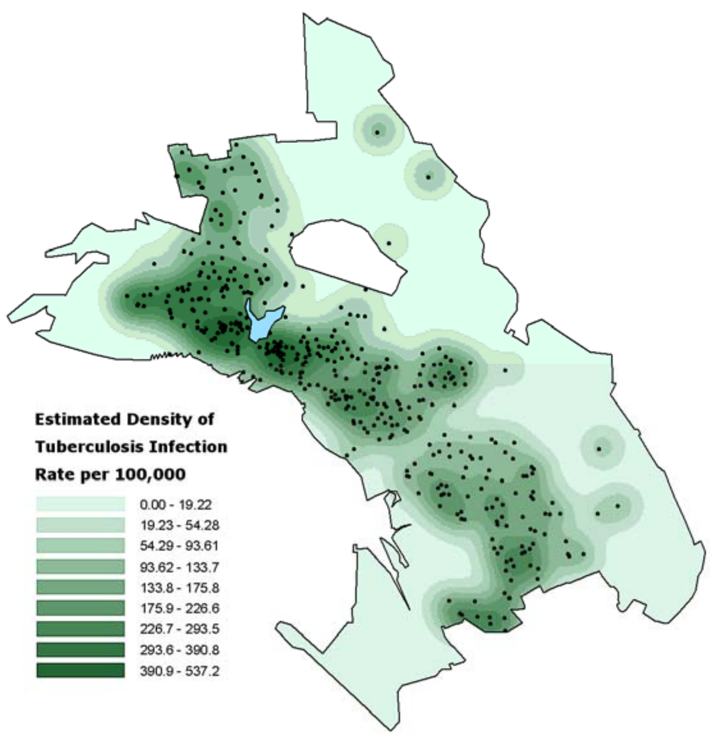
Provided here for comparison are maps (Maps 13 and 14) where the fixed interval used for the density estimation of the total population was 1 mile, instead of the 0.253 miles used for all the other maps, producing a more smoothed denominator for the tuberculosis infection rate calculation. And the result is definitely a reduction in the number of spikes of data and a more smooth representation of the data overall. However, these estimations were not used in our regression analysis, and the maps are provided here only for comparison.

As demonstrated in our study, kernel density estimation is a particularly useful method as it helps to identify the location, spatial extent, and intensity of disease "hot spots," as well as providing excellent techniques for identifying risk when combined with regression analysis. It is also visually attractive, helping to invoke further inquiry and the reasoning behind why disease is concentrated. The method is also less subjective if clear guidelines are followed for the setting of parameters.

Map 13: Estimated Density of Tuberculosis Infection Rate per 100,000



Map 14: Estimated Density of Tuberculosis Infection Rate per 100,000 Overlaid with Locations of Infection



VIII. OVERALL SUMMARY OF FINDINGS

Our findings from the multivariate analysis confirmed our initial hypotheses. We found a positive association between the incidence of tuberculosis and high neighborhood concentrations of foreign-born and non-white residents, residents in poverty and over-crowded housing. We found a small, but statistically significant, association between over-crowded housing rates and the incidence of tuberculosis when all other variables were held constant.

The spatial analysis found highest concentrations of tuberculosis cases in two neighborhoods, one in Downtown and the other in the Lower San Antonio area immediately south of Lake Merritt. Adjusting the disease distribution for the underlying population, the estimated density map of infection rates reveals that the infection rate is highest in the Downtown area immediately west of Lake Merritt. The risk map shows that Downtown, Chinatown, and neighboring areas reaching into West Oakland, as well as the Lower San Antonio and Fruitvale, are areas where prevention efforts should be targeted.

This study found that the combination of kernel density estimation and regression analysis was a useful method to identify the location, spatial extent, and intensity of disease "hot spots," and risk of infection at the neighborhood level. The analysis utilized 69,181 geographic areas, allowing for much greater statistical power than would have been possible hade we used Census Tracts (n=106) or Census Block Groups (n=337). Consequently we were able to very specifically identify areas where the risk of tuberculosis is greatest. The resulting contour maps may be particularly useful for the design of targeted community intervention and prevention programs.

APPENDIX I

Derivation of the Regression Equation

A linear regression model is represented by the equation Y= a + bX,

Where Y' = predicted score on the dependent variable
a = Y-intercept, or point at which he regression line crosses the Y-axis
b = slope, or the change in Y for every one unit increase in X
X = score on the independent variable
From the above, we can represent multivariate equation by:
$Y' = a + b_1X_1 + b_2X_2$

Using this principle, we constructed the following equation for our regression analysis.

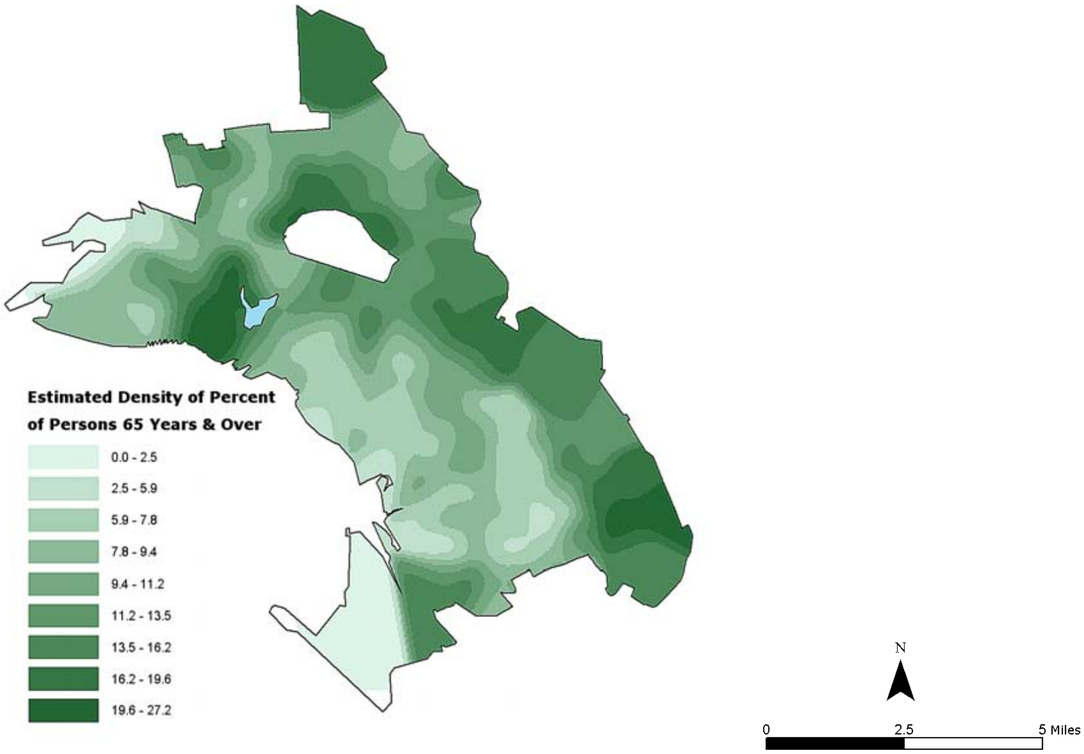
> ***To predict TB infection=***
> ***3127.742-(2966.191)(per_male)+(189.351)(logit of per_pov)+(37.795)(logit of per_nonw)+(635.067)(logit of per_imm)+(277.149)(sqr of c_crowd)+ (17725.612)(sqr of per_65)***

APPENDIX II

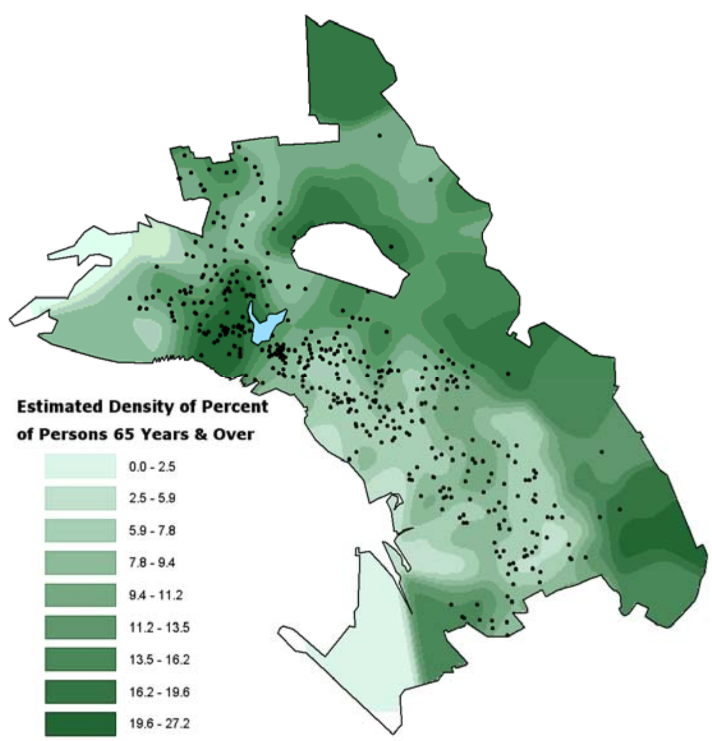
Maps of Demographic Variables Associated with Population Infected with Tuberculosis (Maps 15 to 26)

Statistical analysis of variables within the Oakland tuberculosis data set have revealed the following demographic factors to be associated with much of the population infected with tuberculosis: 1) age of 65 years old and over, 2) foreign born, 3) male, and 4) non-white race. In addition, established findings of tuberculosis infection from literature reviews indicate that 5) poverty and 6) high number of persons per room (overcrowding) also are important indicators of high tuberculosis infection. Therefore, we extracted data on a block group level pertinent to the above six factors from the 2000 census and mapped them using the kernel density routine, using the relevant variable as the intensity variable and the block group centroids as the location points. The even numbered maps are the same as the odd numbered maps, with the locations of tuberculosis infection cases, represented by dots, overlaid over the contours of the census data. These maps are provided here so that visual comparisons can be made between the distribution of the population represented by these variables and the locations of tuberculosis infection.

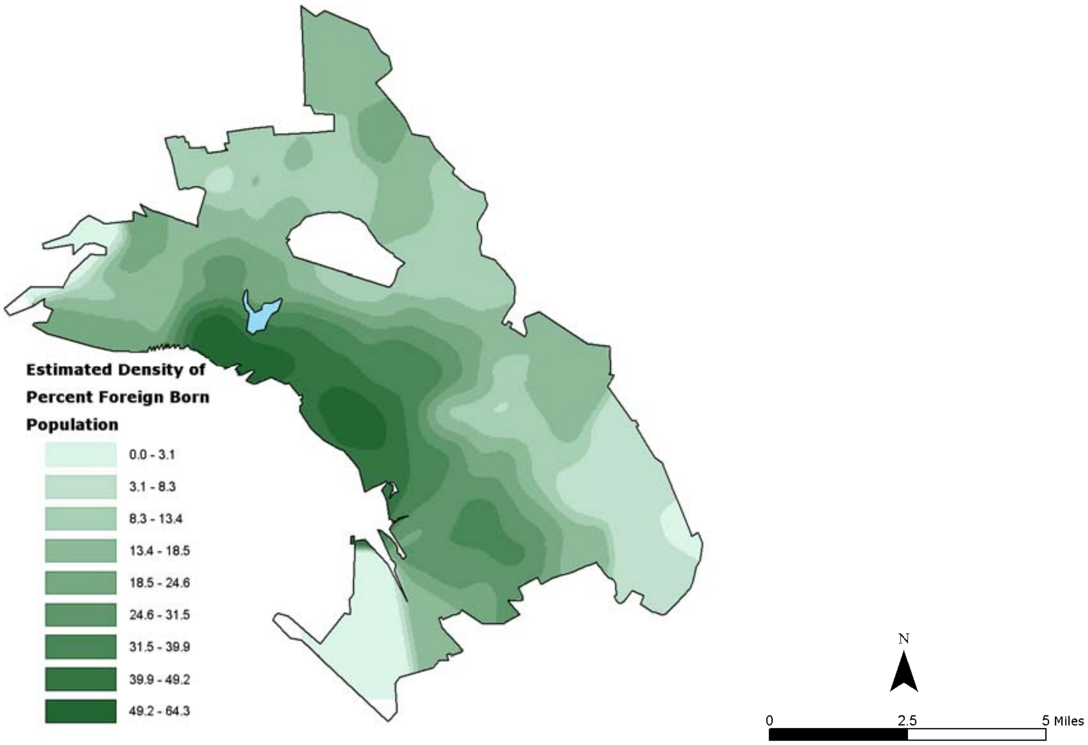
Map 15: Estimated Density of Percent of Persons 65 Years and Over



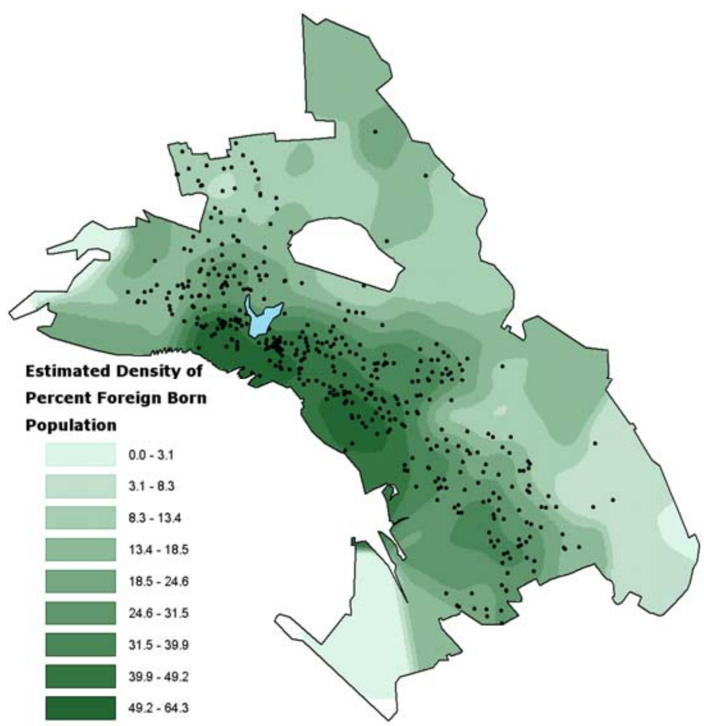
Map 16: Estimated Density of Percent of Persons 65 Years and Over Overlaid with Locations of Infection



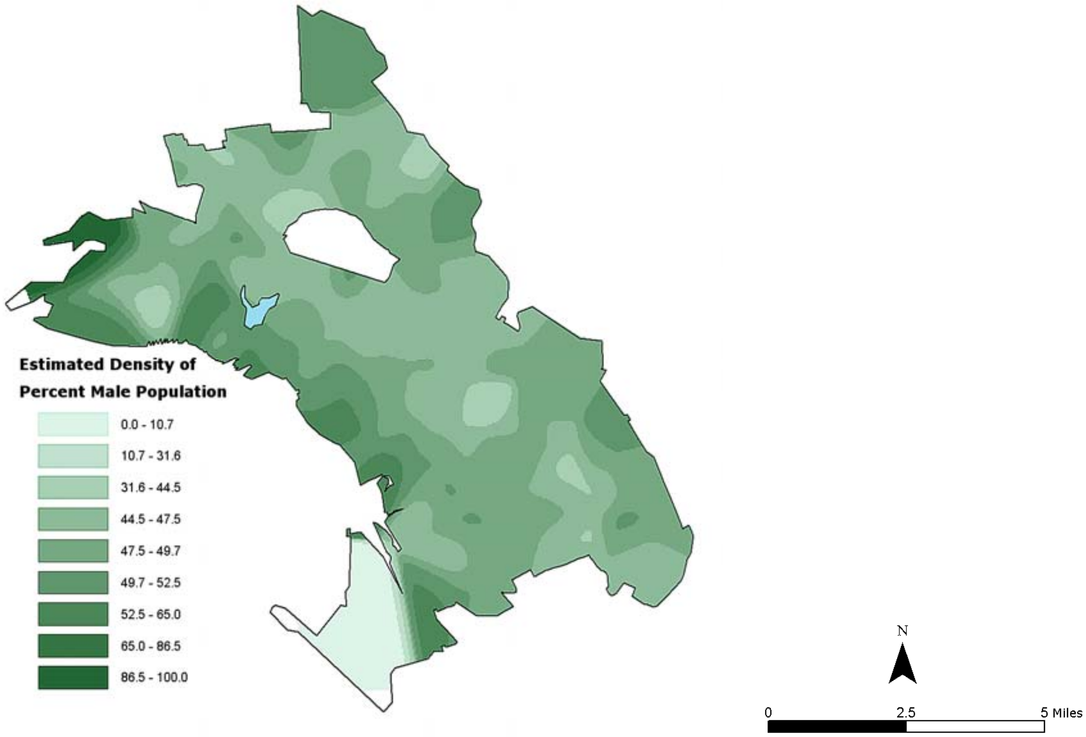
Map 17: Estimated Density of Percent Foreign Born Population



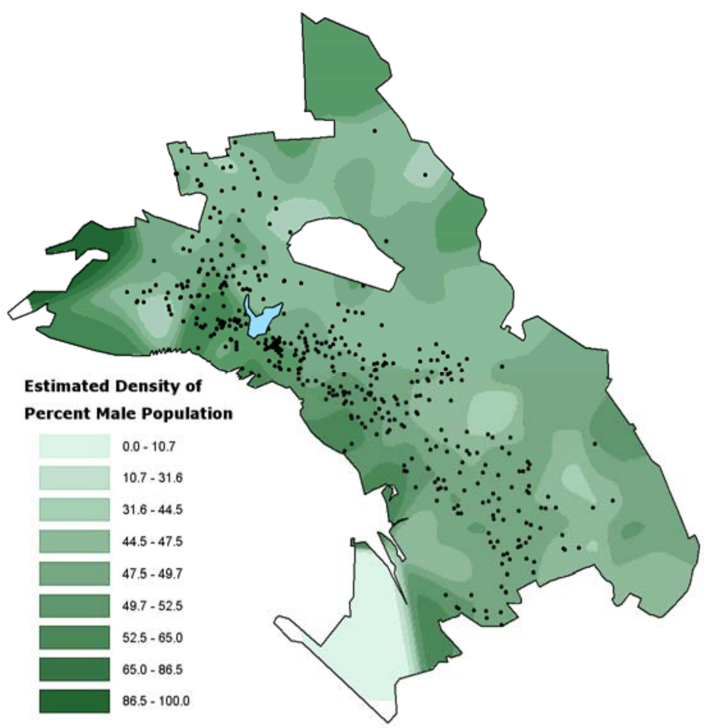
Map 18: Estimated Density of Percent Foreign Born Population Overlaid with Locations of Infection



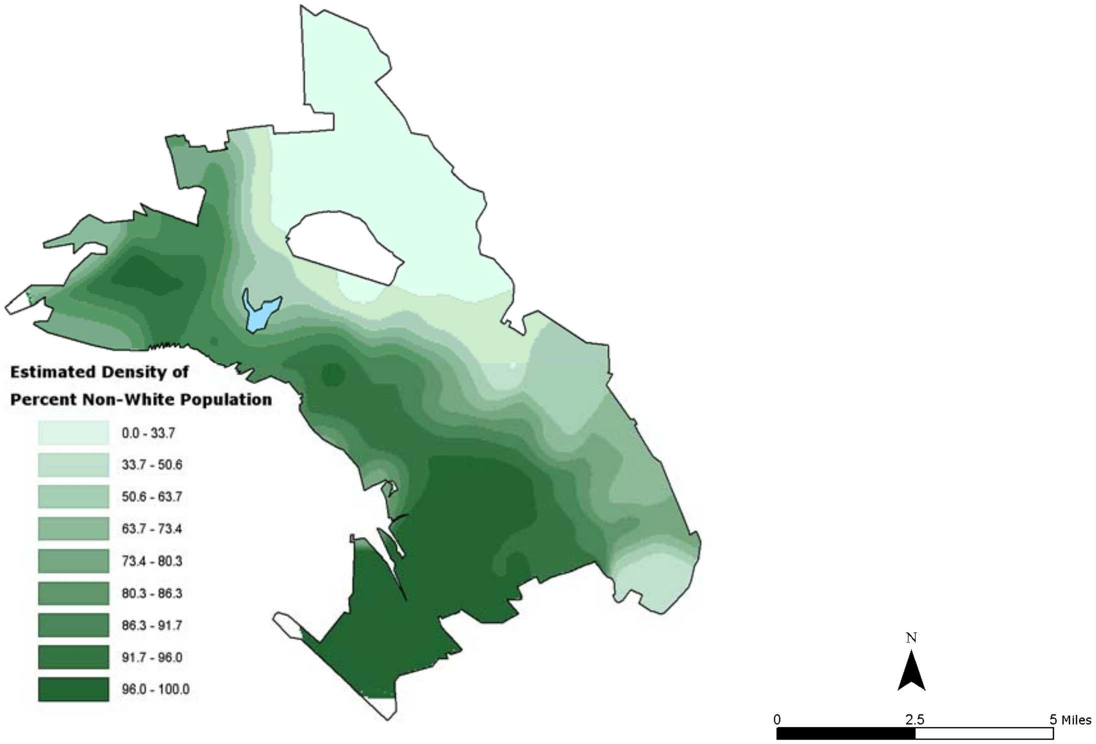
Map 19: Estimated Density of Percent Male Population



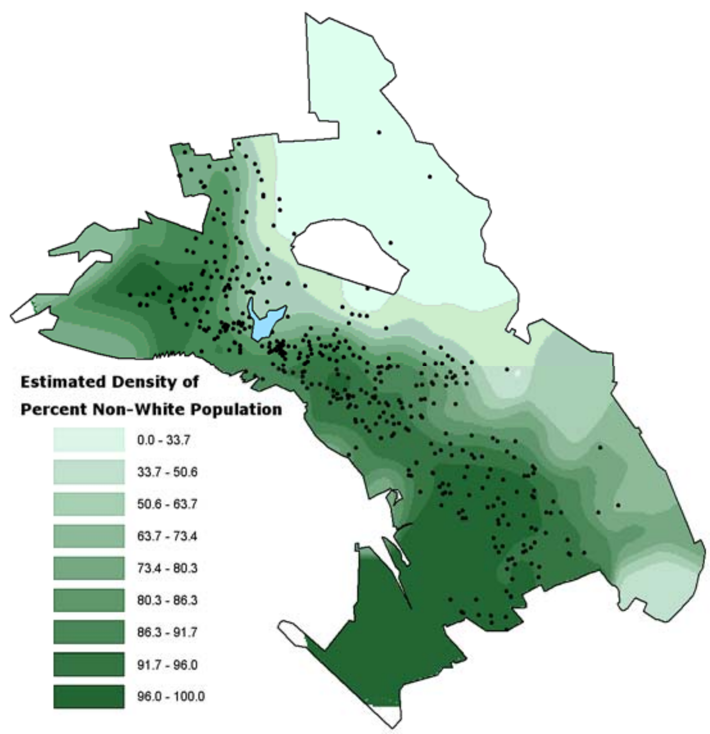
Map 20: Estimated Density of Percent Male Population Overlaid with Locations of Infection



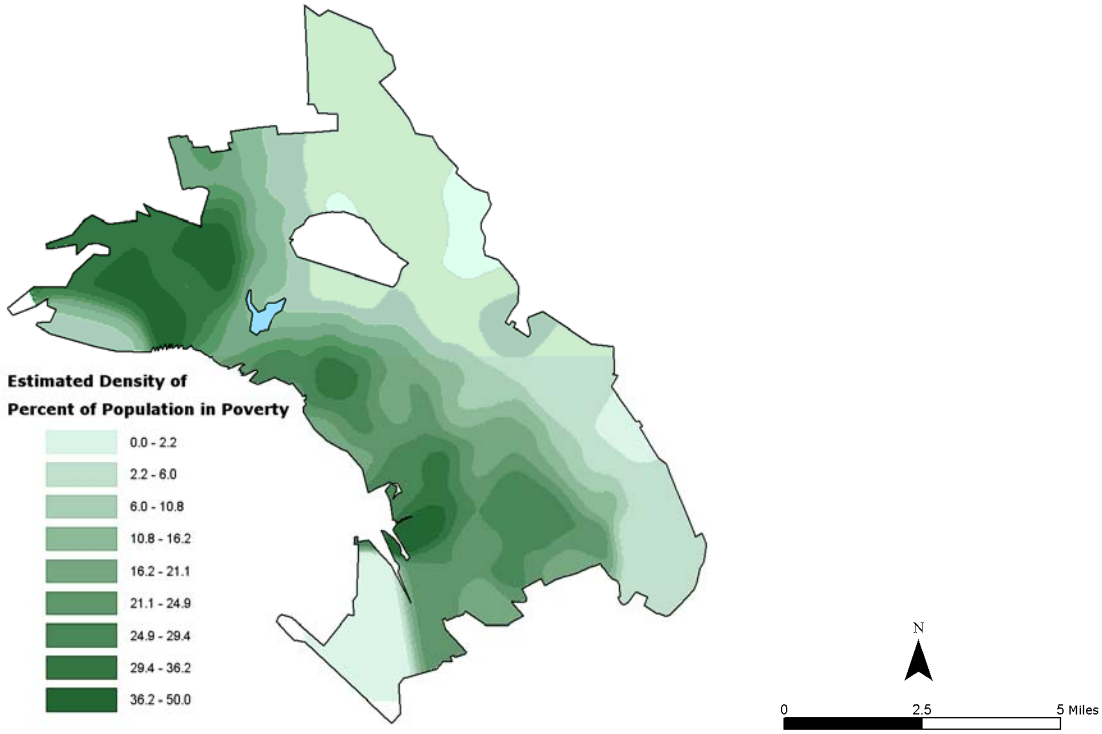
Map 21: Estimated Density of Percent Non-White Population



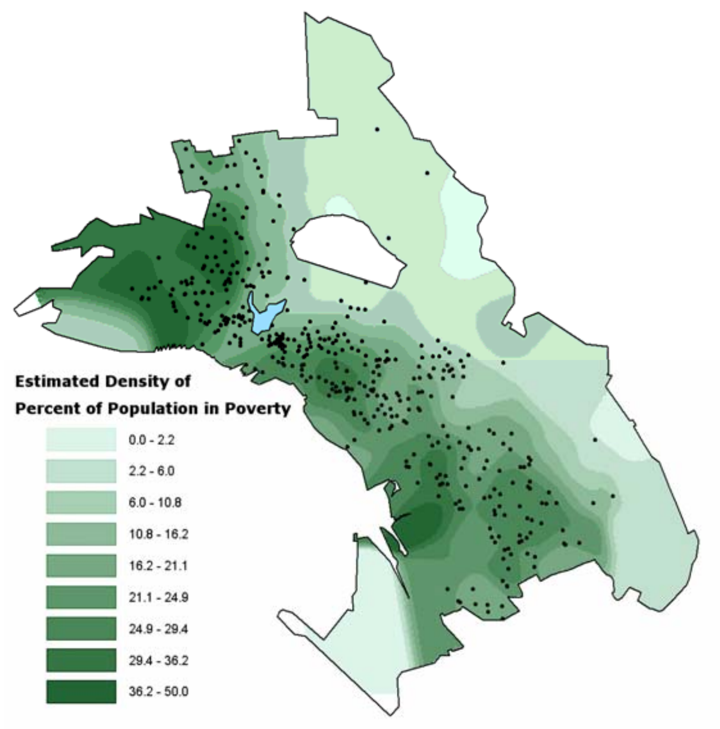
Map 22: Estimated Density of Percent Non-White Population Overlaid with Locations of Infection



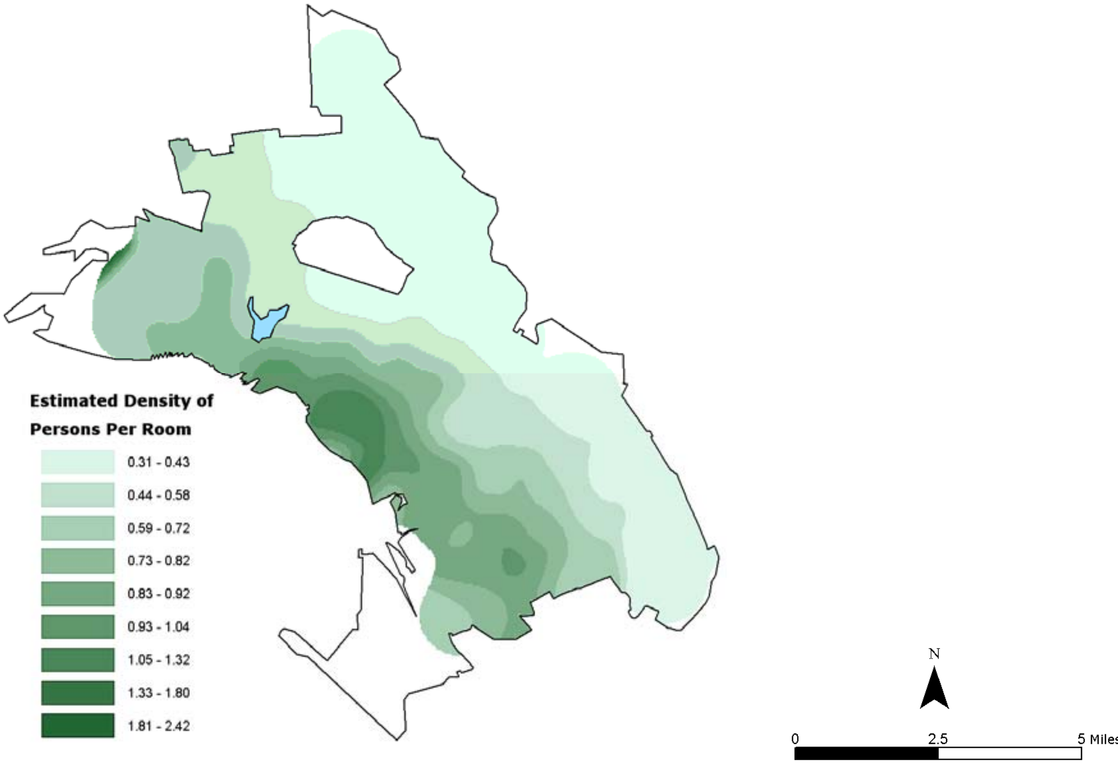
Map 23: Estimated Density of Percent of Population in Poverty



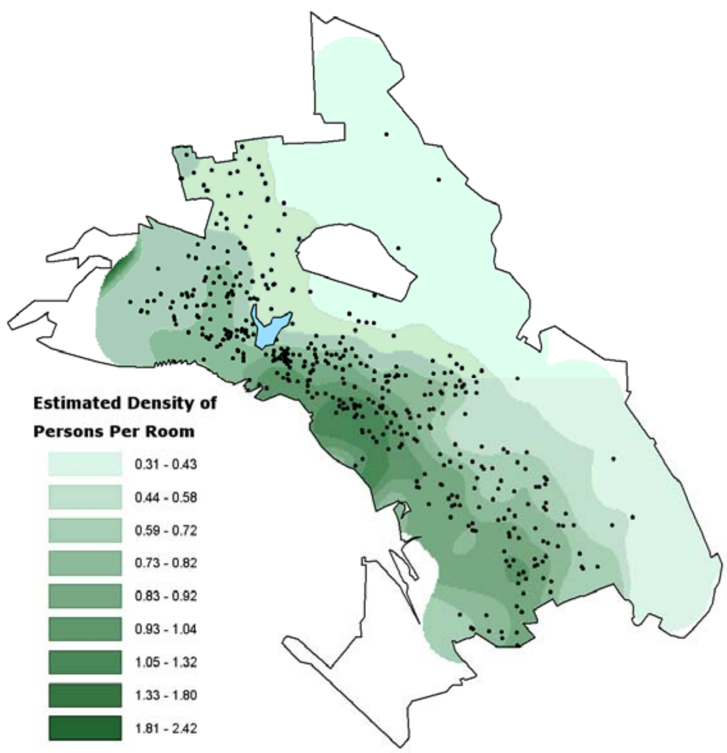
Map 24: Estimated Density of Percent of Population in Poverty Overlaid with Locations of Infection



Map 25: Estimated Density of Persons Per Room



Map 26: Estimated Density of Persons Per Room Overlaid with Locations of Infection



REFERENCES

American Lung Association. Tuberculosis. http://www.lungusa.org/diseases/lungtb.html#who

Barnes, P.F.; Yang, Z.; Preston-Martin, S.; Pogoda, J.M.; Jones, B.E.; Otaya, M.; Eisenach, K.D.; Knowles, L.; Harvey, S.; Cave, M.D. (1997) Patterns of tuberculosis transmission in central Los Angeles. *JAMA*. 278(14): 1159-63.

Center for Disease Control. **http://www.cdc.gov/nchstp/tb/worldtb2002/race.htm**

Centers for Disease Control and Prevention (1990) Tuberculosis among foreign-born persons entering the United States: recommendations of the advisory committee for elimination of tuberculosis. *MMWR*. 39(RR-18): 1-21.

Graham Barr, A.V. Diez-Roux, CA Knirsch, A. Pablos-Méndez Neighborhood Poverty and the Resurgence of Tuberculosis in New York City, 1984-1992 http://www.hopkins-tb.org/news/11-05-2001.shtml#006

Kawachi, I; Kennedy, B. P and Wilkinson, R. G eds. (1999). The society and population health reader Volume 1: income inequality and health, The New Press, New York.

Neter, J., Kutner. M.H., Nachtsheim, C.J., & Wasserman, W. (1996). Applied linear statistical models. Boston MA: McGraw-Hill.

Robert L. Cowie, MD; Jill W. Sharpe, BN Tuberculosis among immigrants: interval from arrival in Canada to diagnosis A 5-year study in southern Alberta. *CMAJ* 1998;158:599-602. http://collection.nlc-bnc.ca/100/201/300/cdn_medical_association/cmaj/vol-158/issue-5/0599.htm

Ruiz, J.D.; Molitor,F.; McFarland,W.; Klausner, J.; Lemp, G.; Page-Shafer, K.; Parikh-Patel, A.; Morrow, S.; and Sun, R.K. (2000) Prevalence of HIV infection, sexually transmitted diseases, and hepatitis and related risk behavior in young women living in low-income neighborhoods of northern California. *Western Journal of Medicine* 172, no. 6: 368-73.

Stone, M. (2001) The Utility of Geographical Information Systems (GIS) and Spatial Analysis In Tuberculosis Surveillance in Harris County, Texas, 1995-1998. http://www.esri.com/library/userconf/health01/papers/hc01_p02a/hc01_p02a.html.

Sumartojo, E. (2000). Structural factors in HIV prevention: Concepts, examples, and implications for research. *AIDS,* 14 (Suppl. 1), S3-S10.

Talbot EA, Moore M, McCray E, Binkin NJ. Tuberculosis among foreign-born persons in the United States, 1993-1998.
Troels Lillebaek, Åse B. Andersen, Asger Dirksen, Else Smith, Lene T. Skovgaard, and Axel Kok-Jensen. Persistent High Incidence of Tuberculosis in Immigrants in a Low-Incidence Country. http://www.cdc.gov/ncidod/eid/vol8no7/01-0482.htm

World Health Organization (1997). Anti-tuberculosis Drug Resistance in the World: The WHO/IUATLD Global Project on Anti-tuberculosis Drug Resistance Surveillance 1994-1997, WHO/TB/97.229.

"Tuberculosis is a serious threat" March 22, 2001.
http://www.action.org/sf/media/oaktrib20010322tb.html