Krieger N, Waterman PD, Chen JT, Soobader M-J, Subramanian SV, Rehkopf D. The Public Health Disparities Geocoding Project: improving monitoring of social inequalities in health in the United States.

Presented by N. Krieger to the Subcommittee on Populations—National Committee on Vital and Health Statistics, Philadelphia, PA, November 8, 2002.

Research funded by: NIH/NICHD (1 R01 HD 3685-01)

[SLIDE 1]

Good afternoon. I'm delighted to present, on behalf of my co-authors and our Public Health Disparities Geocoding Project, our talk on improving monitoring social inequalities in health in the United States.

[SLIDE 2]

I'll begin by acknowledging our study partners, at the Massachusetts Department of Public Health and the Rhode Island of Department of Health, as listed on this slide.

[SLIDE 3]

One premise of our project is that collaborations between universities and health departments is vital to improving monitoring of social inequalities in health. Health departments are aware that this research is needed, yet typically lack the time and resources to do the work. University-based researchers, in turn, are well-placed to get the grants and have the expertise to conduct the needed research, often methodological —and of course need to collaborate with the health departments for access to the data.

The origins of this particular project drew on my experience, as PI, in collaborating previously with SEER cancer registries to geocode and analyze their data, in order to use area-based socioeconomic measures to investigate the impact of socioeconomic position on cancer incidence and survival. Shortly after I arrived in Boston, I contacted the Massachusetts Department of Public Health to extend this methodology to analyze the incidence of AIDS, resulting in the 1st US state-level analysis of AIDS incidence stratified by socioeconomic position, in conjunction with race/ethnicity and gender. Building on this experience, I and the team I assembled worked with relevant staff at MDPH and—because we thought it important to have another state for comparison—the Rhode Island Department of Health to make arrangements to get access to their data, while paying careful attention to all confidentiality stipulations. We then wrote & obtained, on the 1st round, an NIH R01 grant to conduct our study, which is in its last year. Among our project's products are: (a) scientific publications, (b) geocoded public health surveillance system data that we gave back to the state health departments and which otherwise would never have been geocoded, (c) training health department staff in our methods, and (d) once we have finished all the work, we will prepare a final document that we will send to all US health departments summarizing our key findings and methods.

[SLIDE 4]

We undertook our project because of an important problem: apart from data on education in birth & death certificates, most US public health surveillance systems contain no socioeconomic data. The net result is that 85% of the tables on "Health Status and Determinants" in the annual federal report *Health, United States* and 70% of the *Healthy People 2010* objectives fail to include any socioeconomic measures. Instead, the data are typically only racialized, and we have no ability to assess either socioeconomic gradients in health WITHIN racial/ethnic groups, let alone their racial/ethnic health disparities. An obvious need is thus ROUTINE monitoring of health, at the local & national level, for all populations, stratified by socioeconomic position.

[SLIDE 5]

One possible solution, making an intractable problem more tractable, as it were, involves tracts. By this I mean we can geocode our health data and then link both these data AND our population data to—and stratify by—census-derived area-based socioeconomic measures. One problem, however, is that there is NO consensus on which area-based socioeconomic measures, at which level of geography (i.e., census block group, census tract, ZIP Code), should be used. The literature instead is extremely eclectic, with myriad studies using different measures at different geographic levels, thereby precluding meaningful comparison across studies or over time.

[SLIDE 6]

The purpose of our empirical investigation thus was to determine which area-based socioeconomic measures, which we abbreviate as ABSMs, at which level of geography, would be most apt for public health monitoring. Our goal was to come up with valid, robust, easy to construct and easy to interpret ABSMs that could readily be used by any US state health department, for any health outcomes—from birth to death, and for women & men, young & old alike, among any racial/ethnic group. Guided by ecosocial theory, we anticipated different ABSMs might function differently for diverse outcomes, given likely different pathways contingent on the cumulative interplay of exposure, susceptibility, and resistance over the lifecourse. Thus, our outcomes included: mortality (all-cause and cause-specific), cancer incidence (all-sites and site-specific), low birth weight, childhood lead poisoning, sexually transmitted infections, tuberculosis, and non-fatal weapons-related injuries.

[SLIDE 7]

Our *a priori* criteria for evaluating the ABSMs, listed on this slide, pertained to: (1) external validity, (2) robustness, (3) completeness, and (4) user-friendliness.

[SLIDE 8]

This slide lists the key steps we undertook, from establishing our study base to creating our ABSMs, geocoding the health data, linking these records to the ABSMs, and generating rates stratified by the ABSMs at each level of geography. We conducted these analyses first for the total population of each state and are now in the process of additionally stratifying the results by race/ethnicity and gender. We have also been addressing various methodologic issues pertaining to multilevel data and analysis.

[SLIDE 9]

This slide presents data on our study population, defined in terms of people. In 1990, the population of MA was approximately 6 million persons and that of RI, 1 million. The number of records we obtained from each surveillance system varied by outcome, with the total equaling nearly 1 million.

[SLIDE 10]

In both states, approximately 90% of the population was White, and 4 to 5% were Hispanic or Black.

[SLIDE 11]

In terms of areas, as expected, block groups and tracts on average contained approximately 1,000 and 4,500 people respectively, and the ZIP Codes, about 13-14,000 people. Population size was most variable at the ZIP Code level and least at the block group level.

[SLIDE 12]

Listed on this slide are the 19 census-derived ABSMs we generated, 11 single variable and 8 composite, intended to capture diverse domains of socioeconomic position. These included: occupational class, income and income inequality, poverty, wealth, education, crowding, plus combinations of these variables. For the latter, we generated either pre-established indices, such as the UK Townsend and Carstairs indices of economic deprivation or the CDC Index of Local Economic Resources, and also we generated study-specific ABSMs, either via factor analysis or by creating *a priori* categorical combinations of such variables as poverty, wealth, and occupational class.

[SLIDE 13]

Because we knew we would want to display results stratified by the different ABSMs, we created 2 types of cut-points: those based on *a priori* categories vs percentiles, and more specifically quintiles. As shown on this slide, pertaining to % working class, median household income, and % below poverty, only the *a priori* categorical cut-points, highlighted in yellow, are comparable across geographic levels within and across states; cut-points for quintiles, highlighted in green, varied across these different areas.

[SLIDE 14]

It's one thing to see output, however, it's another to have a sense of its relevance to the real world. So, to see whether our measures made any sense, we randomly selected several addresses in Boston and took a look at them, writing down our impressions of the neighborhood—and then compared this to how the area would be characterized by our ABSMs. This first slide, of an economically depressed area in Boston's Chinatown, turned out to be characterized as a highly working class, poor, low income area with high unemployment and few expensive homes.

[SLIDE 15]

This one house in Beacon Hill looked like it was--and turned out to be—in a fairly affluent area: over 75% professionals, low poverty, high income, low unemployment, and lots of expensive homes.

[SLIDE 16]

And, just to give you a sense of the full terrain of our project, this slide shows poverty rates, by census tracts, in both states. The highest rates are clustered in key cities, but poverty is also high in several outlying areas.

[SLIDE 17]

Next, before geocoding our health data, we did a small study to evaluate the accuracy of several candidate geocoding firms. As we reported last year in the *American Journal of Public Health*, we found considerable variation in accuracy and cost and chose the firm that achieved 96% accuracy on a test file. Overall, we were able to geocode 92% of our nearly 1 million records to the block group level, 98% to the census tract level, and 98.2% to the ZIP Code level. Note, however, that 6.1% of the approximately

840,000 records geocoded to the ZIP Code level could not be linked to 1990 census data because their ZIP Codes either were for non-residential sites or else were created or changed after the 1990 census. As we'll show in a moment, and as we likewise reported in the *American Journal of Public Health* earlier this year, this produced some serious discrepancies between ZIP Code vs tract and block group level results.

[SLIDE 18]

Also of note were the problems we encountered coding race/ethnicity, in relation to inconsistencies within health data bases maintained WITHIN a given health department, and also with the 1990 US census categories—which delimited our denominators. Although most databases contained separate fields for "race" and for "ethnicity" (referring to Hispanic origin vs not), several—indicated by the X's in pink—included "Hispanic" in the "race" field; this problem was especially important for the MA STI, RI lead, and MA WRISS data.

[SLIDE 19]

So, having done all this, what did we find? I'll start by showing you one example, using all-cause mortality data for Massachusetts, presenting a table that we know has way too many numbers—but it does get us started. For each of the 11 ABSMs we focused on, this table presents, in the 1st 3 columns, the age-standardized rates for areas with the least resources, for each level of geography, followed by a 2nd set of 3 columns with rates for areas with the most resources. The next set of columns shows the incidence rate ratio, comparing people living in the worst to best off areas. Of note, the findings for the different ABSMs, within and across levels of geography, were actually quite similar, with results aptly summarized by the highlighted data in the last row: overall, persons in the worst-off areas had mortality rates 1.3 to 1.4 times higher than persons in the best-off areas.

[SLIDE 20]

For a different picture, however, consider this slide for colon cancer incidence. If you look at the highlighted data in the last row for the median value for the incidence rate ratio, you will see that whereas both the block group and tract measures indicated persons in the worst off areas were at somewhat less risk of colon cancer than those in the more affluent areas, the ZIP Code results suggested a socioeconomic gradient in the opposite direction.

[SLIDE 21]

One problem, however, with comparing incidence rate ratios is that classifications producing smaller groups at the margins might conceivably lead to larger effects, because a finer discrimination of extremes is achieved. To address this problem, we used an alternative parameter estimate, the relative index of inequality, or RII, which was first used with UK social class data. Its value is that the RII provides a slope estimate of risk across the full range of the distribution of the determinant, taking into account the population size of each stratum, thereby permitting meaningful comparison of gradients across different socioeconomic measures. To consolidate our key findings, we also devised what we call a scaled RII plot, in which we display the RII for a given ABSM divided by the median value for all the ABSMs being compared. This lets us determine which ABSMs are likely to detect RIIs similar to, higher than, or lower than those the median RII.

[SLIDE 22]

Here we show our scaled RII plot for ABSMs at the census tract level, looking across all outcomes for both states. First, note that for each outcome most ABSMs hovered close to the median, suggesting the impact of socioeconomic position on a given health outcome is robust. That said, measures of economic deprivation—such as the percent below poverty (the orange line) and the Townsend index (the blue line)—routinely picked up socioeconomic gradients either at or above the median. Moreover, for several outcomes—most notably HIV mortality, homicide, TB, and STIs—these measures picked up gradients far larger than those detected by the other ABSMs and also consonant with what has been reported in the literature. By contrast, measures of wealth and income inequality generally detected associations falling below the median, while those detected with measures of education hovered around the median.

[SLIDE 23]

What I will now show you are examples of what US public health data could look like if routinely stratified by an apt ABSM. In these slides, we use the tract level variable for percent below poverty—chosen because 98% of records were geocoded to the tract level and, as noted above, the poverty measure worked well in detecting socioeconomic gradients and is readily interpretable. In these slides, each bar represents the population living in a specified socioeconomic stratum, ranging from people in areas with less than 5% below poverty on up to people in areas were 20% or more live below poverty. The height of each bar depicts the rate for the health outcome, the width of the bar is proportional to the amount of the population living in the specified socioeconomic stratum, and the upper x-axis gives the incidence rate ratio, using as referent group the rate among persons in the least impoverished areas. Note that none of the outcomes I am about to show you—except death and birth--could be included in the 1998 *Socioeconomic Status and Health Chartbook* produced by the National Center for Health Statistics, precisely because their health databases lack socioeconomic data. So, starting with all-cause mortality, this figure clearly shows a poverty gradient, with persons living in census-defined poverty areas with more than 20% below poverty experiencing the highest death rates.

[SLIDE 24]

This slides next shows the expected reverse socioeconomic gradient for breast cancer incidence, with rates highest among women in the least impoverished areas.

[SLIDE 25]

Next are data on the percent of low birthweight births. And, of note, the 2-fold excess risk among women in the most compared to the least impoverished areas is equivalent to the 2-fold excess we observed comparing women with less than a high school education to women who had completed 4 or more years of college, using educational data from the birth certificates.

[SLIDE 26]

Here are data on the percent of children with elevated lead levels, showing the over 9-fold excess among those living in the most compared to least impoverished areas.

[SLIDE 27]

Ditto results for tuberculosis, with an excess risk is 8-fold.

[SLIDE 28]

For syphilis, the excess risk jumps to 18-fold.

[SLIDE 29]

And for non-fatal weapons-related injury, it is 11-fold. In other words, for none of these outcomes do we have trivial socioeconomic gradients—yet, in current US public health reports, all of these gradients are routinely ignored and unreported.

[SLIDE 30]

And, particularly germane to this hearing, these next 2 slides hint at what these types of analyses could reveal about socioeconomic gradients within racial/ethnic-gender groups and the contribution of socioeconomic inequalities to racial/ethnic disparities in health. Using the example of premature mortality (before the age of 65), this first slide shows 2 key findings, for 3 census-defined groups: whites (non-Hispanic & Hispanic), blacks (non-Hispanic & Hispanic), and Hispanic (all "races"). The first, shown by the data in red, is that whereas nearly half the white women and men lived in CT with <5% below poverty, half or more African-Americans and Hispanics lived in CT with ≥20% below poverty. Second, shown by the data in blue, within each racial/ethnic group there were marked socioeconomic gradients in premature mortality, ranging from a 2-fold excess among white women, comparing those in CT with ≥20% vs <5% below poverty, up to a 4-fold excess for Black and Hispanic men.

[SLIDE 31]

Next, this slide shows that, first, for men and women, rates of premature mortality were higher among blacks compared to whites (the data in bright red) chiefly in the more impoverished CT. Second, adjusting for CT poverty (data in dark red) reduced the overall age-adjusted excess risk of premature mortality (data in pink) from 1.8 to 1.3 for men and from 1.7 to 1.3 for women.

[SLIDE 32]

So, to start summing up, our data suggest that: (1)) it IS feasible—and informative--to monitor US socioeconomic inequalities in health using ABSMs, and (2) the choice of both ABSM and level of geography matters.

Before you accept our results, however, it is important to consider several possible sources of error and bias. First, our results if anything underestimate, rather than overestimate, likely socioeconomic gradients in health, because if poorer persons were less likely to be geocoded or to be included in these health surveillance systems, we would be missing the worst-off part of the population. Second, suggesting was little bias in ascertainment of the determinant, ABSM data were typically missing for under 1% of the geocoded areas. Third, from a temporal standpoint, the simultaneity of measurement of the ABSMs and outcomes is appropriate, because the point of monitoring is to determine where the burden of disease falls; studies with a more etiologic focus would of course need to take into account etiologic period.

From a spatial standpoint, the results we presented have not taken into account spatial correlation, and we're actually doing analyses right now to examine its impact on our findings.

Of particular importance, however, I want to stress that issues of ecologic fallacy are not germane to the present study design, since individuals constituted the unit of observation for both the dependent variables (health outcomes) and the independent variables (living in an area with certain sociodemographic characteristics). Instead, validity of using ABSMs depends on the extent to which areas constitute meaningful geographic units—a different question from whether they are "proxies" for individual-level socioeconomic data. Lastly, a lack of comparable studies means we can't say how our results are similar to or differ from those in the literature. That said, among the handful of studies comparing gradients detected with individual- vs area-based socioeconomic measures, they typically have found more consistency in results with block group and tract vs ZIP Code level measures.

[SLIDE 33]

So, to offer some interpretations, starting with level of geography, our finding that tract and block group level data behaved similarly, whereas ZIP Code level data were more problematic, was consistent with our expectations. And, in addition to noting problems introduced by the spatiotemporal mismatch between ZIP Codes and US census-defined areas, we remind you that ZIP Codes have been replaced by ZIP Code Tabulation Areas (ZCTAs) in the Year 2000 census. This effectively renders moot the possibility of simply using people's mailing address ZIP Code to link to US census data, because ZIP Codes and ZCTAs sharing the same code may in fact encompass different geographic areas.

Second, with regard to ABSMs, what stands out is the robust impact of socioeconomic position on health—in that, by and large, whatever measure you use, with only a few exceptions, you can document powerful socioeconomic gradients in health. That said, the most sensitive ABSMs, across all outcomes, were those measuring economic deprivation. And here, it is striking to note that the single variable measure "percent below poverty" performed as well or better than virtually all of the more complex composite measures, which are far harder to construct and to explain.

Third, we think it important to flag some unanswered questions we're right now in the midst of addressing. One pertains to whether our results, shown for the total population, hold for different racial/ethnic-gender groups, and our preliminary results suggest they do. The second pertains to the multilevel nature of our data. We're just now completing analyses investigating whether the nesting of block groups within census tracts matters for effect estimates, and our tentative answer is: "sometimes"—i.e., only in cases when BOTH the ABSM AND outcome exhibit strong spatial clustering. We're likewise exploring the contribution of different levels to the spatial distribution of our outcomes, and finding that geographic variation at the tract level at times increases when we take into account individual-level data, contrary to what might be expected. We hope to be reporting on these findings in the scientific literature in the next year.

[SLIDE 34]

And so, based on the evidence you have seen today and our additional analyses underway, our tentative conclusion—drawing on both our *a priori* criteria and also several desirable attributes of indicators, as summarized on this slide—is that efforts to monitor US socioeconomic inequalities in health using ABSMs will be best served by those tract or block group measure that are: (1) the most attuned to capturing economic deprivation, (2) meaningful across regions and over time (and hence use *a priori* categorical cut-points), (3) have little missing data, and (4) are easily understood. In our view, the best candidate ABSM meeting all of these criteria is the "percent below poverty," at the census tract level.

[SLIDE 35]

In conclusion, then, monitoring of social inequalities in health in the United States requires that health departments collect data on BOTH race/ethnicity and class—BOTH matter. The realities of socioeconomic inequality and the impact of past and present racial discrimination, both economic and non-economic, means we must have data on both of these dimensions of social life if we to generate informative data on distributions and determinants of population health. Moreover, we need identical data on race/ethnicity and socioeconomic position for the numerators and denominators, hence the critical importance of working with US census categories and counts.

Our overall recommendations are, first, that more work needs to be done to ensure the consistency of coding of race/ethnicity across US public health data bases and the census. And, second, as importantly, all US public health data bases should be routinely geocoded and employ standard ABSMs that can be compared across states and over time, so that we can ROUTINELY monitor socioeconomic disparities in health WITHIN racial/ethnic groups and ROUTINELY assess the contribution of socioeconomic inequality to racial/ethnic disparities in health.

[SLIDE 36]

Lastly, as a resource on the methodology we propose, we refer you to the following papers from our Public Health Disparities Geocoding Project, some already published, some in press, with a promise of more underway. There are also 2 additional slides, also available on handouts, with references for prior studies I have published jointly examining racial/ethnic and socioeconomic disparities in health, using area-based socioeconomic measures, as well as articles pertaining to conceptualizing, measuring and analyzing social inequalities in health involving race/ethnicity, class, and gender.

Thank you very much.

[SLIDES 37 & 38: Additional references]

THE PUBLIC HEALTH DISPARITIES GEOCODING PROJECT:

IMPROVING MONITORING OF SOCIAL INEQUALITIES IN HEALTH IN THE UNITED STATES

Nancy Krieger, Pamela D. Waterman, Jarvis Chen,
Mah-Jabeen Soobader, SV Subramanian, David Rehkopf
Harvard School of Public Health

Funded by: NIH/NICHD (1 R01 HD36865-01)

Presented to the Subcommittee on Populations -
National Committee on Vital and Health Statistics,
Nov 8, 2002
Philadelphia, PA

STUDY PARTNERS: MA & RI STATE HEALTH DEPTS

Massachusetts Department of Public Health (MDPH):

Dr. Daniel Friedman (Assistant Commissioner, Bureau of Health Statistics, Research and Evaluation);

Alice Mroszczyk, Program Coordinator for 24A/B/111B Review Committee;

***Cancer Registry:* Dr. Susan Gershman, Director; Mary Mroszczyk, Geocoding/Special Projects Coordinator; Ann R. MacMillan, Data Analyst;**

***Registry of Vital Records and Statistics:* Elaine Trudeau, Registrar of Vital Records; Charlene Zion, Public Information Office;**

***Infectious Diseases:* Alfred DeMaria, Assistant Commissioner; Yuren Tang, Chief, Surveillance Program, Sharon Sharnprapai, TB epidemiologist;**

***Weapons-Related Injury Surveillance System:* Victoria Vespe Ozonoff, Program Director; Beth Hume, Data Manager/Analyst; Patrice Cummins, Epidemiologist; Laurie Janelli, Site Coordinator**

Rhode Island Department of Health (RIDOH):

Dr. Jay Buechner (Chief, Office of Health Statistics, Rhode Island Department of Health);

***Vital Statistics:* Roberta Chevoya, State Registrar of Vital Records;**

***Division of Disease Prevention and Control:* Dr. John Fulton, Associate Director, Ted Donnelly, Senior Public Health Epidemiologist; and Michael Goscminksi, Epidemiologist;**

***Environmental Health Risk Assessment:* Susan Feeley, Public Health Epidemiologist (no longer at RIDOH);**

***Childhood Lead Poisoning Prevention:* Magaly Angeloni, Program Manager.**

VITAL COLLABORATIONS:
UNIVERSITIES + HEALTH DEPARTMENTS

- Premise: research vital to improving monitoring social inequalities in health
  - Health depts: aware of this, lack time & resources
  - University researchers: can get grants, improve methods
- Process:
  - Drew on PI's experience with working with SEER cancer registries & getting grants to geocode & analyze their data
  - Contacted MDPH & RIDOH, worked with all data stakeholders on confidentiality & data use issues
- Resources:
  - University: obtained NIH R01 grant; time; expertise
  - Health depts: staff time to access data
- Products:
  - Scientific manuscripts & report to all state health depts
  - Training of MDPH & RIDOH staff in methods

THE PROBLEM

- Limited socioeconomic data in US public health data
  - Birth & death certificates: education only since 1989
  - Cancer, tuberculosis, AIDS, lead, and related registries, plus most hospital data: NO socioeconomic information
- Effect on US federal reports & public health objectives
  - Health, United States: 85% of the 73 tables on "Health Status and Determinants" have NO socioeconomic data (exception: 1998 "Socioeconomic Status and Health Chartbook")
  - Healthy People 2010: 70% of 467 public health objectives include NO socioeconomic data
- Limits analysis of racial/ethnic disparities in health
  - Cannot assess socioeconomic gradients in health WITHIN racial/ethnic groups or how they contribute to racial/ethnic disparities in health.
- Needed: routine monitoring of health, at local level, for all populations, stratified by socioeconomic position

POSSIBLE SOLUTION:
GEOCODING & AREA-BASED SOCIOECONOMIC DATA

- **Geocoding:** can link health data (numerators) & population data (denominators) to--and stratify by--census-derived area-based socioeconomic measures (thereby making an intractable problem more tractable …)
- **Problem: NO** standard validated US area-based socioeconomic measures for research or monitoring
  - no consensus on <u>which</u> area-based socioeconomic measures, at <u>which</u> level of geography (census block group, census tract, ZIP Code), should be used
  - US literature: extremely eclectic, with studies using different measures and different geographic levels, precluding meaningful comparisons across studies or over time.

PURPOSE OF PROJECT

- **Empirical investigation:** to determine which area-based socioeconomic measures, at which level of geography, are apt for monitoring US socioeconomic inequalities in health
- **Goal:** to generate <u>valid</u>, <u>robust</u>, <u>easy to construct</u>, & <u>easy to interpret</u> area-based socioeconomic measures (ABSMs) that can be readily used:
  - by any US state health department
  - for any health outcome, from birth to death
  - for women & men, young & old, among any racial/ethnic group
  - Ecosocial rationale: anticipated different ABSMs might function differently for diverse outcomes, given likely different pathways contingent on the cumulative interplay of exposure, susceptibility, and resistance over the lifecourse
- **Outcomes:** mortality (all-cause, cause-specific), cancer incidence (all-sites, site-specific), low birth weight, childhood lead poisoning, sexually transmitted infections (STI), tuberculosis (TB), and non-fatal weapons-related injuries

A PRIORI CRITERIA FOR ABSMs

- **External validity:** do the measures find gradients in the direction reported in the literature, i.e., positive, negative, or none, and across the full range of the distribution?
- **Robustness:** do the measures detect expected gradients across a wide range of outcomes?
- **Completeness:** is the measure relatively unaffected by missing data?
- **User-friendliness:** how easy is the measure to understand and explain?

METHODS: OVERVIEW

- Steps:
  1) Generate study base of population, areas, and outcomes centered around the 1990 census
  2) Create diverse ABSMs at each level of geography
  3) Geocode health records and link to the ABSMs
  4) Stratify health outcomes by these ABSMs at each level of geography and compare detected socioeconomic gradients (including to expected gradients, based on extant literature)
- Conduct analyses for total population, plus stratified by race/ethnicity & gender
- Address methodologic issues pertaining to multilevel data and analysis

STUDY POPULATION: PEOPLE

(final analytic data sets)

	MA	RI
1990 population	6,016,425	1,003,464
Mortality data* (1989-1991)	155,764	27,287
Cancer data** (primary invasive; 1988/9-92)	140,610	19,798
Birth data (MA: 1989-1991; RI: 1987-1993)	267,311	96,138
Childhood lead screening (1994-1995)	na	62,514
STIs (MA: 1994-1998; RI: 1994-1996)	39,144	6,403
TB (MA: 1993-1998; RI: 1985-1994)	1,793	576
Non-fatal weapons related injury (1995-1997)	5,571	na

•*all-cause, plus analyses of top 5 causes by race/ethnicity: heart disease, malignant neoplasm, cerebrovascular disease, pneumonia and influenza, chronic obstructive pulmonary disease, unintentional injury, diabetes, HIV, and homicide and legal intervention.

*MA: 1988-1992; RI: 1989-1992; all-cause, plus analyses of breast, cervix, colon, lung, prostate

RACIAL/ETHNIC COMPOSITION: MA & RI, 1990 CENSUS

Race/ethnicity (by rank order)	Massachusetts		Rhode Island	
White	5,405,374	(90%)	917,375	(91%)
Hispanic Origin (of any race)	287,549	(5%)	45,752	(5%)
Black	300,130	(5%)	38,861	(4%)
Other	155,288	(3%)	24,832	(2%)
Asian or Pacific Islander	143,392	(2%)	18,325	(2%)
American Indian, Eskimo or Aleut	12,241	(0.2%)	4,071	(0.4%)

STUDY POPULATION: AREAS

	N	Mean population size (SD)	Range
Massachusetts			
Block group	5,603	1,085.4 (665.2)	5 to 10,096
Census tract	1,331	4,571.8 (2,080.0)	18 to 15,411
ZIP Code	424	12,719.7 (12,244.1)	14 to 65,001
Rhode Island			
Block group	897	1,137.7 (670.8)	7 to 5,652
Census tract	235	4,325.3 (1,810.9)	26 to 9,822
ZIP Code	70	14,335.2 (13,234.8)	63 to 53,763

19 CENSUS-DERIVED ABSMs

Construct	Measure (quintiles & *a priori* categories)
11 Single variable:	
--Occupational class	% working class, % unemployed
--Income	median household (HH) annual income low income: % HH <50% median income high income: % HH ≥ $150,000 per year Gini coefficient (income inequality)
--Poverty	% < Poverty
--Wealth	% owner-occupied homes > $300,000
--Education	low: % adults < high school education high: % adults ≥ 4 yrs college education
--Crowding	% crowded households
8 Composite:	--Townsend index --Carstairs index --CDC Index of Local Economic Resources --SEP1, SEP2, factor 1, factor 2, SEP index

CUT-POINTS:
CATEGORICAL VS PERCENTILE

Single variable		L U	area	Q1 L U	Q2 L U	Q3 L U	Q4 L U	Q5 L U
	C1:	(0.0 , 49.9)						
Working	C2:	(50.0 , 65.9)						
class (%)	C3:	(66.0 , 74.9)						
(categorical)	C4:	(75.0 , 100.0)						
			MA BG	(4,999 , 26,110)	(26,111 , 33,749)	(33,750 , 40,798)	(40,799 , 49,903)	(49,904 , 150,001)
			MA CT	(4,999 , 26,471)	(26,472 , 33,162)	(33,163 , 39,286)	(39,287 , 47,124)	(47,125 , 102,797)
Median			MA ZC	(9,726 , 30,624)	(30,625 , 36,246)	(36,247 , 41,396)	(41,397 , 48,841)	(48,842 , 94,898)
household			RI BG	(4,999 , 22,088)	(22,089 , 30,293)	(30,294 , 35,567)	(35,568 , 41,204)	(41,205 , 150,001)
income($)			RI CT	(6,462 , 23,667)	(23,668 , 31,032)	(31,033 , 35,300)	(35,301 , 40,606)	(40,607 , 78,666)
(quintile)			RI ZC	(8,787 , 29,548)	(29,549 , 33,614)	(33,615 , 36,921)	(36,922 , 41,356)	(41,357 , 60,705)
	C1:	(0.0 , 4.9)						
Poverty(%)	C2:	(5.0 , 9.9)						
(categorical)	C3:	(10.0 , 19.9)						
	C4:	(20.0 , 100.0)						

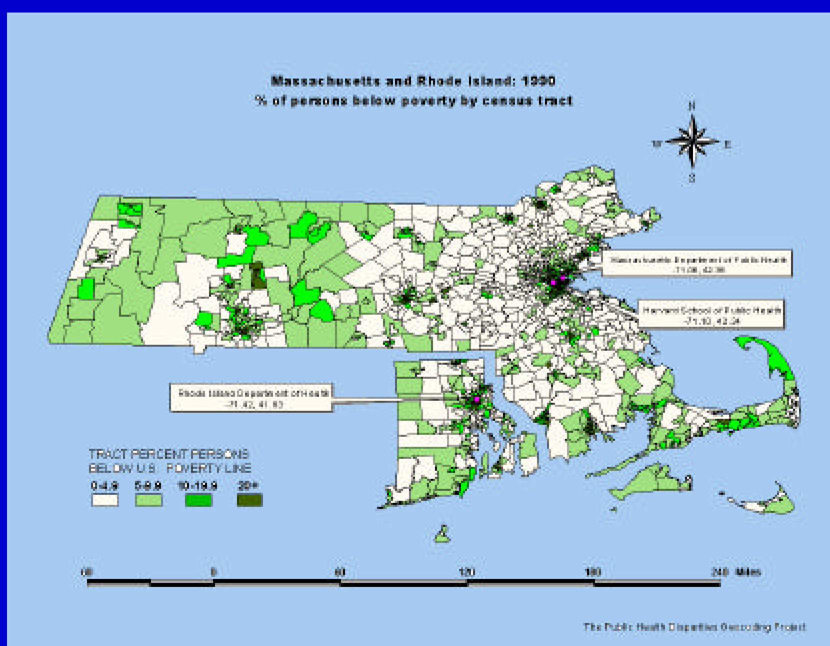
96 Tyler Street

- 86.4 % working class
- 15.6 % unemployed
- 26.5 % below poverty line
- $18,607 median household income
- 5.1 % owner-occupied homes valued >$300,000

138 Mount Vernon St

(this is one home, not an apartment building)

- 26.4 % working class
- 5.4 % unemployed
- 8.0 % below poverty line
- $84,959 median household income
- 40.2 % owner-occupied homes valued >$300,000



GEOCODING RESULTS

1) Evaluation of accuracy: did pilot test and selected firm which accurately geocoded 96% of a random sample of records

2) % of records geocoded:

Database	N	% GEOCODED TO: BG	CT	ZC	% NOT geocoded
TOTAL	970,086	92%	98%	98.2%	1.8%

NB:

a) the % geocoded was similar across most databases and did not vary notably by age, gender, or race/ethnicity, and

b) 6.1% of the 839,748 records geocoded to the ZC level could not be linked to 1990 census data because their ZIP Codes either were for non-residential sites (e.g., agencies, businesses with high mail volume, or post offices and their P.O. Boxes) or else were ZIP Codes created or changed after the 1990 census.

PROBLEMS CODING RACE/ETHNICITY:
INCONSISTENCIES WITHIN HEALTH DEPTS & WITH 1990 US CENSUS

"Race/ethnicity" fields		Death		Birth		Cancer		STIs		TB		Weapons-related injury	Lead
		MA	RI	MA	RI	MA	RI	MA	RI	MA	RI	MA	RI
"Race"	White	X	X	X	X	X	X	X	X	X	X	X	X
	Black	X	X	X	X	X	X	X	X	X	X	X	X
	American Indian	X	X	X	X	X	X	X	X			X	X
	Asian/Pacific Islander	X	X	X	X	X	X	X	X	X	X	X	X
	Other	X	X	X	X	X			X			X	X
	Hispanic	X						X				X	X
"Ethnicity"	Hispanic	X	X	X			X		X	X	X		

X = INCLUDES HISPANIC ORIGIN; X = HISPANIC + PORTUGEUSE
X = INCLUDES HAITIANS AND CAPE VERDEANS

% OF RECORDS WITH "RACE" LISTED AS "HISPANIC":
MA DEATH = 1%, MA STI = 17%, RI LEAD = 14%, MA WRISS = 20%

MA: ALL-CAUSE MORTALITY
(1989-1991)

Mortality Outcome	Area-based socioeconomic measure	Rate: least resources			Rate: most resources			Incidence Rate Ratio (IRR) (95% CI): least/most		
		BG	CT	ZC	BG	CT	ZC	BG	CT	ZC
All cause	Working class (categorical)	929.7	966.6	900.3	718.9	749.8	647.1	1.29 (1.23, 1.36)	1.29 (1.22, 1.36)	1.39 (1.30, 1.49)
	Median household income (quintile)	954.9	1006.7	927.0	747.9	781.1	698.9	1.28 (1.22, 1.34)	1.29 (1.23, 1.35)	1.33 (1.26, 1.39)
	Poverty (categorical)	1030.7	1060.4	1070.5	763.3	800.1	766.8	1.35 (1.29, 1.42)	1.33 (1.26, 1.39)	1.40 (1.32, 1.47)
	Gini (quintile)	865.5	937.1	884.3	840.2	854.9	822.7	1.03 (0.98, 1.08)	1.10 (1.04, 1.15)	1.07 (1.01, 1.14)
	Wealth (categorical)	834.3	886.1	880.5	703.7	751.1	665.9	1.19 (1.13, 1.24)	1.18 (1.13, 1.23)	1.32 (1.26, 1.39)
	Crowding (categorical) ‡	1119.4	1024.6	944.7	782.7	837.6	803.5	1.43 (1.23, 1.67)	1.22 (1.00, 1.5)	1.18 (0.69, 2.00)
	Low education (categorical)	962.4	986.6	960.8	752.3	780.4	734.9	1.28 (1.22, 1.34)	1.26 (1.20, 1.33)	1.31 (1.23, 1.39)
	Townsend index (quintile)	1001.9	1049.9	938.2	743.2	777.8	733.3	1.35 (1.28, 1.42)	1.35 (1.28, 1.42)	1.28 (1.21, 1.35)
	Index of Local Economic Resources (quintile)	952.5	1005.9	953.3	726.7	769.8	681.5	1.31 (1.25, 1.37)	1.31 (1.25, 1.37)	1.40 (1.34, 1.46)
	SEP1 (categorical)	1025.6	1036.3	1043.9	687.4	741.7	646.2	1.49 (1.38, 1.61)	1.40 (1.30, 1.51)	1.62 (1.43, 1.82)
	SEP Index (quintile)	934.8	1004.2	934.4	712.1	754.5	672.1	1.31 (1.25, 1.38)	1.33 (1.27, 1.4)	1.39 (1.33, 1.46)
	Median value	954.9	1005.9	938.2	743.2	777.8	698.9	1.31	1.29	1.33

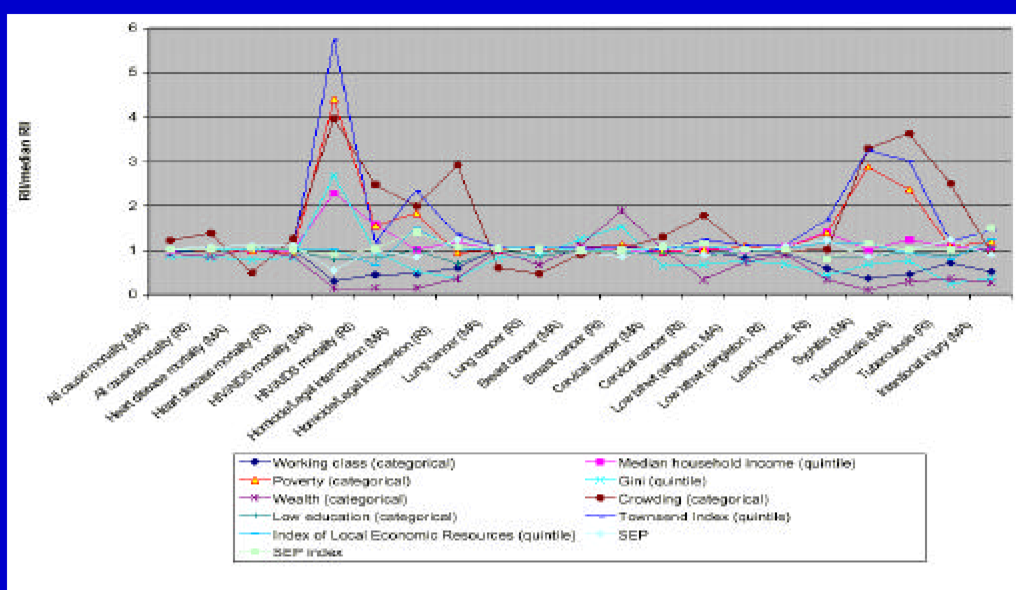
MA: COLON CANCER INCIDENCE
(1988-1992)

Cancer Site	Area-based socioeconomic measure	Rate: least resources			Rate: most resources			Incidence Rate Ratio (IRR) (95% CI): least/most		
		BG	CT	ZC	BG	CT	ZC	BG	CT	ZC
Colon	Working class (categorical)	41.3	42.5	41.1	45.8	48.3	27.9	0.90 (0.76, 1.06)	0.88 (0.73, 1.06)	1.47 (1.14, 1.90)
	Median household income (quintile)	41.0	42.5	42.3	46.3	48.9	37.2	0.89 (0.75, 1.04)	0.87 (0.74, 1.03)	1.14 (0.97, 1.34)
	Poverty (categorical)	41.7	45.6	44.8	43.9	47.4	41.6	0.95 (0.80, 1.13)	0.96 (0.81, 1.15)	1.08 (0.88, 1.32)
	Gini (quintile)	42.4	46.1	39.9	46.3	47.3	44.6	0.92 (0.77, 1.08)	0.97 (0.83, 1.15)	0.89 (0.73, 1.10)
	Wealth (categorical)	42.8	46.5	44.9	43.7	48.6	31.7	0.98 (0.84, 1.14)	0.96 (0.83, 1.10)	1.42 (1.20, 1.67)
	Crowding (categorical) ‡	45.4	35.3	106.5	42.6	47.0	41.4	1.07 (0.59, 1.92)	0.75 (0.32, 1.76)	2.57 (0.78, 8.49)
	Low education (categorical)	39.5	40.8	43.8	45.2	48.0	39.3	0.87 (0.73, 1.05)	0.85 (0.70, 1.03)	1.11 (0.90, 1.38)
	Townsend index (quintile)	43.9	46.1	42.0	44.9	47.6	39.6	0.98 (0.83, 1.16)	0.97 (0.82, 1.15)	1.06 (0.88, 1.28)
	Index of Local Economic Resources (quintile)	40.3	42.6	43.1	45.4	48.7	33.6	0.89 (0.76, 1.04)	0.87 (0.74, 1.03)	1.28 (1.09, 1.50)
	SEP1 (categorical)	41.6	42.5	44.3	45.4	48.2	28.8	0.92 (0.70, 1.21)	0.88 (0.67, 1.16)	1.54 (0.99, 2.40)
	SEP Index (quintile)	40.2	42.5	42.7	46.5	48.4	34.8	0.86 (0.73, 1.02)	0.88 (0.74, 1.04)	1.23 (1.04, 1.45)
	Median value	41.6	42.5	43.1	45.4	48.2	37.2	0.92	0.88	1.23

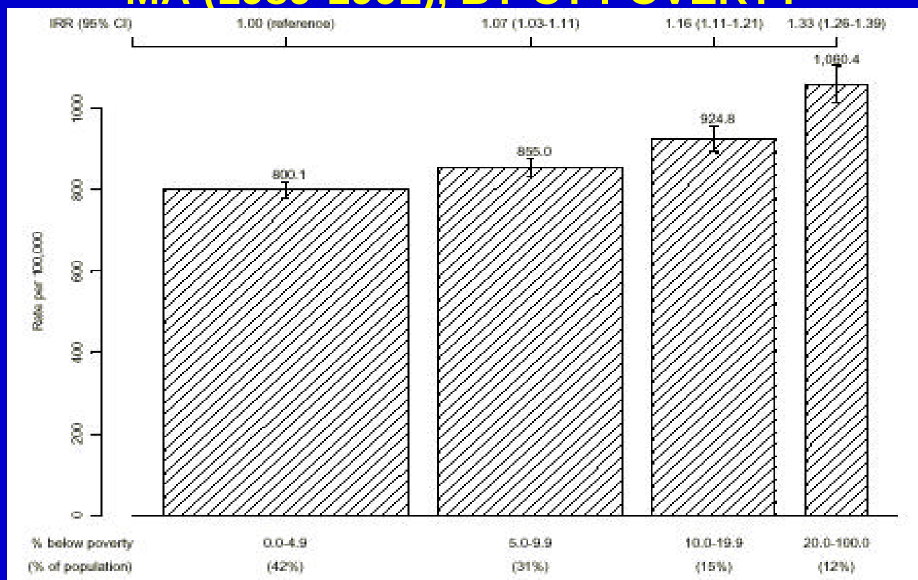
RELATIVE INDEX OF INEQUALITY (RII)

- **Problem with comparisons of IRR:**
  - classifications producing smaller groups at the margins might conceivably lead to larger incidence rate ratios, e.g., comparing the most deprived to the most affluent, because finer discrimination of extremes is achieved
- **Alternative = Relative Index of Inequality (RII)**
  - provides a slope estimate of the risk estimate (e.g., rate ratio) across the full range of the distribution of the determinant, taking into account the population size of each stratum, thereby permitting meaningful comparison of gradients across different socioeconomic measures
  - Scaled RII plot: displays RII for a given ABSM divided by the median value for all the ABSMs being compared

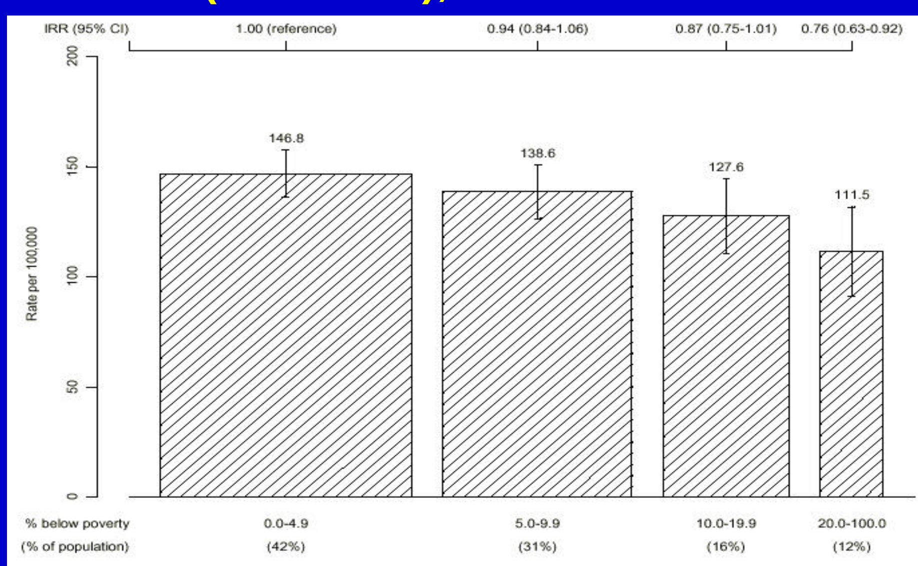
SCALED RII PLOT: CT LEVEL



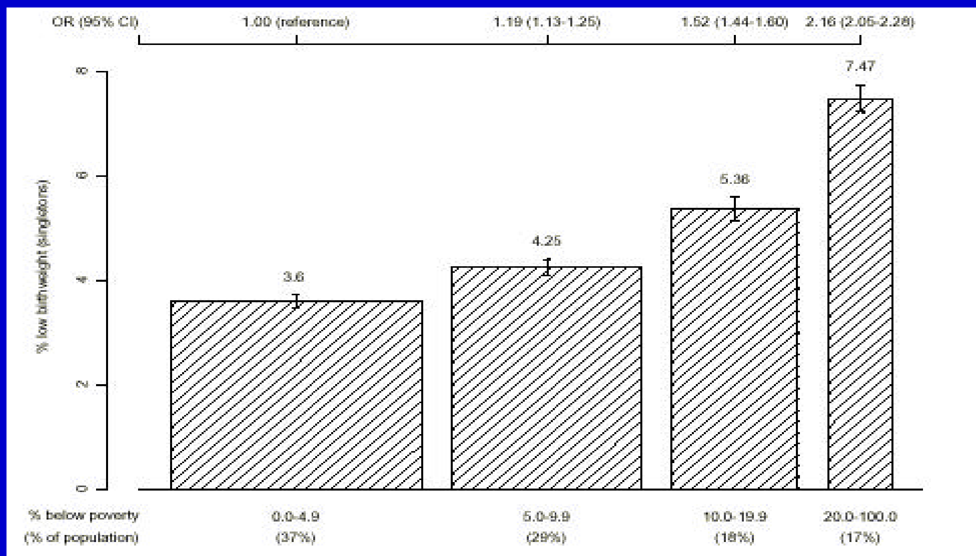
MORTALITY (ALL-CAUSE): MA (1989-1991), BY CT POVERTY



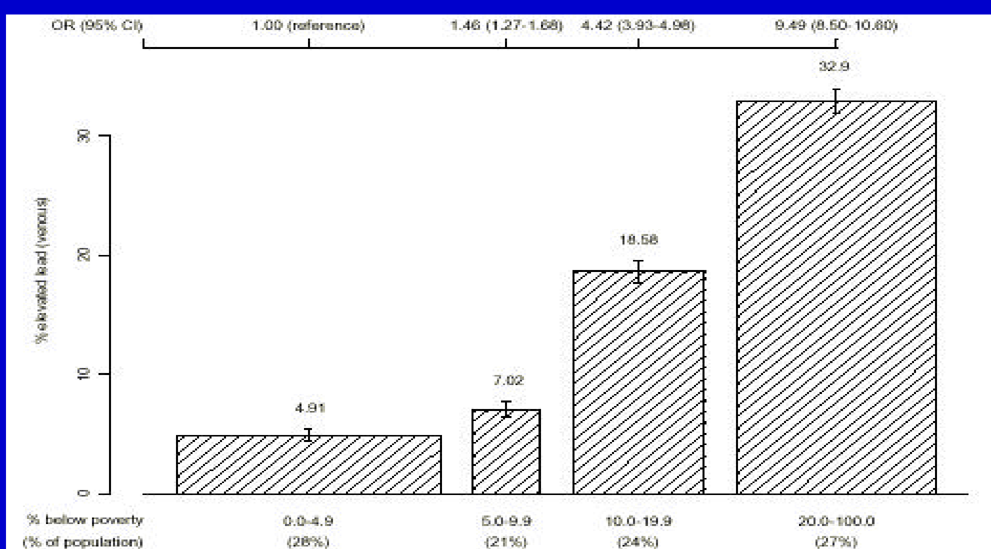
BREAST CANCER INCIDENCE: MA (1998-1992), BY CT POVERTY



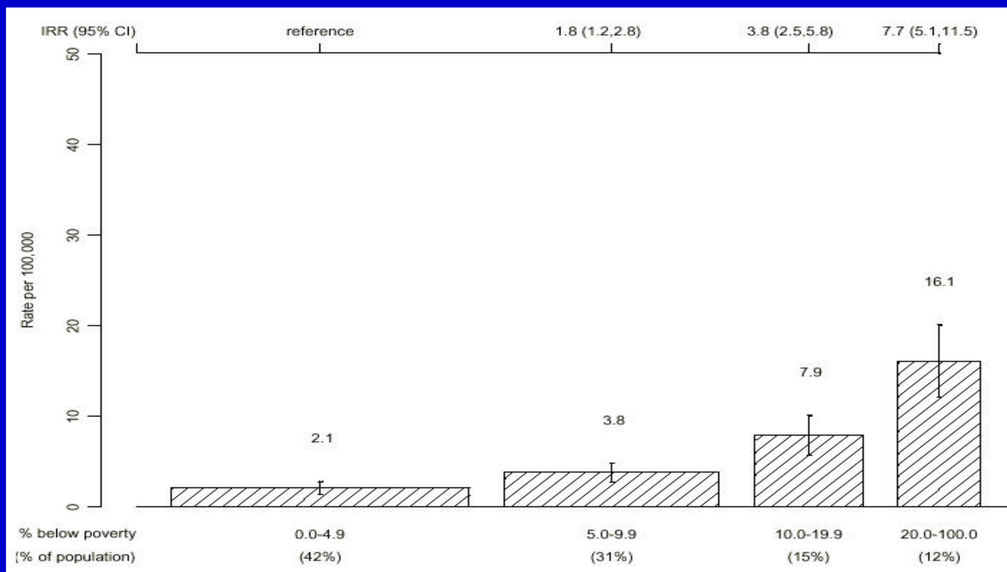
% LOW BIRTHWEIGHT: MA (1989-1991), BY CT POVERTY



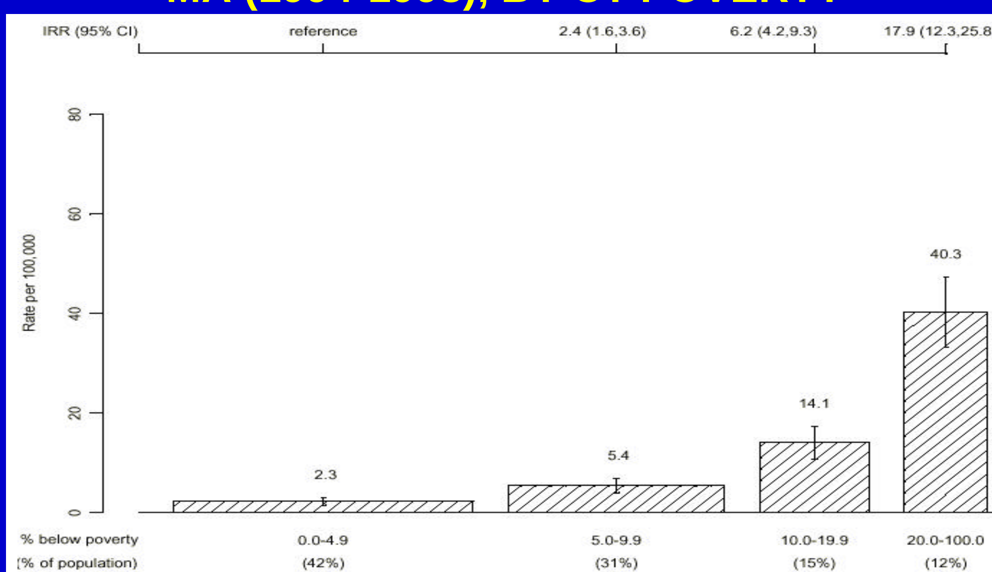
% ELEVATED LEAD IN CHILDREN: RI (1994-1996), BY CT POVERTY



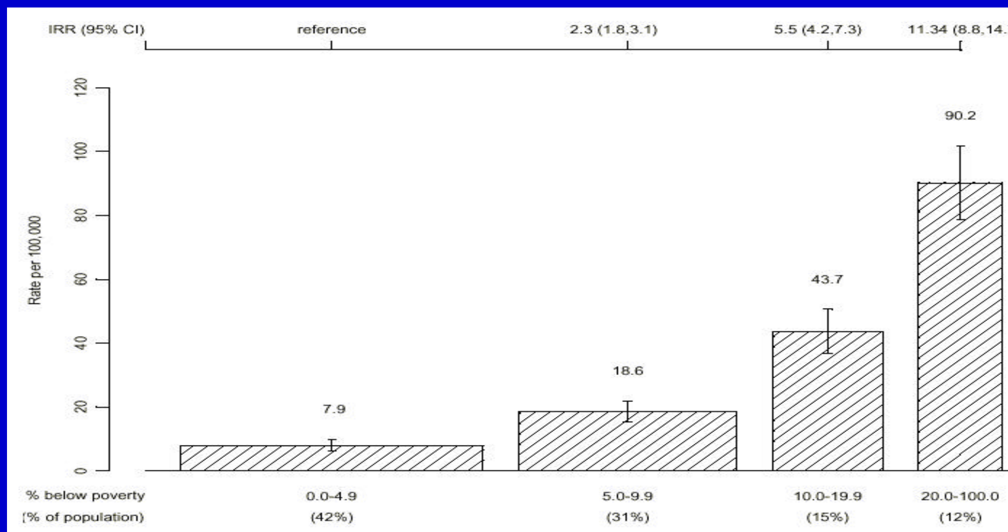
TUBERCULOSIS:
MA (1993-1998), BY CT POVERTY



SYPHILIS:
MA (1994-1998), BY CT POVERTY



NON-FATAL INTENTIONAL WEAPONS-RELATED INJURY: MA (1995-1997), BY CT POVERTY



PREMATURE MORTALITY (< 65 yo)
BY RACE/ETHNICITY & GENDER (CT % < POVERTY) MA, 1989-1991

% below poverty (CT)	N	(%)	Rate*	IRR (95% CI)	RII (95% CI)	N	(%)	Rate*	IRR (95% CI)	RII (95% CI)
	WHITE MEN					WHITE WOMEN				
<5%	3,203,058	(47%)	212.7	(1.0)		3,234,468	(46%)	127.6	(1.0)	
5-9.9%	2,145,999	(31%)	285.3	1.3 (1.2, 1.4)	2.6	2,205,174	(32%)	153.2	1.2 (1.1, 1.3)	1.9
10-19.9%	983,442	(14%)	356.5	1.7 (1.5, 1.8)	(2.5, 2.8)	1,012,152	(14%)	177.6	1.4 (1.2, 1.6)	(1.8, 2.0)
$\geq$20%	537,120	(8%)	481.4	2.2 (2.0, 2.5)		535,689	(8%)	226.8	1.8 (1.5, 2.1)	
	BLACK MEN					BLACK WOMEN				
<5%	42,777	(11%)	180.1	(1.0)		34,530	(8%)	127.8	(1.0)	
5-9.9%	69,978	(17%)	272.7	1.5 (0.8, 2.9)	4.0	69,663	(16%)	180.7	1.4 (0.7, 3.0)	2.2
10-19.9%	96,048	(24%)	422.3	2.3 (1.3, 4.3)	(3.3, 4.8)	101,934	(24%)	232.8	1.8 (0.9, 3.7)	(1.7, 2.8)
$\geq$20%	197,895	(49%)	598.1	3.3 (1.9, 5.8)		220,539	(52%)	276.7	2.2 (1.1, 4.2)	
	HISPANIC MEN					HISPANIC WOMEN				
<5%	41,931	(11%)	93.4	(1.0)		37,938	(9%)	29.8	(1.0)	
5-9.9%	64,854	(16%)	142.6	1.5 (0.6, 4.0)	3.8	64,392	(16%)	86.9	2.9 (0.6, 13.9)	3.2
10-19.9%	81,999	(21%)	263.5	2.8 (1.2, 6.8)	(3.0, 5.0)	84,606	(21%)	96.5	3.2 (0.7, 15.0)	(2.2, 4.8)
$\geq$20%	205,635	(52%)	325.4	3.5 (1.5, 7.9)		221,898	(54%)	134.7	4.5 (1.1, 19.3)	

*Age-standardized rate (per 100,000) to the year 2000 standard million

PREMATURE MORTALITY (<65 yrs old): BLACK/WHITE DISPARITIES

WITHIN & ADJUSTING FOR SOCIOECONOMIC POSITION (CT % < POVERTY), MA, 1989-1991

	Black/White IRR* (95% CI)	
% below poverty (CT)	Men	Women
<5%	0.8 (0.6, 1.1)	1.0 (0.8, 1.4)
5-9.9%	1.1 (0.9, 1.2)	1.3 (1.1, 1.6)
10-19.9%	1.4 (1.2, 1.5)	1.4 (1.2, 1.6)
≥20%	1.4 (1.3, 1.5)	1.2 (1.0, 1.5)
Overall	1.8 (1.8, 1.9)	1.7 (1.6, 1.8)
Adjusted for % < poverty (CT)	1.3 (1.2, 1.4)	1.3 (1.2, 1.4)

*all models adjusted for age, using Poisson regression models

DISCUSSION

- Key findings:
  - feasible to monitor US socioeconomic inequalities in health using ABSMs
  - Choice of <u>both</u> area-based socioeconomic measure (ABSM) <u>and</u> level of geography matters
- Sources of error & bias
  - Geocoding & underregistration of cases: if associated with poverty, then a conservative bias
  - ABSM: very small % missing data
  - Temporal: simultaneity ok for monitoring (burden of disease; not same as etiologic research)
  - Spatial correlation: addressing its impact on findings
  - Ecologic fallacy: not relevant
- Comparison to prior studies: none directly comparable

INTERPRETATION & IMPLICATIONS

- **Level of geography:**
  - similarity of block group & census tract expected
  - ZIP Code data: subject to spatiotemporal mismatch with US census data, plus no ZIP Codes in 2000 census (ZCTA instead …)
- **Choice of area-based socioeconomic measures:**
  - Overall: robust, but measures of economic deprivation: most sensitive, across all outcomes
  - "Percent below poverty": performed as well as composite measures
- **Unanswered questions (analyses underway):**
  - Stratification by race/ethnicity and gender: see same patterns for associations of ABSMs and outcomes across levels of geography?
  - Multilevel models
    - Does ignoring nesting of BG within CT bias effect estimates?
    - Contribution of different levels to spatial distribution of outcomes

TENTATIVE CONCLUSION

- **Based on: (a) our *a priori* criteria, and (b) desirable attributes of an indicator (Rossi & Gilmartin, 1980)**
  - conceptually-based
  - constructed from valid, reliable, and accessible data using appropriate statistical techniques
  - comparable over time and across population groups; and
  - readily understandable, with normative value relevant to timely policy making
- **Efforts to monitor US socioeconomic inequalities in health using area-based socioeconomic measures will be best served by those tract or block group measures that are**
    - most attuned to capturing economic deprivation
    - meaningful across regions and over time (hence use *a priori* categorical cut-points)
    - have little missing data. and
    - easily understood
- **Likely candidate: tract-level "% below poverty"**

CONCLUDING COMMENTS:
RACE/ETHNICITY & CLASS & HEALTH

- **To monitor social inequalities in health:**
  - imperative to collect data on BOTH race/ethnicity AND class
  - need: identical data on race/ethnicity & socioeconomic position for numerators (cases) and denominators (population),
  - hence importance of US census categories and counts

- **Overall recommendation:**
  - ensure consistency of racial/ethnic categories across all US public health data bases and US census
  - routinely geocode all US public health data bases and employ standard area-based socioeconomic measure to:
    - ROUTINELY monitor socioeconomic disparities in health within racial/ethnic groups
    - ROUTINELY assess the contribution of socioeconomic inequality to racial/ethnic disparities in health

REFERENCES:
PUBLIC HEALTH DISPARITIES GEOCODING PROJECT

•Krieger N, Waterman P, Lemieux K, Zierler S, Hogan JW. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health* 2001; 91:1114-1116.

•Krieger N, Waterman P, Chen JT, Soobader M-J, Subramanian SV, Carson R. ZIP Code caveat: bias due to spatiotemporal mismatches between ZIP Codes and US census-defined areas—the Public Health Disparities Geocoding Project. *Am J Public Health* 2002; 92:1100-1102.

•Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV, Carson R. Geocoding and monitoring US socioeconomic inequalities in mortality and cancer incidence: does choice of area-based measure and geographic level matter?—the Public Health Disparities Geocoding Project. *Am J Epidemiol* 2002; 156:471-482.

•Krieger N, Zierler S, Hogan JW, Waterman P, Chen J, Lemieux K, Gjelsvik A.Geocoding and measurement of neighborhood socioeconomic position. In: Kawachi I, Berkman LF (eds). *Neighborhoods and Health*. New York: Oxford University Press (forthcoming).

•Krieger N, Chen JT, Waterman PD, Soobader M-J, Subramanian SV, Carson R. Choosing area-based socioeconomic measures to monitor social inequalities in low birthweight and childhood lead poisoning—The Public Health Disparities Geocoding Project (US). *J Epidemiol Community Health* (in press).

•Krieger N, Waterman PD, Chen JT, Soobader M-J, Subramanian SV. Monitoring socioeconomic inequalities in sexually transmitted infections, tuberculosis, and violence: geocoding and choice of area-based socioeconomic measures—The Public Health Disparities Geocoding Project (US). *Public Health Reports* (in press).

ADDITIONAL REFERENCES:
EMPIRICAL INVESTIGATIONS

•Bassett M, Krieger N. Social class and black-white differences in breast cancer survival. Am J Public Health 1986; 76:1400-1403.

•Krieger N. Social class and the black/white crossover in age-specific incidence of breast cancer: a study linking census-derived data to population-based registry records. Am J Epidemiol 1990; 131:804-814.

•Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. J Epidemiol Community Health 1991; 45:35-42.

•Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. Am J Public Health 1992; 82:703-710.

•Ellen JM, Kohn RP, Bolan GA, Shiboski S, Krieger N. Socioeconomic differences in sexually transmitted disease rates among black and white adolescents, San Francisco, 1990 to 1992. Am J Public Health 1995; 85:1546-1548.

•Krieger N, Quesenberry C Jr, Peng T, Horn-Ross P, Stewart S, Brown S, Swallen K, Guillermo T, Alvarez-Martinez L, Ward F. Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, black, Hispanic, and white residents of the San Francisco Bay Area, 1988-1992 (USA). Cancer Causes and Control 1999; 10:525-537.

•Zierler S, Krieger N, Tang Y, Coady W, Siegfried E, DeMaria A, Auerbach J. Economic deprivation and AIDS incidence in Massachusetts. Am J Public Health 2000; 90:1064-1073.

ADDITIONAL REFERENCES:
CONCEPTS & METHODS

•Krieger N, Rowley DL, Herman AA, Avery B, Phillips MT. Racism, sexism, and social class: implications for studies of health, disease, and well-being. Am J Prev Med 1993; 9 (Suppl):82-122.

•Krieger N, Fee E. Social class: the missing link in U.S. health data. Int J Health Services 1994; 24:25-44.

•Krieger N, Fee E. Measuring social inequalities in health in the United States: an historical review, 1900-1950. Int J Health Services 1996; 26:391-418.

•Krieger N, Williams D, Moss N. Measuring social class in US public health research: concepts, methodologies and guidelines. Annu Rev Public Health 1997; 18:341-378.

•Krieger N, Chen JT, Ebel G. Can we monitor socioeconomic inequalities in health? A survey of U.S. Health Departments' data collection and reporting practices. Public Health Reports 1997; 112:481-491.

•Krieger N. Discrimination and health. In: Berkman L, Kawachi I (eds). Social Epidemiology. Oxford: Oxford University Press, 2000; 36-75.

•Krieger N. Counting accountably: public health implications of new approaches to classifying race/ethnicity in the United States 2000 census—US and global perspectives. Am J Public Health 2000; 90:1687-1689.

•Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol 2001; 30:668-677.

•Krieger N. The ostrich, the albatross, and public health: an ecosocial perspective--or why an explicit focus on health consequences of discrimination and deprivation is vital for good science and public health practice. Public Health Reports 2001; 116:419-423.