Small-Area Life Expectancy: A Comparison of Methods, Relationship to Neighborhood Sociodemographic Factors and Outlier Analysis

Life expectancy (LE) has become increasingly popular as a tool for demonstrating health disparities across countries, US states, and small geographic areas where area-level characteristics may, in part, account for variability in health outcomes. LE varies substantially across U.S. communities and even small areas such as neighborhoods [1, 2]. However, numerous methodological challenges arise when life table methods, which are well suited to larger populations, are applied to smaller populations [3-5]. Additionally, localized, place-based characteristics play an important role in LE and other measures of population health. To date, only a few studies have used a fine geographic scale to identify local assets that might enhance health outcomes [6-8]. Recently, there has been increased interest by policymakers and health officials to improve and refine methods to assess those small-area factors that influence population and neighborhood health. Likewise, there is also an increased need to understand outlier communities—areas that have unusually high or low LE or other health measures that would not be predicted based on their socioeconomic, environmental, and demographic profile.

Therefore, there are three related objectives of our study. First, we compare two methods of LE calculation for census tracts in California: Poisson modeling and traditional life table approaches. This analysis will help us understand which method produces the most stable and valid LE estimates, as well as to understand the advantages and drawbacks of each. Although a substantial amount of research has examined these advantages and drawbacks, few have done so on a fine geographic scale.

Second, we assess the association between poverty and LE on the census tract level. This objective largely will validate previous studies that have established the negative association between poverty and LE. In this objective, we not only examine poverty, race/ethnicity, and

other demographic characteristics, we also examine the potential for interactions of these characteristics to influence life expectancy.

Third, we obtain a set of positive outlier census tracts based on residuals from the LE-poverty models described in the previous objective. These outliers represent census tracts that have unusually high life expectancy given their socioeconomic and demographic characteristics that would otherwise predict lower LE. Using those outliers, we then evaluate the associations between a detailed and comprehensive list of socioeconomic, demographic, and environmental factors and "outlier status".

Methods

OBJECTIVE 1: COMPARISON OF LIFE EXPECTANCY CALCULATION METHODS:

We compared two approaches to estimate LE at the census tract level: the actuarial life table method and Poisson modeling. We abstracted all 1999-2001 and 2009-2011 death records from California vital statistics and geocoded each record to the census tract of the decedent's last residence. Records were aggregated by age group and census tract and combined with 2000 and 2010 US Census Bureau population data, respectively. Age-specific mortality rates were calculated using average annual deaths over the three years divided by the census population count for that year. We used those rates to calculate abridged life tables and resultant census-tract life expectancies at birth with both methods: the abridged life tables and Poisson modeling. Chang's method was used to calculate the variance of LE [9]. The Poisson modeling approach has three advantages for estimating LE of small area: pooling multiple year data, borrowing strengths across geospatial units, and incorporating relationships between covariates. We estimated LE by fitting mixed-effect Poisson regressions based on the same spatial and age

distribution of deaths in California [10, 11]. We also compared the LEs produced by these two methods across several characteristics: consistency across point estimates, standard deviation and confidence bounds, and number of census tract estimates marked as unreliable based on extreme population sizes or tabulated deaths.

Calculation of Life Expectancy

Life Table Methods

Death records for each census tract were aggregated by 13 age groups and combined with US Census Bureau population data. Age-specific mortality rates were calculated using average annual deaths over the three years divided by the respective census population count. We used those rates to calculate abridged life tables and resultant census-tract LEs at birth detailed by the following formula (Chiang, 1984):

$$l_{x+n} = {}_n l_x - {}_n d_x, L_0 = l_1 + a_0 \times d_0 = l_1 + a_0 \times (l_0 - l_1),$$

$${}_n L_x = \frac{n \times (l_x + l_{x+n})}{2}, L_{85} = \frac{l_{85}}{m_{85-}}, T_x = \sum {}_n L_x, e_x = \frac{T_x}{l_x}$$

where x indicates the starting point for an age interval; n is the interval length and m is mortality rate; q is mortality probability; l is expected survival population; d is expected number of deaths; L is survival person-years; T is total survival person-years; E is expected remaining life time; $0 < a_0 < 1$ is a coefficient. We use infant mortality rate for 0-1 age group and the mortality probability for the last age group is set as 1. While some life tables are partitioned into one-year age groups, this study uses 13 age-groups due to the structure of the U.S. Census Bureau's American Community Survey (as accessed through the American FactFinder) allowing for reliable estimation over the broad range of subdomains considered.

Assume that grouped data in the age intervals are independent across intervals and total deaths, D_i, within interval *i* follows a binomial distribution with probability q_i but with unknown number of independent trials. Chiang (1984) developed Taylor-Linearization methods to estimate the variance of LE as

$$\text{var}(\hat{e}_a) = \sum_{i=a}^{w-1} \hat{p}_{ai}^2 [(1-a_i)n_i + \hat{e}_{i+1}]^2 \hat{V}(\hat{p}_i)$$

where w is the final age interval; P_{ai} is the probability of surviving to age x_i given survival to age x_a; $q_i = 1 - p_i, \hat{V}(\hat{p}_i) = \hat{q}_i^2(1-\hat{q}_i)/D_i$.

Poisson modeling

A Poisson modeling approach was also used to estimate LE at birth, which, as previously stated, is advantageous in estimating the LE of small areas. We estimated LE by fitting mixed-effect Poisson regressions based on the same spatial and age distribution of deaths in California [2, 3]. For small areas, Poisson modeling fits better than logistic regression because the latter is biased when the outcome is rare. The model is specified as below:

$$\log(y_{ik}) = \beta_0 + \beta_1 x_{1ik} + \sum_{j=2}^{24} \beta_j x_{jik} + \sum_{j=25}^{33} \beta_j x_{jik} + z_{ik} a_k, i = 1, \cdots, 6850, k = 1, \cdots, 57$$

where y_{ik} is the death count within census tract i and county k.

x_{1ik} is the poverty measure;

x_{jik} (j = 2,…,11) is the population relative frequency of the j^{th} age group;

x_{jik} (j = 12,…,15) is the population relative frequency of the j^{th} marital status group (divorced, never married, married, and separated);

x_{16ik} is the population relative frequency of females;

x_{jik} (j=17,18,19) is the population relative frequency of the j^{th} educational attainment group (less than high school, high school, and some college or associate degree);

x_{20ik} is the population relative frequency of the native population;
x_{21ik} is the population relative frequency of Hispanic ethnicity.

The parameter β_0 is an intercept term, $\beta_i (i = 1, \cdots, 21)$ is the respective slope of the $x_i (i = 1, \cdots, 21)$; Z_{ik} is the random effect for the i^{th} census tract and j^{th} county in California; and a_k is the random coefficient effect for the k^{th} county.

OBJECTIVE 2: SMALL-AREA LIFE EXPECTANCY AND POVERTY

Life table methods were ultimately selected and were used to calculate the LEs of census tracts with a population of at least 3750 individuals, ≥20 deaths, and at least 50% of population not living group quarters such as correctional institutions, nursing home, college dormitories including college quarters off campus, or military quarters. This results in a sample of 6,670 and 7,654 census tracts used in this analysis for 2000 and 2010, respectively. A parsimonious weighted least squares (WLS) regression (model 1) was selected among the following covariates: poverty, education, age, gender, marital status, race, immigrant status, and Hispanic origin from the US Census using reciprocal variance of the life expectancy as regression weights. The variables were selected *a priori* as major predictors of life expectancy in previous studies.

Small-area life expectancy, poverty, and other demographic characteristics

Weighted least squares regression

The dependent variable (LE) is continuous and census tracts have differential variability in LE. Therefore, a WLS is suitable for analysis of such data. The maximum likelihood estimates of the linear component parameters are calculated using SAS/STAT®, version 9.4, PROC GLM. ArcMap v. 10.1 (ESRI, Redlands, CA) was used for mapping.

The model for a given census tract is

$$y_i = \beta_0 + \beta_1 x_1 + \sum_{j=2}^{24} \beta_j x_j + \sum_{j=25}^{33} \beta_j x_j + \varepsilon_i, \text{var}(\varepsilon_i) \sim N(0, \frac{\sigma^2}{w_i}), i = 1, \cdots, 6850,$$

where w_i (i=1,…,6850) is the standard deviation of LE for the i[th] census tract in California;

x_1 is the poverty measure;

x_j (j = 2,…,11) is the population relative frequency of the j^{th} age group;

x_j (j = 12,…,15) is the population relative frequency of the j^{th} marital status group (divorced, never married, married, and separated);

x_{16} is the population relative frequency of females;

x_j (j=17,18,19) is the population relative frequency of the j^{th} educational attainment group (less than high school, high school, and some college or associate degree);

x_{20} is the population relative frequency of the native population;

x_{21} is the population relative frequency of Hispanic ethnicity.

The parameter β_0 is an intercept term, $\beta_i (i = 1, \cdots, 21)$ is the respective slopes of the $x_i (i = 1, \cdots, 21)$.

Associations between life expectancy and poverty

Poverty was measured as the proportion of persons with a total household income below 200% of the federal poverty level. LE was negatively associated with poverty with a correlation of -0.47. The association is illustrated in Figure 1. The scatterplot was smoothed using a 30-point moving average after truncating 15 points on each side, left and right. The relationship between LE and poverty was not linear as would be expected, with the mean LE (approximated by the moving average) plateauing at a high LE on the left and leveling out at a low LE on the right.

OBJECTIVE 3: OUTLIER ANALYSIS

The analysis then examined outlier census tracts to ascertain those community-level assets that may contribute to unexpectedly high life expectancy even after controlling for poverty. In this phase of the analysis, census-tract life expectancies were categorized as three classes:

unexpectedly low LE with a standardized residual value less than or equal to -2.576, unexpectedly high LE with a standardized residual value greater than or equal to 2.576, and expectedly normal LE with an absolute standardized residual value between -2.576 and 2.576. We then used an extensive list of predictor variables from the US Census, and numerous other state and national databases on socioeconomic, environmental, demographic, and health-related factors (Appendix A). A weighted multinomial logistic regression was modeled against each predictor controlling for the significant variables in model 1. WLS is an effective method to make good use of small-area data sets and can handle data points with varying quality in regression analysis. Weights used in the analysis were the reciprocal of the variance associated with the LE calculated in the Phase 1 regression models.

Results

In this section, we first evaluate models 1 and 2 and then summarize the results from these models. We also compare the results from model 1 with those from model 2. Finally, we document the results of outlier analysis.

Objective 1: Comparison of life expectancy calculation methods

The life expectancies produced by these two methods were compared across several characteristics: consistency across point estimates, standard deviation and confidence bounds, number of census tract estimates marked as unreliable based on extreme population sizes or tabulated deaths, and adjusted R-squared values. Summary statistics for both methods and their comparison are shown in Tables 1 and 2. The results, and the respective abilities of the models to predict life expectancy, suggest that the life table approach is more reliable than Poisson modeling. Census tract-specific life expectancies for 2010 are displayed in Figure 1 using this approach.

Figures 2A and 2B indicate that the scatter plots for observed LE against predicted LE and the difference between observed LE and predicted LE by 13 age groups using Poisson modeling. Figure 3A indicates that predicted LE from Poisson modeling overestimates the observed LE. The distribution of the differences between observed LE and predicted LE are centered at zero for the first eleven age groups, but spreads out for the last two age groups. It appears there are some potential outliers in the last two age groups.

Based on these results, the subsequent analyses will be conducted using the life table approach for calculating LE.

Objective 2: Small-area life expectancy and poverty ("Phase 1")

The results showed a monotonic negative association: life expectancy decreased as the population with incomes below 200% of the Federal poverty threshold increased, and this association held even after adjusting for confounders. Overall, 71% of the variability in life expectancy was explained by the Phase 1 variables (Table 3). Scatterplots are shown to illustrate how outliers were identified (census tracts with large standardized residuals that depart from the regression equation) and eliminated outliers suspected to be erroneous based on implausible LE values (Figure 4).

Objective 3: Outlier analysis ("Phase 2")

We identified a listing of area-based measures (Appendix A) with the potential to account for some of the outlier census tract's higher than expected life expectancies, including environmental, socioeconomic, detailed demographic, and health-related variables, beyond the variables used in the Phase 1 model. Some variables are modifiable, while others are non-modifiable, including many demographic variables. Recall that weighted multinomial logistic regression was used to model outlier status as a function of these variables.

Prediction of Negative Outliers (Low LE)

Table 4 displays the frequencies of both positive and negative outliers for data 2000 and 2010. Some of the strongest predictors of negative outliers (unexpectedly low LE given low poverty levels) were the ratio of whites to non-whites (odds ratio (OR) 0.85, 90% CI [0.72, 0.98]), and percent of population living in nursing homes (OR 23.4, 90% CI 7.84, 71.4), although census tracts with a majority of nursing home residents were removed a priori from the analysis. Positive predictors were less common, however.

Prediction of Positive Outliers (High LE)

Much of the remainder of the analysis focuses on community assets that predicted positive outliers- those census tracts with unexpectedly high life expectancy after controlling for the basic demographic and socioeconomic variables in Phase I. First, using weighted logistic regression, we regressed positive outlier status on each of the "Phase 2" variables individually in what we define as the "bivariate analysis". Second, we conducted a multivariable logistic regression analysis of those outliers including all significant predictors ($p < 0.2$) from the bivariate analysis. The bivariate and multivariate analyses were conducted on the 2000 and 2010 data separately.

Tables 5 and 6 represent bivariate of outliers for 2000 and 2010, respectively. For the 2000 analysis, some of the strongest associations between predictors and positive outliers (unexpectedly high LE) included percent of population who bikes or walks to work (OR 1.85, 90% CI: 1.19, 2.69). This result means that, controlling the phase 1 variables, a 10 percent increase in the population who bike or walk to work will increase the likelihood of being an outlier census tract (high LE, high poverty) by 85%. Factors that significantly reduced the likelihood of outlier status included cancer risk due to air pollutants (OR 0.76, 90% CI: 0.68, 0.84), foster care rate (OR 0.06, 90% CI: 0.02, 0.25), percent of householders without children

(OR 0.59, 90% CI: 0.35, 0.99), and percent of households that were overcrowded (>= 1.01 persons/room, OR 0.45, 90% CI: 0.19, 1.00). Factors that were positively associated with outlier status included percent population living in rural areas (OR 1.23, 90% CI: 1.14, 1.33) and sex ratio (female to male) at birth (OR 1.63, 90% CI: 1.04, 2.49).

Similarly, for 2010 data some of the strongest associations between predictors and positive outliers (unexpectedly high LE) included single-parent families (OR 0.20, 90% CI: 0.11, 0.33), dual parent and single father families (OR 2.42, 90% CI: 1.69, 3.48), annual mean PM 2.5 concentrations (OR 0.54, 90% CI: 0.39, 0.75), illiteracy rate (OR 0.71, 90%CI: 0.54, 0.93), presence of college dormitories including college quarters off campus (OR 1.04, 90% CI: 1.01, 1.07), the presence of an acute care hospital (OR 2.60, 90% CI: 1.48, 4.31), The presence of a specialty clinic (OR 2.31, 90% CI: 1.32, 3.78), CT is predominantly open space (OR 2.52, 90% CI: 1.45, 4.12), and percent of the population living within 1/2 mile of a park, beach, open space, or coastline (OR 0.95, 90% CI: 0.90, 1.00).

The results of multiple logistic regression models are summarized in Tables 7 and 8 for data 2000 and 2010, respectively. After adjustment, several factors remained significant in 2000. Census tracts considered rural were about 19% more likely to be a positive outlier than more urban census tracts (OR 1.19, 90% CI: 1.08, 1.31), for example. Negative predictors of outlier status included cancer risk due to air pollutants (OR 0.65, 90% CI: 0.51, 0.84) and low travel time to work: 30 minutes or less (OR 0.65, 90% CI: 0.48, 0.86) and 30-60 minutes (OR 0.52, 90% CI: 0.35, 0.77). In 2010, the model was more robust, yielding many more significant factors associated with outlier status. In addition to several family structure variables, being a high LE/high poverty outlier was associated with the presence of at least one acute care hospital (OR 2.11, 90% CI: 1.08, 4.12) and at least one specialty clinic (OR 2.06, 90% CI: 1.07, 3.96) in

that census tract. Higher inequality, as measured through the Gini coefficient, (OR 1.67, 90% 1.12, 2.49) was also associated with an increased likelihood of being an outlier. Two environmental variables were also predictive of outlier status: open space was positively associated (OR 2.55, 90% CI: 1.39, 4.66) and PM 2.5 concentrations were negatively associated (OR 0.88, 90% CI: 0.80, 0.97).

Discussion

There were numerous options available in selecting the most appropriate manner in which to calculate LE, and several strengths and drawbacks of each method. Despite these limitations, the results of Objective 1 show that reliable LE estimation is possible when properly applied to small geographic areas, with several important caveats. One such caveat is that in the analysis each census tract was considered to be independent of the others. Therefore, spatial autocorrelation that may account for some of the observed patterns was not taken into account. Second, as with any small area estimate of population health, the predicted point estimates of LE were subject to statistical uncertainty. In addition, deaths were counted at the address of the decedent's last known residence. Therefore, the factors that may contribute to life expectancy may come from other census tracts in which decedents may work or otherwise spend more time. Lastly, there were numerous counterintuitive findings, particularly in the multivariate models (Tables 7 and 8) for both years. In the multivariate models, it is possible that highly correlated predictors may result in one predictor being a "positive" predictor of outlier status while the other may emerge as an unexpected "negative" predictor to compensate statistically. An example of this potential limitation may have occurred in 2000, where the maltreatment allegation rate was actually a positive predictor of outlier status. This unexpected finding may be due to collinearity with other, similar variables in the model that

may be associated with outlier status in the direction we would expect. Other examples include unexpected findings regarding income inequality, commuting time to work, and proximity to parks, beaches, and other open spaces.

In Objective 2, our findings validate previous studies that have shown consistent associations between summary measures of population health—namely LE—and socioeconomic status, as measured by poverty. In Objective 3, we found additional, less well- established associations that might account for some of the extreme values of LE observed above and beyond the contribution of poverty to explain LE. Our preliminary results suggest that factors that may help communities enjoy longer life expectancy include cleaner air, access to health care facilities, employment, and literacy. Such findings may be useful to policymakers who cannot rectify core social determinants such as household income or education, but are able to institute policies that may buffer their adverse effects. For example, the finding that census tracts with high illiteracy rates were significantly less likely to be a positive outlier could have potential importance for policymakers. Some factors that were also found to predict outlier status, such as rural locations, open space, and the prevalence of single-parent households are less amenable to change by policymakers, however. The next step in the analysis is to delve deeper into assessing those census tracts that remain outliers after accounting for other socioeconomic, environmental, and demographic factors in Objective 3. These remaining census tracts will now undergo quantitative spatial analysis and qualitative inquiry (key informant interviews) to further clarify positive assets that might explain their more favorable outcomes.

These positive assets may influence health directly or indirectly through facilitating or impeding healthy behaviors. Although unhealthy behavior is a matter of personal choice to some degree, the influence of neighborhood and community factors on health behavior and population

health is well-documented [12-14]. Personal health behaviors, such as eating a healthy diet and actively exercising, are intrinsically tied to socioeconomic and social and physical environment factors, such as some of those assessed in this analysis. Still other community assets are not determinative of behaviors.

The debate over individual versus area-level factors and their influence on population health remains unresolved. The findings of this study underscore the potential for community-based assets to affect health, above and beyond the contributions of traditional socioeconomic factors, such as income and poverty. Moreover, some community-level factors are more than simply aggregated individual-level factors and could not be measured on the individual level, particularly environmental and related factors.

Our findings highlight the utility, strengths, and drawbacks of life expectancy methods in these small geographic areas and validate the myriad of past studies on the association between population health and socioeconomic status. In this study we also present a novel way of classifying outliers from the poverty-life expectancy association and assessing the neighborhood and environmental characteristics that may contribute to unexpectedly high or low life expectancy. Assessing variation and ultimately reducing geographic disparities in population health is a major challenge today. Our study examines several potentially modifiable factors that can improve health, save lives, and reduce disparities through evidence-supported policy.

What this study adds:

- A comparison of life expectancy calculation methods for small geographic areas
- A description of some of the strengths and drawbacks to small-area life expectancy calculation

- A novel statistical method for categorizing and describing small areas with unexpectedly high life expectancy and examining predictors of those outliers
- A detailed assessment of small area characteristics that potentially predict life expectancy

Table 1: Summary statistics of life expectancy (LE) from two approaches, 2010

Type of LE calculation	N	mean	SE	min.	median	max.
Observed LE	6435	77.93	3.67	60.82	78.03	98.51
LE from Poisson	6435	80.69	3.29	66.30	80.62	122.81

Table 2: Comparison of life expectancy (LE) using three approaches, 2010

Population (low bound)	N	Mean	SE	Median
951	94	2.1	6.36	2.03
5000	374	2.52	5.37	2.39
7500	915	2.63	4.63	2.58
10000	2477	2.67	4.05	2.43
15000	1762	2.76	3.86	2.37
20,000	801	2.98	3.57	2.69
25,000	320	3.91	4.04	3.61

Table 3: Results of Phase 1 analysis "Best model": Predictors of life expectancy based on poverty and other community-level sociodemographic characteristics for 2010

Variable	Parameter estimate (beta)	Standard error	T-score	P-value
Intercept	75.45816	1.73755	43.43	<.0001
Poverty				
Below 200% of FPL	0.17736	0.01883	9.42	<.0001
Poverty squared	-0.000762	0.000163	-4.68	<.0001
Sex				
Percent female	0.02045	0.01067	1.92	0.0553
Age				
Percent < 5 years old	-0.03055	0.01651	-1.85	0.0643
Percent 5 to 9 years old	0.06367	0.03108	2.05	0.0406
Percent 20 to 24 years old	0.08159	0.01427	5.72	<.0001
Percent 60 to 64 years old	0.06867	0.01934	3.55	0.0004
Percent 65 to 74 years old	0.07746	0.01661	4.66	<.0001
Percent 75 to 84 years old	0.13465	0.01924	7.00	<.0001
Marital status				
Percent divorced	0.13874	0.02434	5.70	<.0001
Percent never married	0.12149	0.01811	6.71	<.0001
Percent married	0.12832	0.01434	8.95	<.0001
Education				
Percent with no high school degree	-0.11769	0.00757	-15.54	<.0001
Percent with high school diploma only	-0.07456	0.01121	-6.65	<.0001
Percent with some college and associates degree	-0.08622	0.00600	-14.37	<.0001
Ethnicity				
Percent Native-born	-0.02546	0.00567	-4.49	<.0001
Percent Hispanic	0.00901	0.00572	1.58	0.1150
Interactions				
poverty*age2	-0.00215	0.000929	-2.32	0.0206
poverty*educ2	-0.00141	0.000358	-3.93	<.0001
poverty*Divorced	-0.00289	0.000629	-4.60	<.0001
poverty*Never_married	-0.00131	0.000293	-4.47	<.0001
poverty*native	-0.00150	0.000176	-8.46	<.0001
poverty*Hispanic	0.00050	0.000139	3.57	0.0004

Table 4: Frequency of positive and negative outliers, 2000 and 2010. Positive outliers are those census tracts that had unexpectedly high LE given high poverty, whereas negative outliers are those census tracts with unexpectedly low life expectancy and low poverty

	2000		2010	
Outlier type	Frequency	Percent	Frequency	Percent
-1	50	0.75	86	1.12
0	6538	98.02	7491	97.87
1	82	1.23	77	1.01

Table 5: Results of marginal weighted logistic regression analysis adjusted for Phase 1 variables (2000)

Variable	Label	Unit	Life expectancy (1)			
			Odds Ratio	90% CI		p-value
Ainsured	percent of adults insured	10.0	1.00	0.62	1.63	0.9868
Allinsured	percent of all age insured	10.0	1.01	0.60	1.71	0.9762
Allrate	maltreatment allegation rate(per 1000 children)	10.0	0.96	0.85	1.09	0.6268
Armed_force	Work in the military	10.0	0.41	0.05	1.71	0.4338
BikWaL	Walk or bike to work	10.0	1.85	1.19	2.69	0.0114
Cancer	Cancer risk due to air pollutants (per million)	10.0	0.76	0.68	0.84	<.0001
Employed	work but not in the military	10.0	0.65	0.42	1.08	0.1335
Entrrate	enter foster care rate(per 1000 children)	10.0	0.06	0.02	0.25	0.0010
FFR	Number of Fast Food Restaurants	10.0	1.22	0.69	1.96	0.5414
Fami11	Two-parent family householders	10.0	1.23	0.69	2.16	0.5531
Fami12	single-parent family householders	10.0	0.63	0.25	1.42	0.3884
Fami13	no children family householders	10.0	0.59	0.35	0.99	0.0874
Fami14	non-family householders	10.0	1.41	0.93	2.15	0.1838
Fami21	Two-parent families	10.0	1.71	0.94	3.04	0.1340
Fami22	single-parent families	10.0	0.87	0.44	1.66	0.7335
Fami23	no children families	10.0	0.63	0.35	1.17	0.2104
Fami31	Single mother families	10.0	0.96	0.62	1.49	0.8950
Fami32	Single father or two-parent families	10.0	1.04	0.67	1.63	0.8935
GINI		0.1	2.82	1.47	5.41	0.0088
HFP	% health food	10.0	1.00	0.83	1.23	0.9999
Immun	Percent of kindergarteners with all immunizations	0.5	0.93	0.03	32.8	0.9712
Kinsured	percent of under age 18 insured	10.0	1.01	0.56	1.84	0.9800
MHPR	Mental public health provider ratio	10.0	0.92	0.84	1.01	0.1565

Neuro	Neurological risk due to air pollutants (HQ)	1.0	0.00	0.00	0.02	<.0001
PCPR	PCP rate	10.0	0.92	0.84	1.01	0.1483
PM25_00	PM2.5 in 2000	10.0	1.02	0.86	1.19	0.8767
PUnheal	Percent of at least unhealthy days	10.0	0.99	0.88	1.11	0.8570
Quart1	correctional institutions	10.0	0.58	0.07	1.44	0.4872
Quart2	nursing homes	10.0	0.16	0.01	1.94	0.3065
Quart3	college dormitories including college quarters off campus	10.0	2.07	0.65	4.89	0.2001
Quart5	other institutional and no institutional group quarters	10.0	2.21	1.10	4.24	0.0510
Quart6		10.0	0.72	0.45	1.21	0.2645
QuartP	either group quarters present or absent	1.0	0.96	0.63	1.49	0.8701
RFR	% recreation facility rate	10.0	1.40	0.79	2.39	0.3185
RSega	the ratio between White and nonwhite	1.0	1.02	0.95	1.09	0.5731
Race1	White	10.0	1.01	0.85	1.21	0.9295
Race2	Black	10.0	0.69	0.44	0.97	0.1158
Race3	American Indian, Eskimo, or Aleut	10.0	1.82	0.53	4.08	0.2375
Race4	Asian and Pacific Islander	10.0	1.19	0.92	1.56	0.2704
Race5	Other	10.0	1.22	0.64	2.34	0.6195
RentMort	burden and overburden for Rent or mortgage	10.0	0.97	0.69	1.36	0.8696
RentMort1	burden and overburden for Rent or mortgage excluding the no computed units	10.0	0.89	0.63	1.26	0.5776
Respir	Respiratory risk due to air pollutants (HQ)	1.0	0.93	0.90	0.97	0.0011
RoomO1	% of households in overcrowded (>= 1.01 persons/room)	10.0	0.45	0.19	1.00	0.1084
RoomO2	severely overcrowded (>= 1.5 persons per room)	10.0	1.11	0.65	1.87	0.7529
Rural		10.0	1.23	1.14	1.33	<.0001
Sexratio	Sex ratio at birth	0.1	1.63	1.04	2.49	0.0676

Subsrate	substantiation rate(per 1000 children)	5.0	0.67	0.53	0.84	0.0039
TravT1	percent of population(not work at home)to travel less than 30 minutes	10.0	1.02	0.87	1.20	0.8277
TravT2	percent of population(not work at home)to travel 30 to 60 minutes	10.0	0.75	0.60	0.93	0.0300
TravT3	percent of population(not work at home)to travel more than 60 minutes	10.0	1.50	1.15	1.94	0.0102
Unemployed		10.0	1.95	1.12	3.28	0.0399
Urban		10.0	0.81	0.75	0.88	<.0001
illiterate	percent of illiterate population	10.0	0.71	0.54	0.93	0.0395

Table 6: Results of marginal weighted logistic regression adjusted for Phase 1 variables (2010)

Variable	Label		LE(1)			
		Unit	Odds ratio	L	U	Prob
Ainsured	percent of adults (19-64) insured	10.0	0.97	0.72	1.30	0.8470
Allinsured	percent of all age(under 65) insured	10.0	0.94	0.64	1.38	0.7866
Allrate	maltreatment allegation rate (per 1000 children)	1.0	0.99	0.98	1.00	0.1418
Armed_force		10.0	1.26	0.40	2.56	0.6852
BikWaL	Walk or bike to work	10.0	1.38	0.94	1.94	0.1437
Cinsured	Percent of children insured	10.0	0.00	0.00	1E25	0.6616
Diesel_PM	Diesel PM emissions from on-road and non-road sources	10.0	0.80	0.67	0.95	0.0358
Employed	Percent employed	10.0	1.20	0.77	1.92	0.5038
Entrrate	enter foster care rate(per 1000 children)	10.0	0.43	0.12	1.54	0.2880
Fami11	Two-parent family householders	5.0	1.25	1.06	1.47	0.0280
Fami12	single-parent family householders	10.0	0.16	0.08	0.32	<.0001
Fami13	no children family householders	10.0	1.26	0.86	1.83	0.3246
Fami14	non-family householders	10.0	0.91	0.71	1.15	0.4976
Fami21	Two-parent families	10.0	2.08	1.44	3.00	0.0010
Fami22	single-parent families	10.0	0.20	0.11	0.33	<.0001
Fami23	no children families	10.0	1.12	0.78	1.58	0.6110
Fami31	Single mother families	10.0	0.41	0.29	0.59	<.0001
Fami32	Single father or two-parent families	10.0	2.42	1.69	3.48	<.0001
Immun	Percent of kindergarteners with all immunizations	0.1	1.00	0.99	1.00	0.4793

Kinsured	percent of under age 18 insured	10.0	0.46	0.18	1.13	0.1584
PM25	annual mean PM 2.5 concentrations	5.0	0.54	0.39	0.75	0.0019
PM25_10	PM2.5 in 2010	2.0	0.93	0.84	1.02	0.2496
PUnheal	Percent of at least unhealthy days	1.0	0.99	0.98	1.01	0.4103
Pesticides	Total pounds of selected active pesticide ingredients used in production-agriculture	10.0	1.00	1.00	1.00	0.0137
Quart2	nursing homes	0.1	0.99	0.97	1.01	0.3862
Quart3	college dormitories including college quarters off campus	0.5	1.04	1.01	1.07	0.0059
Quart5	other institutional and non-institutional group quarters	10.0	0.61	0.11	2.03	0.5880
QuartP	either group quarters present or absent	1.0	1.22	0.82	1.86	0.4235
RSega	the ratio between White and nonwhite	1.0	0.99	0.93	1.03	0.7509
RentMort	burden and overburden for Rent or mortgage	10.0	1.17	0.95	1.43	0.2075
RoomO1	% of households in overcrowded (>= 1.01 persons/room)	10.0	1.27	0.75	2.10	0.4493
RoomO2	severely overcrowded (>= 1.5 persons per room)	0.5	1.00	0.97	1.03	0.8840
Rural		10.0	0.98	0.86	1.09	0.7616
Sexratio	Sex ratio at birth	1.0	0.28	0.00	117	0.7505
Subsrate	substantiation rate(per 1000 children)	10.0	0.96	0.62	1.46	0.8749
TravT1	percent of population(not work at home)to travel less than 30 minutes	10.0	1.06	0.93	1.21	0.4402
TravT2	percent of population(not work at home)to travel 30 to 60 minutes	10.0	0.86	0.73	1.03	0.1630
TravT3	percent of population(not work at home)to travel more than 60 minutes	1.0	1.01	0.98	1.03	0.5751
Unemployed	Percent unemployed	10.0	0.77	0.47	1.24	0.3704

Urban		10.0	1.02	0.92	1.16	0.7616
bed3	Presence of beds for inpatient mental health (versus absent)	1.0	3.03	0.71	8.56	0.1280
bed5	Presence of beds for long-term care facilities either (versus absent)	1.0	0.90	0.51	1.48	0.7308
bed7	Presence of beds for acute care hospitals (versus absent)	1.0	2.60	1.48	4.31	0.0031
cancer	Cancer risk due to air pollutants (per million)	0.1	0.00	0.00	2E30	0.1483
cost1	child care center infant cost	10.0	1.00	1.00	1.00	0.3001
cost2	child care center preschooler cost	10.0	1.00	1.00	1.00	0.2877
cost3	family child care home infant cost	0.1	1.00	1.00	1.00	0.2292
cost4	family child care home preschooler cost	10.0	1.00	1.00	1.00	0.3154
develop	area is developed or not	1.0	1.52	0.96	2.51	0.1500
enrol1	Percent of age group enrolled in school	10.0	1.00	0.94	1.07	0.9906
enrol2	Percent of age group enrolled in public school	1.0	1.00	0.99	1.01	0.8666
enrol3	Percent of age group enrolled in private school	10.0	1.01	0.95	1.06	0.8659
facip1	specialty clinics either present or absent	1.0	2.31	1.32	3.78	0.0084
facip3	inpatient mental health either present or absent	1.0	3.03	0.71	8.56	0.1280
facip4	community/free clinic either present or absent	1.0	1.08	0.60	1.83	0.8103
facip5	long-term care facilities either present or absent	1.0	0.90	0.51	1.48	0.7308
facip6	home health/hospice either present or absent	1.0	0.81	0.46	1.33	0.5108
facip7	acute care hospitals either present or absent	1.0	2.60	1.48	4.31	0.0031
farm	area is cultivated or not	1.0	0.36	0.10	0.92	0.1150

gini	Gini coefficient	10.0	4E26	2E12	2E40	0.0018
lead1	0-5 years with elevated blood lead level	1.0	26E6	0.00	9E48	0.7820
lead2	6-20 years with elevated blood lead level	1.0	0.00	0.00	7E6	0.4029
lead3	all children under 20 with elevated blood level	1.0	17E3	0.00	1E45	0.8741
neuro	Neurological risk due to air pollutants (HI)	0.1	0.86	0.50	1.38	0.6253
openspace	area is open space versus not	1.0	2.52	1.45	4.12	0.0034
park	Percent of population within 1/2 mile of park, beach, open space, or coastline	10.0	0.95	0.90	1.00	0.0825
pollution_burd	Average of percentiles from the pollution burden indicators	10.0	0.95	0.83	1.09	0.5138
race1	White	10.0	1.00	0.89	1.13	0.700
race2	Black	10.0	0.41	0.25	0.64	0.0018
race3	American Indian, Eskimo, or Aleut	10.0	1.15	0.27	2.84	0.8543
race4	Asian and Pacific Islander	10.0	1.22	1.00	1.49	0.0998
race5	Other	10.0	1.10	0.88	1.38	0.4816
respir	Respiratory risk due to air pollutants (HI)	10.0	0.86	0.28	2.47	0.8260
snap	percent of household receiving snap	10.0	0.58	0.34	0.96	0.0856
traffic	Traffic density in vehicle kilometers per hour per road length within 150 meters	10.0	1.00	1.00	1.00	0.5889

Table 7: Phase-II Results from "best" multiple logistic regression model controlling for phase-I variables (2000)

Effect	Label	Unit	LE(1)			
			Odds Ratio	90% CI		p-value
Allrate	maltreatment allegation rate (per 1000 children)	10.0	1.39	1.11	1.74	0.02
BikWaL	Walk or bike to work	10.0	1.15	0.62	2.15	0.71
Cancer	Cancer risk due to air pollutants (per million)	10.0	0.65	0.51	0.84	<0.001
Employed	work but not in the military	1.0	1.07	0.89	1.29	0.54
Entrrate	enter foster care rate(per 1000 children)	1.0	0.63	0.51	0.78	<0.001
Fami13	no children family householders	10.0	0.96	0.19	4.93	0.97
Fami14	non-family householders	10.0	0.93	0.34	2.53	0.90
Fami21	Two-parent families	10.0	0.59	0.24	1.44	0.33
Fami23	no children families	10.0	0.33	0.08	1.25	0.17
GINI		0.1	1.26	0.62	2.56	0.60
Immun	Percent of kindergarteners with all immunizations	0.1	0.83	0.56	1.21	0.41
MHPR	Mental public health provider ratio	10.0	0.87	0.76	0.99	0.08
Neuro	Neurological risk due to air pollutants (HQ)	0.1	1.05	0.71	1.54	0.84
PCPR	PCP rate	10.0	1.13	0.96	1.32	0.21
Quart1	correctional institutions	1.0	0.96	0.80	1.15	0.69
Quart2	nursing homes	1.0	0.77	0.57	1.06	0.18
Quart5	other institutional and non-institutional group quarters	1.0	1.04	0.91	1.19	0.64
Quart6		1.0	0.96	0.85	1.09	0.61
QuartP	either group quarters present or absent	1.0	0.91	0.59	1.41	0.74
RFR	% recreation facility rate	10.0	0.36	0.12	1.04	0.11
RSega	the ratio between White and nonwhite	1.0	0.92	0.83	1.01	0.15
Race2	Black	10.0	0.64	0.40	1.03	0.12
Respir	Respiratory risk due to air pollutants (HQ)	1.0	1.08	1.02	1.15	0.04
RoomO1	% of households in overcrowded (>= 1.01 persons/room)	10.0	0.58	0.24	1.37	0.30

RoomO2	severely overcrowded (>= 1.5 persons per room)	10.0	1.09	0.64	1.87	0.79
Rural		10.0	1.19	1.08	1.31	0.00
Sexratio	Sex ratio at birth	0.0	1.04	1.00	1.08	0.09
Subsrate	substantiation rate(per 1000 children)	5.0	0.84	0.60	1.17	0.39
TravT1	percent of population(not work at home)to travel less than 30 minutes	10.0	0.65	0.48	0.86	0.01
TravT2	percent of population(not work at home)to travel 30 to 60 minutes	10.0	0.52	0.35	0.77	0.01
Unemployed	Percent unemployed	1.0	1.12	0.92	1.37	0.34
illiterate	Percent illiterate	1.0	0.98	0.94	1.02	0.43

Table 8: Phase-II Results from "best" multiple logistic regression model controlling for phase-I variables (2010)

Variable	Label	Unit	LE(1)			
			Odds ratio	l	u	Prob
Diesel_PM	Diesel PM emissions from on-road and non-road sources	1.0	0.99	0.97	1.01	0.53
Fami11	Two-parent family householders	1.0	0.80	0.70	0.91	0.006
Fami12	single-parent family householders	1.0	1.55	1.12	2.14	0.027
Fami21	Two-parent families	1.0	1.32	1.16	1.51	0.0004
Fami22	single-parent families	1.0	0.52	0.39	0.71	0.0004
Fami32	Single father or two-parent families	1.0	0.83	0.75	0.92	0.0032
Kinsured	percent of under age 18 insured	1.0	0.84	0.74	0.96	0.03
PM25	annual mean PM 2.5 concentrations	1.0	0.88	0.80	0.97	0.02
Pesticides	Total pounds of selected active pesticide ingredients used in production-agriculture	1.0	1.00	1.00	1.00	0.07
Quart3	college dormitories including college quarters off campus	1.0	1.03	0.95	1.10	0.55
bed7	beds for acute care hospitals either present or absent	1.0	2.11	1.08	4.12	0.07
facip1	specialty clinics either present or absent	1.0	2.06	1.07	3.96	0.07
facip3	inpatient mental health either present or absent	1.0	2.55	0.62	10.5	0.28
farm	area is cultivated or not	1.0	0.30	0.09	1.02	0.11
gini		0.1	1.67	1.12	2.49	0.03
openspace	area is open space or not	1.0	2.55	1.39	4.66	0.01
park	Percent of population within 1/2 mile of park, beach, open space, or coastline	1.0	1.00	0.99	1.00	0.34
race2	Black	1.0	0.95	0.90	1.00	0.11
race4	Asian and Pacific Islander	1.0	1.03	1.01	1.06	0.03
snap	percent of household receiving snap	1.0	0.99	0.93	1.05	0.77

Figure 1: Census tract life expectancy in California, 2010

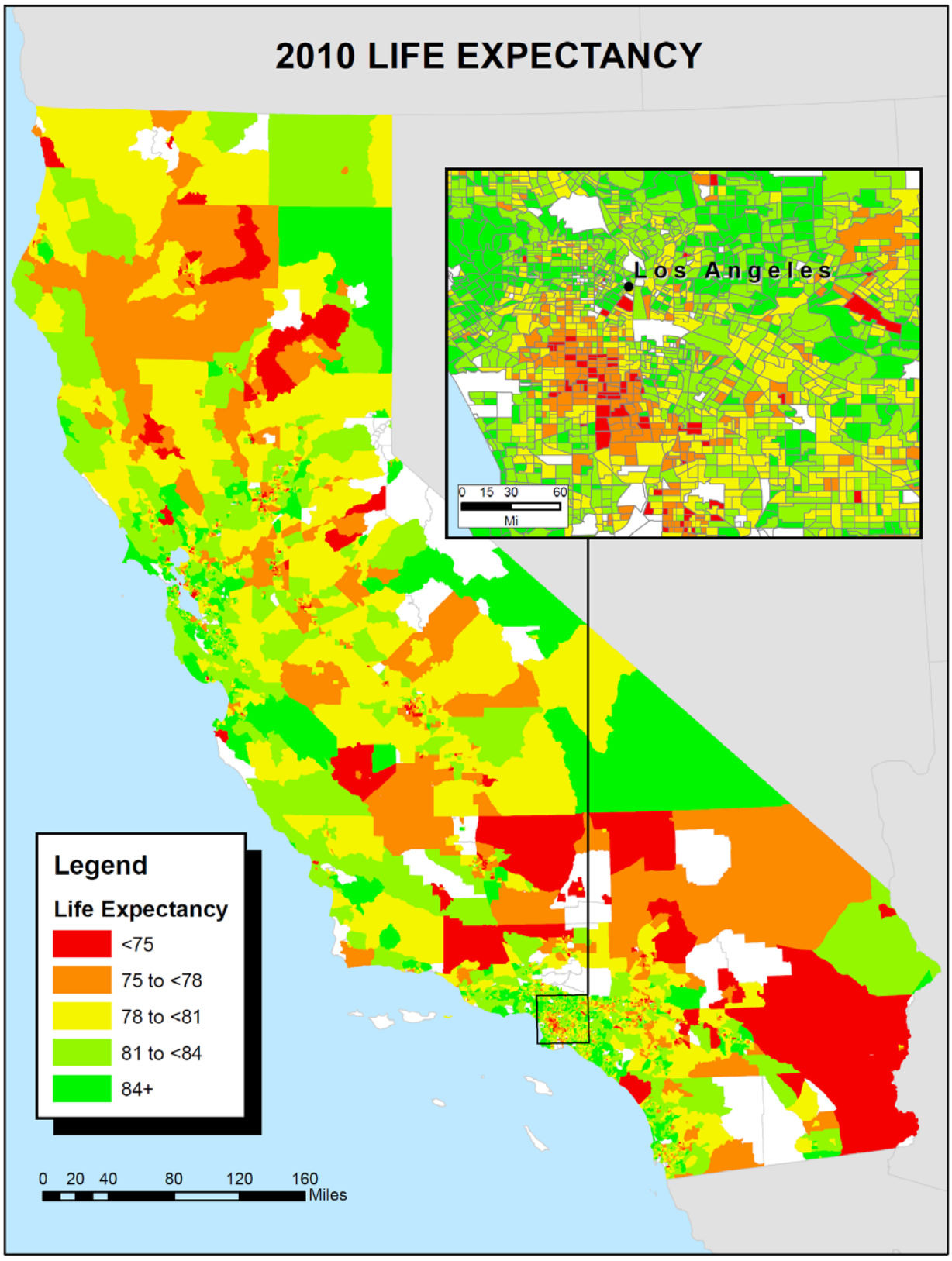
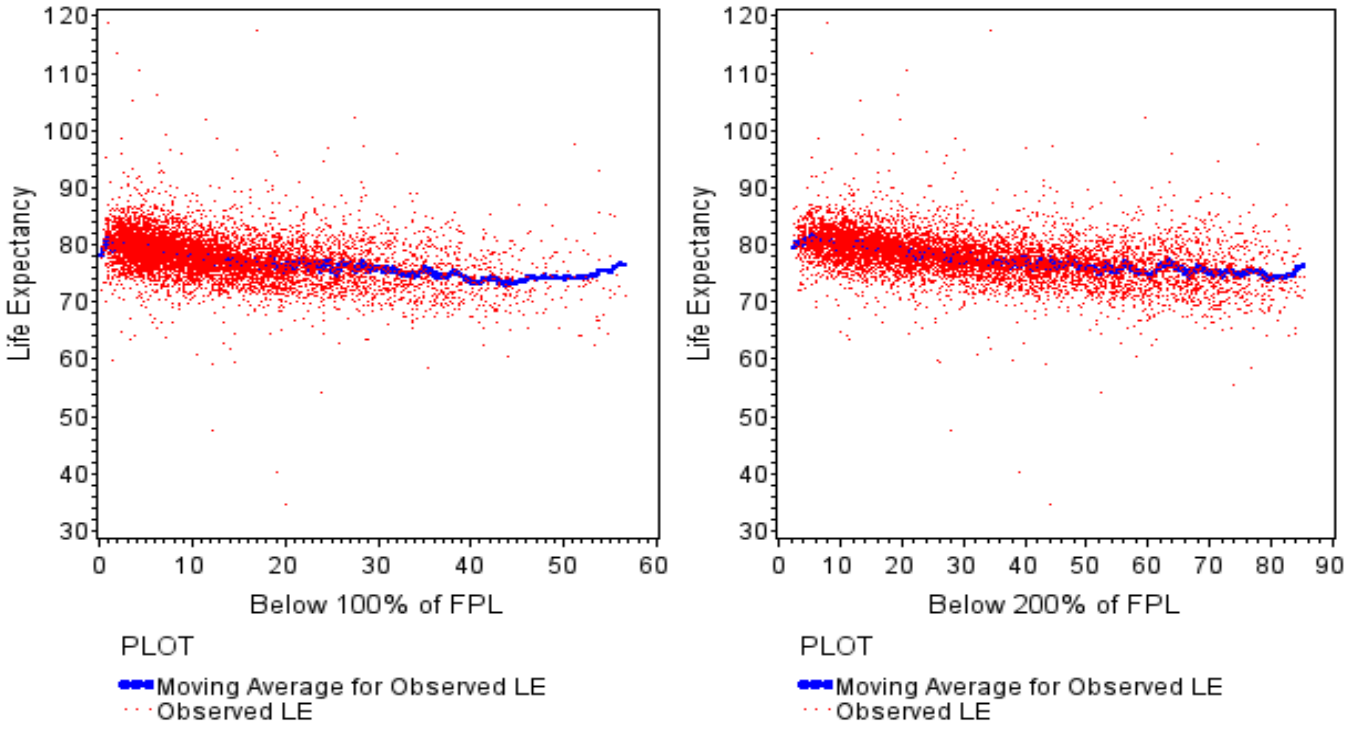
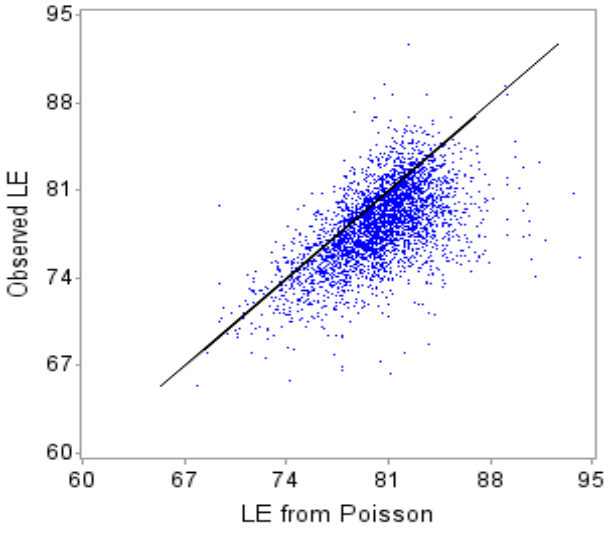


Figure 2: Life expectancy and percent poverty by census tract: Percent of households below 100% of poverty line (Panel A) and Percent of households below 200% of poverty line (Panel B) (2010)

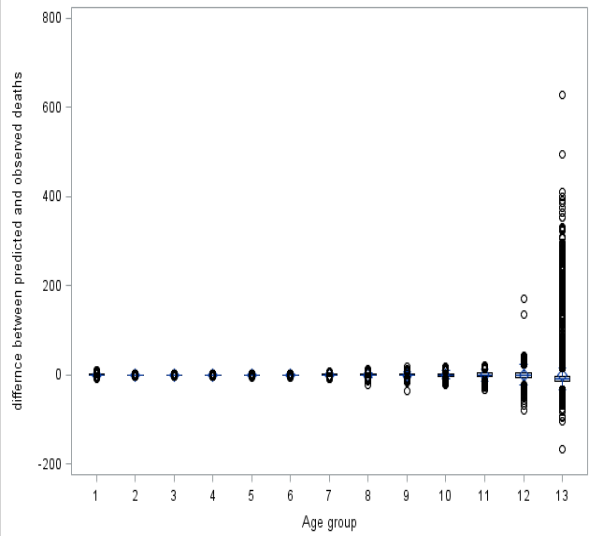


Panel A **Panel B**

Figure 3: Observed life expectancy and life expectancy from Poisson modeling. Scatterplot of life expectancy (Panel A) and difference of observed and predicted life expectancy from Poisson Models (Panel B)

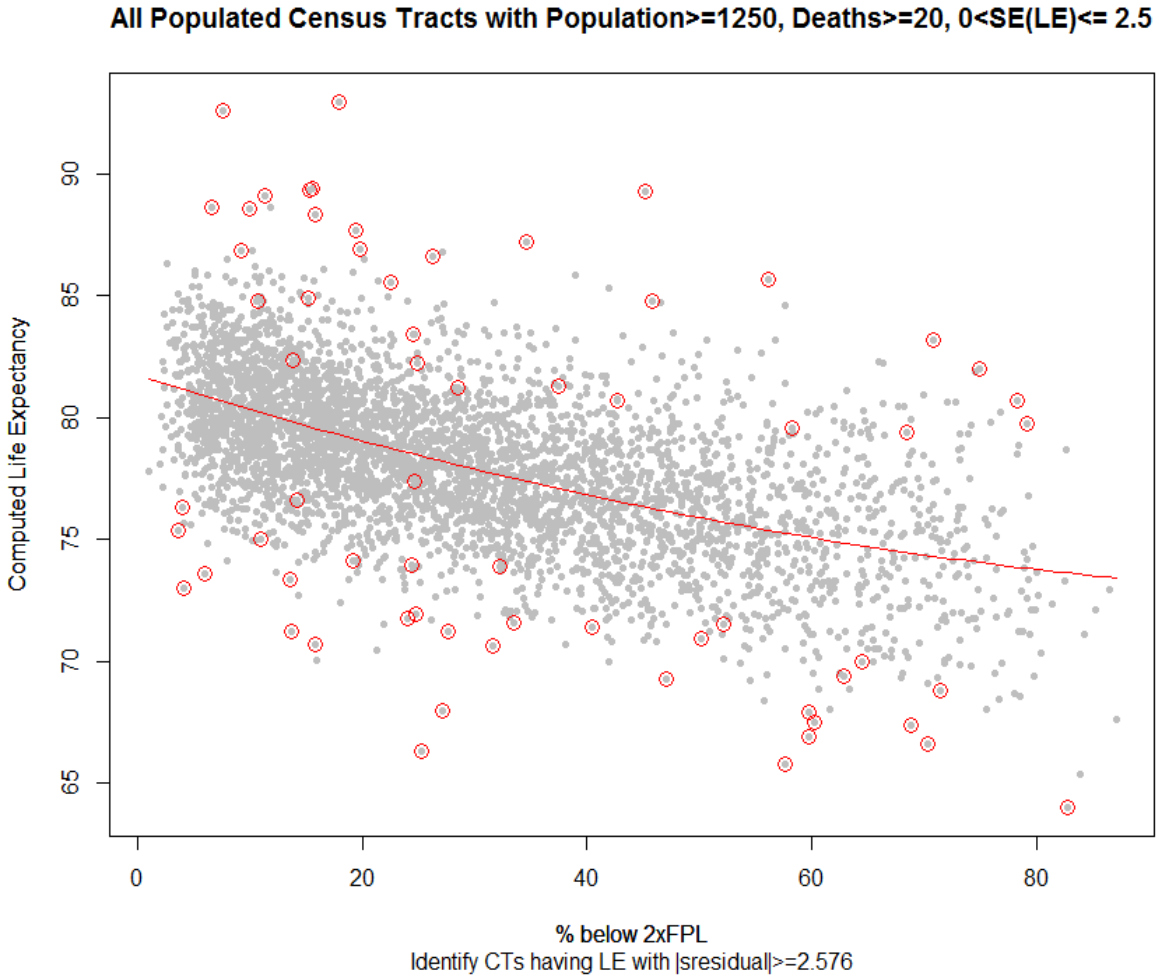


Panel A



Panel B

Figure 4: Association between life expectancy and percent below 2x poverty level for selected California census tracts meeting inclusion criteria. Census tracts surrounded by red circles were those considered "outliers" in the analysis.



References

1. Geronimus AT, Bound J, Waidmann TA, Colen CG, Steffick D. Inequality in life expectancy, functional status, and active life expectancy across selected black and white populations in the United States. Demography. 2001; 29(2):227-51.

2. Auchincloss AH, Hadden W. The health effects of rural-urban residence and concentrated poverty. Rural Health. 2002; 18(2):319-36.

3. Eayres D, Williams ES. Evaluation of methodologies for small area life expectancy estimation. J Epidemiol Community Health. 2004 Mar; 58(3):243-49.

4. Bravo JM, Malta J. Estimating life expectancy in small population areas. UN Statistical Commission and Economic Commission for Europe: Conference of European Statisticians, Lisbon, Portugal, April 2010.

5. Toson B, Baker A. Life expectancy at birth: methodological options for small populations. National Statistics Methodological Series No. 33, 2005.

6. Clarke CA, Miller T, Chang ET, Yin D, Cockburn M, Gomez SL. Racial and social class gradients in life expectancy in contemporary California. Soc Sci Med. 2010;70(9):1373-80.

7. Gomez SL, Glaser SL, McClure LA, Shema SJ, Kealey M, Keegan TH, Satariano WA. The California Neighborhoods Data System: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. Cancer Causes Control. 2011; 22(4):631-47.

8. Diez Roux AV. Investigating neighborhood and area effects on health. Am J Public Health. 2001Nov; 91(11):1783-89.

9. Chiang. CL The Life Table and Its Applications. Malabar, Florida: Robert E. Krieger Publishing Company, 1984.

10. Lee RD, Carter LR. Modeling and forecasting US mortality. J Am Stat Assoc. 1992; 87(419): 659-71.

11. Sandeep CK, Alison LR, Majid E, Christopher JL M. Falling behind: life expectancy in US counties from 2000 to 2007 in an international context. Population Health Metrics 2011, 9:16.

12. Woolf SH, Dekker MM, Byrne FR, Miller WD. Citizen-centered health promotion: building collaborations to facilitate healthy living. Am J Prev Med. 2011; 40(1 Suppl 1):S38–47.

13. Braveman P, Egerter S, Williams D. Social determinants of health: coming of age. Annu Rev Public Health. 2011; 32: 381–98.

14. Adler NE and Rehkopf DH. U.S. disparities in health: descriptions, causes, and mechanisms. Annu Rev Public Health. 2008; 29: 235–52.

Acknowledgements

We thank Tony Iton and Jim Keddy The California Endowment for funding this project, Matt Beyers of the Alameda County Public Health Department for his assistance in obtaining death record data, and Dr. Ivan Suen of the Virginia Commonwealth University Wilder School of Government for his spatial analysis expertise. We also thank Dr. Derek Chapman for his guidance with this research project.

This study was conducted by the Virginia Commonwealth University Center on Society and Health.

APPENDIX A: Variable lists

Table A1: Data Sources (2000)

Phase	Variable	Data Source	Geographic Scale	Year
1	Deaths	California Department of Public Health	Census tract	1999-2001
1	% of population aged >= 25 years by educational attainment	American Community Survey	Census tract	City (1, 3 years); Census tract (5 years)
1	Gender	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2000
1	Age Group Distribution	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2000
1	Income-to-Poverty Ratio	U.S. Census Bureau,American Community Survey:American FactFinder: st00007.uf3	Census tract	2000
1	Marital status	U.S. Census Bureau,American Community Survey:American FactFinder: Table DP-2	Census tract	2000
1	Immigration status	American Community Survey	Census tract	2000
1	Hispanic Origin	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2000
2	Race	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2000
2	Percent of unhealthy days	http://www.epa.gov/airdata/ad_rep_aqi.html	County	2000
2	% of children less than 18 reported with neglect, physical, or sexual abuse	Child Welfare Dynamic Report System . University of California, Berkeley and Dept. of Social Services.	Census Tract	2000
2	Group Quarters	American Community Survey	Census Tract	2000
2	Cancer risk due to air pollutants (per million)	National Air Toxic Assessment (http://www.epa.gov/ttn/atw/nata1999/tables.html)	Census Tract	1999
2	Respiratory risk due to air pollutants (HI)	National Air Toxic Assessment (http://www.epa.gov/ttn/atw/nata1999/tables.html)	Census Tract	1999
2	Neurological risk due to air pollutants (HI)	National Air Toxic Assessment	Census Tract	1999
2	Time traveled to work	American Community Survey	Census Tract	2000
2	Rural vs Urban Location	American Community Survey	Census Tract	2000
2	percent of residents that walk or bike to work	American Community Survey. Detailed data sets, annual, 3-year, 5-year	Census tract	2000
2	Access to healthy food	MATCH 2011	County	2008
2	Retail food environment index (e.g. number of fast-food restaurants and convenience stores/total number of supermarkets and produce vendors	California Board of Equalization	Census tract	2011
2	insured Population	American Community Survey	County	2000
2	Insured children under 20	California Healthy Kids Survey	County	2001
2	Immunization status	California Healthy Kids Survey	County	2001

2	Mental public health provider ratio	MATCH 2011	County	2008
2	Primary care providers rate	MATCH 2011	County	2006
2	household type	American Community Survey. Detailed data sets, annual, 3-year, 5-year and California Department of Housing and Community Development. Building Blocks for Effective Housing Elements, housing needs	Census tract	2000
2	Percent of Household income spent on rent or mortgage using benchmarks of >30% (burdened) and >50% severely burdened	American Community Survey. Detailed data sets, annual, 3-year, 5-year and California Department of Housing and Community Development. Building Blocks for Effective Housing Elements. Housing needs	Census tract	2000
2	Income Inequality: Gini coefficient describing the amount of total annual community income generated by the number of households	American Community Survey. Detailed data sets, annual, 3-year, 5-year. U.S Bureau of Census.	Census tract	2000
2	employment status	American Community Survey. Detailed data sets, annual, 3-year, 5-year. U.S Bureau of Census. AND 29. Monthly Labor Force Data for Cities and Census Designated Places (CDP). Employment Development Department.	Census tract	2000
2	% of households in overcrowded (>= 1.01 persons/room) and severely overcrowded (>= 1.5 persons per room) conditions	American Community Survey. Detailed data sets, annual, 3-year, 5-year: AND California Department of Housing and Community Development. Building Blocks for Effective Housing Elements. Housing needs	Census tract	2000
2	recreation facility rate	MATCH 2011	County	2008
2	Sex ratio at birth	California Healthy Kids Survey	County	2000
3	Mentally unhealthy days	MATCH 2011	County	2003-2009
3	Diabetic	MATCH 2011	County	2003-2006
3	Adults obesity	MATCH 2011	County	2006-2008
3	Smokers	MATCH 2011	County	2002-2008
3	physical unhealthy days	MATCH 2011	County	2002-2008
3	physical inactivity	MATCH 2011	County	
3	Teen birth rate	MATCH 2011	County	2000-2006
3	Percent of population located <1/2 mile of a regional bus/rail/ferry and <1/4 mile local bus /light rail	University of California Davis (ULTRANS), University of California Berkeley (SafeTREC). Transit asset inventories (Link 1 and link 2).	Census Tract	Annual

3	% of households/populations near busy roadways	American Community Survey. Detailed data sets, annual, 3-year, 5-year. U.S Bureau of Census. Federal Highway Administration. FHWA Functional Classification Guidelines. Department of Transportation; 2008 in conjunction with public or commercial GIS files of roadways.	Census tract	Annual
3	Hospitalization rate for ambulatory care	MATCH 2011	County	2005-2006
3	HIV rate	MATCH 2011	County	
3	STI rate	MATCH 2011	County	

Table A2: Data Sources (2010)

Phase	variable	Data Source	Geographic Scale	Years/Frequency of Update
1	Deaths	California Department of Public Health	Census tract	
1	% of population aged >= 25 years by educational attainment	American Community Survey. Detailed data sets, annual, 3-year, 5-year. U.S Bureau of Census, and Condition of education. U.S. Department of Education National Center for Educational Statistics.	Census tract	2006-2010
1	Gender	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2006-2010
1	Age Group Distribution	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2006-2010
1	Income-to-Poverty Ratio	U.S. Census Bureau,American Community Survey:American FactFinder: st00007.uf3	Census tract	2006-2010
1	Marital status	U.S. Census Bureau,American Community Survey:American FactFinder: Table DP-2	Census tract	2006-2010
1	immigrant status	American Community Survey	Census tract	2006-2010
1	Hispanic Origin	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2006-2010
2	Race	U.S. Census Bureau,American Community Survey:American FactFinder Table DP-1	Census tract	2006-2010
2	percent of tree canopy coverage (urban areas)	National Land Cover Database. USGS, EPA, NOAA, DOI, NASA, USFS, US Park Service (Multi-Resource Land Consortium) and other public and commercial satellite imaging projects.	Census tract	2010
2	Percent of unhealthy days	http://www.epa.gov/airdata/ad_rep_aqi.html	County/ Air basin	2010
2	% of children less than 18 reported with neglect, physical, or sexual abuse	Child Welfare Dynamic Report System . University of California, Berkeley and Dept. of Social Services.	Census Tract	2010
2	Group Quarters	American Community Survey	Census Tract	2010
2	Cancer risk due to air pollutants (per million)	National Air Toxic Assessment (http://www.epa.gov/ttn/atw/nata1999/tables.html)	Census Tract	2005
2	Respiratory risk due to air pollutants (HQ)	National Air Toxic Assessment (http://www.epa.gov/ttn/atw/nata1999/tables.html)	Census Tract	2005
2	Neurological risk due to air pollutants (HQ)	National Air Toxic Assessment	Census Tract	2005
2	Time traveled to work	American Community Survey	Census Tract	2006-2010
2	Rural vs Urban Location	American Community Survey	Census Tract	2006-2010
2	percent of residents that walk or bike to work	American Community Survey. Detailed data sets, annual, 3-year, 5-year	Census tract	2006-2010
2	Access to healthy food	MATCH 2013	county	2006
2	Retail food environment index (eg number of fast-food restaurants and convenience stores/total number of supermarkets and produce vendors	California Board of Equalization	Census tract	2011
2	Uninsured Population	American Community Survey	County	2010
2	Insured children under 20	California Healthy Kids Survey	County	2009
2	Immunization status	California Healthy Kids Survey	County	2010
2	Mental public health provider ratio	MATCH 2013	county	2011-2012
2	Primary care providers rate	MATCH 2013	county	2006
2	household type	American Community Survey. Detailed data sets, annual, 3-year, 5-year and California Department of Housing and Community Development. Building Blocks for Effective Housing Elements, housing needs	Census tract	2006-2010
2	Percent of Household	American Community Survey. Detailed data sets,	Census tract	2006-2010

	income spent on rent or mortgage using benchmarks of >30% (burdened) and >50% severely burdened	annual, 3-year, 5-year and California Department of Housing and Community Development. Building Blocks for Effective Housing Elements. Housing needs		
2	Income Inequality: Gini coefficient describing the amount of total annual community income generated by the number of households	American Community Survey. Detailed data sets, annual, 3-year, 5-year. U.S Bureau of Census.	Census tract	2006-2010
2	employment status	American Community Survey. Detailed data sets, annual, 3-year, 5-year. U.S Bureau of Census.	census tract	2006-2010
2	% of households in overcrowded (>= 1.01 persons/room) and severely overcrowded (>= 1.5 persons per room) conditions	American Community Survey. Detailed data sets, annual, 3-year, 5-year: AND California Department of Housing and Community Development. Building Blocks for Effective Housing Elements. Housing needs	Census tract	2006-2010
2	recreation facility rate	MATCH 2013	County	2010
2	Sex ratio at birth	California Healthy Kids Survey	County	2010
2	child blood lead level	California Healthy Kids Survey	county	2010
2	child care cost	California Healthy Kids Survey	county	2010
2	without emotional or social support	MATCH 2013	county	2005-2010
2	could not see doctor due to cost	MATCH 2013	county	2005-2011
2	PM 25		county	2010
2	Number of free/community clinics	California Office of Statewide Health Planning and Development: Healthcare Atlas	Census Tract	2012
2	number of specialty clinics	California Office of Statewide Health Planning and Development: Healthcare Atlas	Census Tract	2012
2	number of hospice/home care services	California Office of Statewide Health Planning and Development: Healthcare Atlas	Census Tract	2012
2	Number of acute care hospitals	California Office of Statewide Health Planning and Development: Healthcare Atlas	Census Tract	2012
2	Number of in-patient psychiatric services	California Office of Statewide Health Planning and Development: Healthcare Atlas	Census Tract	2012
2	Number of out-patient psychiatric services	California Office of Statewide Health Planning and Development: Healthcare Atlas	Census Tract	2012
2	Number of nursing homes	California Office of Statewide Health Planning and Development: Healthcare Atlas	Census Tract	2012
2	PM 2.5 (ZCTA)	CalEnviroScreen	ZCTA	2009-2010
2	Diesel PM	CalEnviroScreen	ZCTA	2009-2010
2	Pesticides	CalEnviroScreen	ZCTA	2009-2010
2	Traffic Density	CalEnviroScreen	ZCTA	2009-2010
2	Pollution Burden	CalEnviroScreen	ZCTA	2009-2010
2	% population within 1/2 mile of park, beach or open space	CalEnviroScreen	ZCTA	2009-2010
3	Mentally unhealthy days	MATCH 2013	county	2005-2011
3	Diabetic	MATCH 2013	county	2010
3	Adults obesity	MATCH 2013	county	2009
3	Smokers	MATCH 2013	county	2005-2011
3	physical unhealthy days	MATCH 2013	County	2005-2011
3	physical inactivity	MATCH 2013	County	2009
3	Teen birth rate	MATCH 2013	County	2004-2010
3	Percent of population located <1/2 mile of a regional bus/rail/ferry	University of California Davis (ULTRANS), University of California Berkeley (SafeTREC). Transit asset inventories (Link 1 and link 2).	Census Tract	Annual

	and <1/4 mile local bus /light rail			
3	% of households/populations near busy roadways	American Community Survey. Detailed data sets, annual, 3-year, 5-year. U.S Bureau of Census. Federal Highway Administration. FHWA Functional Classification Guidelines. Department of Transportation; 2008 in conjunction with public or commercial GIS files of roadways.	Census tract	Annual
3	Hospitalization rate for ambulatory care	MATCH 2013	county	2005-2006
3	HbA1c	MATCH 2013	County	2009
3	STI rate	MATCH 2013	County	2010
2	SNAP	American Community Survey	Census tract	2007-2011
2	School Enrollment	American Community Survey	Census tract	2007-2011
2	Transportation to work	California Department of Public Health	Census tract	2006-2010