THEORY AND METHODS

Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US)

N Krieger, J T Chen, P D Waterman, M-J Soobader, S V Subramanian, R Carson

....

See end of article for authors' affiliations

. . . .

Correspondence to:
Professor N Krieger, Department of Health and Social Behavior, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA; nkrieger@hsph.harvard.edu

Accepted for publication 14 June 2002

. . . .

Study objectives: To determine which area based socioeconomic measures can meaningfully be used, at which level of geography, to monitor socioeconomic inequalities in childhood health in the US.
Design: Cross sectional analysis of birth certificate and childhood lead poisoning registry data, geocoded and linked to diverse area based socioeconomic measures that were generated at three geographical levels: census tract, block group, and ZIP code.
Setting: Two US states: Massachusetts (1990 population=6 016 425) and Rhode Island (1990 population=1 003 464).
Participants: All births born to mothers ages 15 to 55 years old who were residents of either Massachusetts (1989–1991; n=267 311) or Rhode Island (1987–1993; n=96 138), and all children ages 1 to 5 years residing in Rhode Island who were screened for lead levels between 1994 and 1996 (n=62 514 children, restricted to first test during the study period).
Main results: Analyses of both the birth weight and lead data indicated that: (a) block group and tract socioeconomic measures performed similarly within and across both states, while ZIP code level measures tended to detect smaller effects; (b) measures pertaining to economic poverty detected stronger gradients than measures of education, occupation, and wealth; (c) results were similar for categories generated by quintiles and by a priori categorical cut off points; and (d) the area based socioeconomic measures yielded estimates of effect equal to or augmenting those detected, respectively, by individual level educational data for birth outcomes and by the area based housing measure recommended by the US government for monitoring childhood lead poisoning.
Conclusions: Census tract or block group area based socioeconomic measures of economic deprivation could be meaningfully used in conjunction with US public health surveillance systems to enable or enhance monitoring of social inequalities in health in the United States.

Despite the well known existence of socioeconomic inequalities in childhood health within the United States,[1] efforts to monitor trends in these disparities are hampered by two problems. The first is the scant data available on socioeconomic position in US public health surveillance systems.[2] US birth certificates, for example, contain only data on parents' educational level[2 3] and no socioeconomic data are included in public health data systems to monitor childhood lead poisoning.[4 5] Substantial research indicates, however, that not only is material deprivation an important determinant of both low birth weight and increased blood lead concentrations in children,[3–6] but also that different economic indicators may vary in their associations with these and other health outcomes.[7–11] The implication is that, without adequate socioeconomic data, US public health monitoring systems are compromised in their ability to track disparities and trends in childhood health.

One solution to the paucity or absence of socioeconomic data in US public health surveillance systems gaining increasing recognition involves geocoding and use of area based socioeconomic measures.[7 12 13] Of note, these area based measures can be conceptualised as meaningful indicators of socioeconomic context in their own right and not merely "proxies" for individual level data, providing information on not only the area's residents (its composition) but also area level characteristics not reducible to the individual level (for example, concentration of poverty, absence of a nearby clinic, adjacency to a toxic waste site, respectively understood as contextual,

public goods, and environmental attributes).[7 12–14] Especially relevant to children's health, these measures can likewise be equally applied, unlike individual level educational or occupational measures, to all persons, regardless of age, gender, and employment status.[7 12]

This potential solution, however, raises a second problem: a lack of consensus as to which kinds of area based socioeconomic measures, obtained at which level of geography, are meaningful for monitoring socioeconomic inequalities in health.[7 8 13 14] Recent US studies on low birth weight and related birth outcomes, for example, have used an eclectic array of single indicator area based measures (for example, poverty rate, per capita income, percentage unemployed, percentage of adults with less than a high school education, crowding, and ratio of home owners to renters[15–20]) plus assorted indices (for example, summed z scores combining data on median family income and education[16]), all variously measured at the level of the census block group,[18] census tract,[15 17] ZIP code,[20] and larger "community areas" defined by local health departments.[16 19] Similar inconsistencies regarding type and geographical level of area based socioeconomic measures are evident in studies on population screening of children's blood lead concentrations.[5 11 21 22] Although a plurality of measures may be useful for aetiological research, in the case of monitoring, such heterogeneity impedes comparing results across studies and across outcomes, let alone tracking changes over time.

Table 1 Study population: people and areas and percent of cases geocoded, by level of geography: births in Massachusetts (1989–1991; n=267,273) and Rhode Island (1987–1993; n=94,175), and childhood lead screening* in Rhode Island (1994–1996; n=62514)

| | N (%) | | % geocoded to: | | | | | | | |
| | | | BG | | CT | | ZC | | not geocoded | |
Data source: people	MA	RI	MA	RI	MA	RI	MA	RI	MA	RI
Birth:	267273 (100.0)	94175 (100.0)	94.7	97.4	100.0	100.0	100.0	99.9	0.0	0.0
Gender of infant:										
Boy	137208 (51.3)	48258 (51.2)	94.6	97.3	100.0	100.0	100.0	99.9	0.0	0.0
Girl	130282 (48.7)	45917 (48.8)	94.8	97.4	100.0	100.0	100.0	99.9	0.0	0.0
Age of mother:										
15–19	17578 (6.6)	9588 (10.2)	95.1	97.8	100.0	100.0	100.0	99.9	0.0	0.0
20–24	50363 (18.8)	22482 (23.9)	94.5	97.7	100.0	100.0	100.0	99.9	0.0	0.0
24–29	83356 (31.2)	30647 (32.5)	94.3	97.2	100.0	100.0	100.0	99.9	0.0	0.0
30–34	78198 (29.3)	22690 (24.1)	95.0	97.1	100.0	100.0	100.0	99.9	0.0	0.0
35–39	32119 (12.0)	7719 (8.2)	95.0	97.2	100.0	100.0	100.0	99.8	0.0	0.0
40–44	5485 (2.1)	1003 (1.1)	95.0	97.2	100.0	100.0	100.0	99.9	0.0	0.0
45–55	174 (0.1)	46 (0.0)	93.6	100.0	100.0	100.0	100.0	100.0	0.0	0.0
Mother's race/ethnicity†:										
White, non-Hispanic	209023 (78.5)	38751 (76.2)	94.1	98.8	100.0	100.0	100.0	99.8	0.0	0.0
Black, non-Hispanic	20375 (7.7)	3904 (7.7)	96.9	99.0	100.0	100.0	100.0	100.0	0.0	0.0
Other, non-Hispanic	27836 (10.5)	2260 (4.4)	96.8	98.9	100.0	100.0	100.0	99.9	0.0	0.0
Hispanic	9030 (3.4)	5924 (11.7)	96.0	99.0	100.0	100.0	100.0	100.0	0.0	0.0
Mother's educational level (highest attained)†:										
< high school	44632 (16.8)	18515 (20.0)	95.8	97.9	100.0	100.0	100.0	99.9	0.0	0.0
high school	85031 (32.0)	32489 (35.1)	94.2	97.4	100.0	100.0	100.0	99.9	0.0	0.0
> high school, <4 y college	63281 (23.8)	17587 (19.0)	94.1	97.0	100.0	100.0	100.0	99.9	0.0	0.0
4+ y college	73004 (27.4)	23955 (25.9)	95.1	97.2	100.0	100.0	100.0	99.9	0.0	0.0
Plurality:										
singleton	239,564 (89.6)	84995 (90.3)	94.8	97.4	100.0	100.0	100.0	99.9	0.0	0.0
multiple	27,709 (10.4)	9810 (9.7)	93.9	96.5	100.0	100.0	100.0	100.0	0.0	0.0
Birth weight†:										
<1500 g	2834 (1.1)	1090 (1.2)	95.5	97.4	100.0	100.0	100.0	99.9	0.0	0.0
1500–2499 g	12603 (4.7)	4798 (5.1)	95.2	97.7	100.0	100.0	100.0	99.9	0.0	0.0
2500–3499 g	134605 (50.4)	49081 (52.1)	94.8	97.5	100.0	100.0	100.0	99.9	0.0	0.0
3500–4499 g	111637 (41.8)	37522 (39.8)	94.5	97.1	100.0	100.0	100.0	99.9	0.0	0.0
≥4500 g	5594 (2.1)	1684 (1.8)	94.2	97.6	100.0	100.0	100.0	99.8	0.0	0.0
low birth weight (<2500 g)	15437 (5.8)	5888 (6.3)	95.2	97.7	100.0	100.0	100.0	99.9	0.0	0.0
mean (SD)	3394.9 (591.6)	3363.4 (594.5)								

| | N (%) | | BG | | CT | | ZC | | not geocoded | |
Data source: people	venous	capillary	venous	capillary	venous	capillary	venous	capillary	venous	capillary
Lead: RI only (n)	31076 (100.0)	31438 (100.0)	92.9	94.2	93.6	95.1	93.5	94.9	6.4	4.9
Age:										
1 to <2 y	12238 (39.4)	13440 (42.8)	93.7	94.2	94.4	95.3	94.2	95.1	5.6	4.7
2 to <3 y	6793 (21.9)	6467 (20.6)	92.8	94.7	93.6	95.4	93.5	95.2	6.4	4.6
3 to <4 y	5863 (18.9)	5601 (17.8)	93.6	94.3	93.3	95.0	93.1	94.8	6.7	5.0
4 to <5 y	6182 (19.9)	5930 (18.9)	91.6	93.5	92.3	94.5	92.2	94.3	7.7	5.5
Gender:										
Girls	14841 (47.8)	15093 (48.0)	94.1	94.2	94.8	95.2	94.7	95.0	5.2	4.8
Boys	15460 (49.6)	15790 (50.2)	94.1	94.2	95.0	95.1	94.8	94.8	5.1	4.9
unknown	775 (2.5)	555 (1.8)	43.9	93.9	44.4	95.3	44.1	95.0	55.6	4.7
Race/ethnicity‡:										
White	14822 (47.7)	21039 (66.9)	93.4	93.7	94.5	94.7	94.2	94.5	5.5	5.3
Black	2117 (6.8)	1785 (5.7)	97.5	96.8	98.1	97.3	97.8	96.9	1.9	2.7
Hispanic	5538 (17.8)	3156 (10.0)	97.9	98.1	98.1	98.2	98.1	98.1	1.9	1.8
Asian	911 (2.9)	393 (1.3)	99.1	98.7	99.1	99.0	99.1	98.5	0.9	1.0
American Indian	38 (0.1)	84 (0.3)	100.0	84.5	100.0	90.5	97.4	90.5	0.0	9.5
Other	288 (0.9)	572 (1.8)	97.9	92.3	98.6	94.8	98.6	93.4	1.4	5.2
Unknown	7362 (23.7)	4409 (14.0)	85.5	92.7	86.4	93.7	86.3	93.6	13.6	6.3

Table 1 continued

Data source: people	N (%)		% geocoded to:							
			BG		CT		ZC		not geocoded	
	MA	RI	MA	RI	MA	RI	MA	RI	MA	RI
Blood lead level (µg/dl):										
<2.0	3406 (11.0)	1735 (5.5)	93.8	94.5	94.6	95.9	94.4	95.7	5.4	4.1
2.0–4.9	11643 (37.5)	10390 (33.1)	93.8	94.1	94.7	95.1	94.6	94.8	5.3	4.9
5.0–9.9	10678 (34.4)	12444 (39.6)	94.0	93.9	94.7	94.8	94.5	94.6	5.3	5.2
10.0–14.9	3105 (10.0)	4040 (12.9)	93.2	94.5	93.7	95.4	93.6	95.1	6.3	4.6
15.0–19.9	1162 (3.7)	1484 (4.7)	90.4	95.0	90.8	95.3	90.8	95.3	9.2	4.7
≥20.0	1082 (3.5)	1345 (4.3)	70.0	95.8	70.4	96.4	70.3	96.4	29.6	3.6
Increased lead (≥10 µg/dl)	5349 (17.2)	6869 (21.9)	87.9	94.8	88.4	95.6	88.3	95.4	11.6	4.4
mean (SD)	6.2 (5.9)	7.2 (6.5)	5.9 (5.4)	7.2 (6.5)	5.9 (5.4)	7.2 (6.5)	5.9 (5.4)	7.2 (6.5)	9.5 (10.6)	6.9 (6.0)
median	5.0	5.0	5.0	6.0	5.0	5.0	5.0	5.0	5.0	5.0

Data source: areas		Population size				Population size		
	N	Mean (SD)	Range		N	Mean (SD)	Range	
Block groups	5603	1085.4 (665.2)	5 to 10096		897	1137.7 (670.8)	7 to 5652	
Census tracts	1331	4571.8 (2,080.0)	18 to 15411		235	4325.3 (1810.9)	26 to 9822	
zip codes	474	12719.7 (12,244.1)	14 to 65001		70	14335.2 (13234.8)	63 to 53763	

*Restricted to a child's first screening test during the study interval. †in MA, data were missing on mother's race/ethnicity for 0.4% and on mother's education for 0.5% of births; in RI, data on race/ethnicity were not included with the birth certificate data until 1990, so values for race/ethnicity refer only to births between 1990 and 1993 and do not include births from 1987 through 1989, while for mother's education, data were missing for 1.8% of RI births during the total study interval (1987–1993). ‡Race/ethnicity: Among venous cases, Hispanics includes 333 persons identified as Portuguese (0.6%); the category "Asian" included persons categorised as Vietnamese (n=33; 3.6%), Laotian (n=248; 27.2%), Cambodian (n=512; 56.2%), and Hmong (n=118; 13.0%) Among capillary cases, Hispanics included 399 persons identified as Portuguese (1.3%); the category "Asian" included persons categorised as Vietnamese (n=40; 0.1%), Laotian (n=155; 0.5%), Cambodian (n=170; 0.5%) and Hmong (n=28; 0.1%).

This study, concerning low birth weight and childhood lead poisoning, accordingly constitutes one component of our *Public Health Disparities Geocoding Project*, designed to investigate which area based socioeconomic measures, at which levels of geography, can meaningfully be used for monitoring socioeconomic disparities in US health, across a wide variety of health outcomes.[25] A priori criteria for assessing the area based socioeconomic measures under investigation were: (a) external validity (do the measures find gradients reported in the literature?), (b) robustness (do the measures detect expected gradients across a wide range of outcomes?), (c) completeness (are the measures comparatively unaffected by missing data?), and (d) user friendliness (how easy are the measures to understand and explain?). Guided by both an ecosocial framework[23] and our previous empirical research,[12 13 24 25] we hypothesised that area based measures reflecting economic deprivation would exhibit the most pronounced socioeconomic gradients in health, compared with measures of affluence, with effects more consistent at the block group and census tract, as compared with ZIP code, level.

METHODS
Data sources
The study base consisted of the populations and areas in two US states, Massachusetts (MA) and Rhode Island (RI), for the calendar period surrounding the 1990 census[26 27] (table 1). Birth certificate data were provided by both the Massachusetts Department of Public Health (MDPH) and the Rhode Island Department of Health (RIDOH); childhood lead screening data were provided by RIDOH only. Use of these data was approved by all relevant Institutional Review Boards/Human Subjects Committees at the Harvard School of Public Health, MDPH, and RIDOH, with the study conforming to the principles embodied in the Declaration of Helsinki. Geocoding of these databases to the block group, census tract, and ZIP code levels was conducted by a commercial geocoding firm selected for its accuracy (determined to be 95% in a pilot study conducted for this project).[28]

Birth certificate data were obtained for all births to mothers who were residents of either Massachusetts (1989–1991; n=267 531) or Rhode Island (1987–1993; n=97 383) at the time of birth. Analogous to government reports, this excluded (for MA and RI combined) births to mothers under age 15 or older than age 55, and we also excluded births with missing birth weight (n=212) or weighing <150 g (n=970), multiple births including five or more births (n=324), and births that were not geocoded (n=2001); the final analytical dataset thus included 267 311 MA births and 96 138 RI births. Following standard conventions, low birth weight was defined as births <2500 g.[1 3] Mother's race/ethnicity and educational level were obtained by self report on the birth certificate, using closed format questions.

Data on blood lead concentrations from the Rhode Island lead surveillance system, mandated in January 1993,[21] were obtained for all tests (n=135 567) performed on children aged 1 to 5 years who were residents of Rhode Island and screened between 1994 and 1996. The final analytic dataset included 62 514 cases, after excluding cases: geocoded to outside of RI (n=20); under 12 or over 60 months old (n=30 372); missing lead data (n=1741); and repeat tests (n=40 920), removed to minimise the likelihood of artificially attenuating relations between socioeconomic position and risk of increased blood lead concentrations due to treatment after initial high values. Blood specimens were obtained two ways, at the screening physician's discretion: venous and capillary/fingerstick. Because the second method may be subject to contamination (for example, lead dust on the pricked finger),[4 21] we analysed the two types of samples separately (venous: n=31 076; capillary/fingerstick =31 438). All blood lead specimens were analysed at the RIDOH Laboratory using the Graphite Furnace Atomic Absorption Spectrophotometry Method,[29] which has a minimum detection level of 1 µg/dl. The specimens were kept refrigerated until analysed. Following guidelines issued by the Centers for Disease Control and Prevention (CDC) in 1997,[4] raised blood lead concentrations were defined as ≥10 µg/dl.

Table 2 Area based socioeconomic measures: constructs and operational definitions, using 1990 US census data[30]

Construct	Operational definition	Census variable
(A) Occupational class		
(1) Working class[7]	–Percentage of persons employed in predominantly working class occupations, that is., as non-supervisory employees, operationalised as percentage of persons employed in the following 8 of 13 census based occupational groups: administrative support; sales; private household service; other service (except protective); precision production, craft, repair; machine operators, assemblers, inspectors; transportation and material moving; handlers, equipment cleaners, labourers.	P78
(2) Unemployment	–Percentage of persons age 16 and older in the labour force who are unemployed (and actively seeking work)	P71
(B) Income		
(3) Median household income	–Median household income in year prior to the decennial census (for US in 1989 = $30056)	P80A
(4) Low income[39]	–Percentage of households with income <50% of the US median household income (that is., <$15000)	P80
(5) High income	–Percentage of households with incomes ≥400% of the US median household income (that is, ≥$150000)	P80
(6) Gini coefficient[40]	–A measure of income inequality, regarding the share of income distribution across the population, calculated using the standard algorithm used by the US Bureau of Census to extrapolate the lower and upper ends of the income distribution	P80, P80A, P81
(C) Poverty		
(7) Below poverty[30,47]	–Percentage of persons below federally defined poverty line, a threshold which varies by size and age composition of the household, and on average equalled $12647 for a family of 4 in 1989[30]	P117
(D) Wealth		
(8) Expensive homes	–Percentage of owner occupied homes worth ≥$300000 (400% of the median value of owned homes in 1989)	H61
(E) Education		
(9) Low: < high school	–Percentage of persons, age 25 and older, with less than a 12th grade education	P57
(10) High: ≥4 y of college	–Percentage of persons, age 25 and older, with at least 4 years of college	P57
(F) Crowding		
(11) Crowded households	–Percentage of households with ≥1 person per room	H69, H49
(G) Composite measures		
(12) Townsend index[38,41–42]	–UK deprivation measure consisting of a standardised z score combining data on percentage crowding, percentage unemployment, percentage no car ownership, and percentage renters	H69, H49, H40, H8
(13) Carstairs index[37–38,43]	–UK deprivation measure consisting of a standardised z score combining data on percentage crowding, percentage male unemployment, percentage no car ownership, and percentage low social class (US census categories for: transportation and material moving; handlers, equipment cleaners, and laborers; household service).	H69, H49, H40, P78
(14) Index of Local Economic Resources[44]	–A "summary index" based on: "white collar employment, unemployment, and family income"	P78, P71, P107A
(15) SEP1	–A composite categorical variable based on percentage < poverty, working class, and expensive homes	{see above}
(16) SEP2	–A composite categorical variable based on percentage < poverty, working class, and high income	{see above}
(17) factor 1*	–A factor pertaining to economic resources; highly correlated with poverty, median household income, home ownership, and car ownership	{see above}
(18) factor 2*	–A factor pertaining to occupation and education; highly correlated with percentage working class, < high school, and ≥4 y college	{see above}
(19) SEP index	–A summary deprivation measure consisting of a standardised z score combining data on percentage working class, unemployed, < poverty, < high school, expensive homes, and median household income†	{see above}

*Variables used in the factor analysis: percentage working class, unemployed, < poverty, home ownership, car ownership, no telephone, expensive homes, < high school education, ≥ four years of college education, household crowding, households with only one room, no kitchen, no private plumbing, and also median household income and proportion of total income in the area derived from interest, dividends, and net rent. †Values for "expensive homes" and "median household income" were reversed before computing z score so that a higher score on the SEP index would correspond to a higher degree of deprivation.

We obtained 1990 census data for census tracts and block groups from US Bureau of Census Summary Tape File 3A and ZIP code data from Summary Tape File 3B.[30] The block group, a subdivision of the census tract, is the smallest geographical census unit for which census socioeconomic data are tabulated and on average contains 1000 people[31] (page G-6). A census tract, in turn, on average contains 4000 persons and is defined by the US Bureau of Census to be "small, relatively permanent statistical subdivision of a county … designed to be relatively homogeneous with respect to population characteristics, economic status, and living conditions".[31] (pages G-10, G-11). ZIP codes, by contrast, on average contain 30 000 persons and are "administrative units established by the United States Postal Service . . . for the most efficient delivery of mail, and therefore generally do not respect political or census statistical area boundaries".[30] (page A-13)

Three considerations guided our development of area based measures of socioeconomic position (SEP): (a) a priori conceptual definitions of SEP and social class[7]; (b) US and UK evidence emphasising detrimental effects of material deprivation on health[1 32–36]; and (c), the need for measures that can be meaningfully compared over time and space, so as to permit valid monitoring and contrasts in relation to time period and region.[7 13 25 37 38] As shown in table 2, the 11 single variable and eight composite area based socioeconomic measures we generated meeting these criteria, at each level of geography for each state, reflected six domains of SEP [7 8 39 40]: occupational class, income, poverty, wealth, education, and crowding,

premised on the understanding that social class, as a social relationship, fundamentally drives the distribution of these manifest aspects of SEP.[7][25] Additionally, in accordance with CDC guidelines for targeting childhood lead screening,[4] we also created a variable pertaining to the percentage of housing units built prior to 1950.

Among the composite variables, two were US analogues of the UK Townsend[38][41][42] and Carstairs[37][43] deprivation indices, one used the algorithm for the US Center for Disease Control and Prevention's "Index of Local Economic Resources",[44] and five were created exclusively for our study.[13][25] To mirror the skewed population distribution of socioeconomic resources, "SEP1" and "SEP2" simultaneously combined categorical data on poverty, working class, and either wealth or high income. "Factor 1" and "Factor 2" were generated by factor analysis using a maximum likelihood approach[44][45] applied to 15 census variables (see footnote to table 2), using the rank values of the census data, rather than impose arbitrary transformations to normalise their often considerably skewed distribution, with tied values assigned an average rank. The two factor model was selected as the most appropriate description of the underlying factor structure, with correlations between the factors ranging from 0.420 to 0.564 after oblique rotation. Finally, the "SEP index", a standardised z scores akin to the Townsend index, was generated using inputs identified by the factor analysis. Cut off points for categorical area based socioeconomic measures were based on both their centile distribution (for example, quintiles) and a priori considerations (for example, the federal definition of "poverty areas" as regions where ≥20% of the population is below the US poverty line[47][48]).

Data analysis

Our analytic plan involved five steps. In Step (1), we assessed the distribution and missingness of both the health and census data. In Step (2), we calculated the proportion of births that were low birth weight and the proportion of children screened who had high lead concentrations, stratified by the area based socioeconomic measures at each level of geography for each state. In Step (3), we visually inspected and quantified the socioeconomic gradients detected for each outcome at each level of geography using each area based socioeconomic measure, comparing outcomes for infants and children residing in areas with the least and most resources. As measures of effect, we calculated both the odds ratio (OR) and the relative index of inequality (RII). The RII provides a slope estimate of the risk estimate (for example, OR) 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.[49–51] For the birthweight analyses in Step (3), we analysed both data for all births and also for singleton births only, given the growing proportion of multiple births (at increased risk for being low birth weight, regardless of SEP) among older and more affluent women who became pregnant using in vitro fertilisation techniques.[52] Because both sets of analyses yielded similar results, we report only data for singleton births (data not shown available upon request). In Step (4), we restricted analyses to persons geocoded to all three levels of geography; because results were equivalent to those obtained in Step (3), we report only the former, and do not include equivalent variables. For example, given similar results, we present data for only the categorical and not quintile version of the poverty data and omit the variable for low income (data not shown available upon request).

In Step (5), we then summarised findings across socioeconomic measures within and across levels of geography, in relation to our a priori criteria pertaining to external validity, robustness, completeness, and user friendliness. As an additional check on external validity, we also analysed the birth data in relation to mother's educational level and restricted these analyses, to be compatible with US government reports, to mothers aged 20 and older.[1] Also relevant to standard public health practice, we ascertained patterns of lead poisoning in relation to both the CDC's recommended area based screening characteristic (≥27% of housing built before 1950[4]) and the poverty measure, separately and combined. All analyses were conducted in SAS.[53]

RESULTS

Low birth weight (<2500 g) births comprised 5.8% and 6.3% of births, respectively, in Massachusetts and Rhode Island during the study interval (table 1). Increased blood lead concentrations (≥10 μg/dl) were detected among 17% and 22% of Rhode Island children ages 1 to 5 years old who underwent, respectively, venous and capillary/fingerstick tests (table 1).

Geocoding of births was successfully accomplished to the block group level for over 95% of records in Massachusetts and Rhode Island, to the census tract level for 100% of records for both states, and to the ZIP code level for 100% of the Massachusetts and 99.9% of the Rhode Island records (table 1). These results were independent of mother's age, race/ethnicity, and educational level, and also birth weight of the infant. For lead screening, the overall percentage of children geocoded did not vary by age or gender and equalled 94%–95% for the venous and capillary tests; modest variation, however, occurred by race/ethnicity, type of lead test, and lead concentration (table 1). Among records geocoded to the ZIP code level, 8.1% of Massachusetts births, 2.5% of Rhode Island births, and 1.9% of the childhood lead poisoning records could not be linked to 1990 census data because their ZIP codes were either for non-residential sites or else were created or changed after the 1990 census. Less than 1% of areas were missing data on the specified area based socioeconomic measures; the one exception was for measures with data on wealth (affecting only 2% to 4% of areas; data not shown and available upon request).

As table 3 illustrates, in the case of low birth weight, both choice of area based measure and level of geography mattered. Firstly, considering singleton births in Massachusetts (table 3), the strongest gradients (OR ≥ 2.0; RII ≥ 2.5) were observed for census tract and block group measures of economic deprivation (poverty, median household income, Townsend index, Index of Local Economic Resources, SEP index and SEP1); the weakest gradient occurred for wealth (OR ~1.3–1.4; RII ~1.8). Effect estimates detected with ZIP code measures exhibited similar patterns but slightly lesser magnitudes. Equivalent patterns were evident for Rhode Island (table 3). In both states, moreover, these effect estimates were similar to those based on individual level educational data, comparing births to mothers with less than a high school education versus college graduates (MA: OR=1.90, 95% CI 1.80 to 2.01; RI: OR=2.03, 95% CI 1.86 to 2.22).

Similar patterns occurred for the lead data, with census tract and block group measures of economic deprivation detecting socioeconomic gradients either captured to a lesser extent or not at all by the other measures (table 4). Of note, socioeconomic gradients were steeper for the venous compared with capillary specimens. This occurred because a higher proportion of children living in areas with more resources were identified as having high lead concentrations by the capillary compared with the venous test (possibly reflecting greater contamination of the capillary samples), whereas for children in poorer areas, both tests identified a similarly high proportion of children having raised lead concentrations.

Summarising key aspects of these analyses, figures 1A–1C and 2A–2C depict, respectively, socioeconomic gradients for singleton births in Massachusetts and for increased blood lead

Table 3 Low birthweight (LBW) singleton births among mothers 15 to 55 years old, stratified by the census block group (BG), census tract (CT), and zip code (ZC) area based socioeconomic measures for births in areas with the least and most resources: proportion, odds ratio (OR), and relative index of inequality (RII) for Massachusetts (1989–1991; n=239564) and Rhode Island (1987–1993; n=84995)*

Health outcome	Area based socioeconomic measure	Proportion %: least resources			Proportion %: most resources			OR: least/most (95% CI)			RII (95% CI)		
		BG	CT	ZC	BG	CT	ZC	BG	CT	ZC	BG	CT	ZC
% LBW (<2500 g) singleton: MA	Working class (categorical)	6.3	6.4	5.6	3.5	3.5	3.6	1.85 (1.73 to 1.98)	1.87 (1.74 to 2.01)	1.35 (1.23 to 1.48)	2.13 (1.98 to 2.29)	2.11 (1.97 to 2.26)	1.91 (1.77 to 2.05)
	Unemployed (quintile)	6.5	6.8	6.2	3.9	3.5	3.5	1.72 (1.61 to 1.84)	1.99 (1.86 to 2.12)	1.58 (1.46 to 1.70)	2.09 (1.95 to 2.24)	2.37 (2.21 to 2.53)	2.19 (2.04 to 2.36)
	Median household income (quintile)	6.6	6.7	6.3	3.3	3.3	3.3	2.06 (1.93 to 2.19)	2.07 (1.94 to 2.20)	1.61 (1.52 to 1.71)	2.51 (2.34 to 2.69)	2.48 (2.32 to 2.66)	2.41 (2.24 to 2.59)
	Poverty (categorical)	7.4	7.5	7.6	3.7	3.6	3.7	2.08 (1.98 to 2.19)	2.16 (2.05 to 2.28)	1.90 (1.80 to 2.00)	2.69 (2.51 to 2.89)	2.74 (2.55 to 2.93)	2.61 (2.43 to 2.80)
	Gini (quintile)	6.1	6.5	6.2	3.9	4.0	3.7	1.59 (1.49 to 1.69)	1.66 (1.56 to 1.76)	1.57 (1.47 to 1.67)	1.80 (1.68 to 1.93)	1.88 (1.76 to 2.01)	2.07 (1.93 to 2.22)
	Wealth (categorical)	4.9	5.0	5.1	3.8	3.6	3.7	1.32 (1.23 to 1.42)	1.41 (1.31 to 1.51)	1.23 (1.15 to 1.31)	1.79 (1.62 to 1.98)	1.79 (1.63 to 1.96)	1.80 (1.64 to 1.98)
	Crowding (quintile)	6.7	7.1	6.3	3.9	3.7	3.5	1.76 (1.69 to 1.84)	2.01 (1.89 to 2.14)	1.59 (1.47 to 1.71)	2.42 (2.25 to 2.60)	2.55 (2.38 to 2.73)	2.39 (2.22 to 2.57)
	Low education (categorical)	6.9	6.9	6.2	3.6	3.6	3.5	1.97 (1.86 to 2.08)	1.99 (1.88 to 2.12)	1.61 (1.51 to 1.72)	2.48 (2.31 to 2.66)	2.43 (2.27 to 2.60)	2.33 (2.17 to 2.50)
	High education (categorical)	6.0	6.1	6.0	3.6	3.7	3.9	1.71 (1.61 to 1.82)	1.67 (1.57 to 1.78)	1.40 (1.32 to 1.50)	2.15 (2.00 to 2.31)	2.11 (1.97 to 2.26)	1.83 (1.71 to 1.97)
	Townsend index (quintile)	7.0	7.2	6.1	3.3	3.3	3.4	2.19 (2.05 to 2.34)	2.24 (2.10 to 2.39)	1.57 (1.46 to 1.68)	2.85 (2.65 to 3.06)	2.79 (2.60 to 2.99)	2.50 (2.32 to 2.69)
	Index of Local Economic Resources (quintile)	6.7	6.9	6.3	3.4	3.4	3.3	2.05 (1.93 to 2.19)	2.09 (1.97 to 2.22)	1.62 (1.53 to 1.72)	2.59 (2.42 to 2.78)	2.55 (2.39 to 2.73)	2.36 (2.20 to 2.54)
	Factor 1 (quintile)	6.6	6.9	6.1	3.3	3.4	3.3	2.03 (1.91 to 2.16)	2.11 (1.99 to 2.25)	1.59 (1.49 to 1.70)	2.61 (2.43 to 2.81)	2.65 (2.47 to 2.83)	2.49 (2.31 to 2.68)
	Factor 2 (quintile)	6.3	6.5	6.1	3.4	3.5	3.6	1.93 (1.81 to 2.06)	1.91 (1.79 to 2.04)	1.50 (1.40 to 1.59)	2.33 (2.17 to 2.51)	2.27 (2.12 to 2.43)	2.04 (1.90 to 2.19)
	SEP Index (quintile)	6.6	6.9	6.1	3.2	3.4	3.2	2.11 (1.98 to 2.26)	2.09 (1.96 to 2.23)	1.60 (1.51 to 1.71)	2.64 (2.45 to 2.83)	2.55 (2.38 to 2.73)	2.42 (2.25 to 2.61)
	SEP1 (categorical)	7.6	7.4	6.8	3.3	3.5	3.6	2.40 (2.19 to 2.64)	2.21 (2.03 to 2.41)	1.65 (1.47 to 1.85)	2.46 (2.28 to 2.65)	2.54 (2.36 to 2.73)	2.36 (2.18 to 2.57)
	Median value	6.6	6.9	6.2	3.5	3.5	3.6	1.97	2.01	1.59	2.46	2.48	2.36
% LBW (<2500 g) singleton: RI	Working class (categorical)	6.6	6.8	6.9	3.6	3.4	5.0	1.91 (1.65 to 2.22)	2.06 (1.71 to 2.48)	1.40 (1.15 to 1.69)	2.22 (1.98 to 2.49)	2.30 (2.06 to 2.58)	1.98 (1.77 to 2.21)
	Unemployed (quintile)	7.0	7.2	6.8	4.7	4.2	3.5	1.51 (1.37 to 1.67)	1.75 (1.59 to 1.93)	2.00 (1.77 to 2.27)	1.79 (1.61 to 1.99)	2.18 (1.95 to 2.43)	2.11 (1.89 to 2.35)
	Median household income (quintile)	7.5	7.5	6.7	3.5	3.7	3.6	2.28 (2.05 to 2.53)	2.10 (1.90 to 2.32)	1.93 (1.68 to 2.23)	2.71 (2.43 to 3.03)	2.63 (2.36 to 2.93)	2.35 (2.10 to 2.63)
	Poverty (categorical)	7.6	7.7	7.6	4.0	4.1	3.6	1.97 (1.82 to 2.13)	1.96 (1.81 to 2.14)	2.17 (1.95 to 2.43)	2.52 (2.26 to 2.81)	2.52 (2.26 to 2.81)	2.44 (2.18 to 2.72)
	Gini (quintile)	6.3	6.6	6.6	4.4	4.5	4.3	1.48 (1.34 to 1.64)	1.51 (1.37 to 1.66)	1.59 (1.33 to 1.89)	1.64 (1.47 to 1.83)	1.63 (1.47 to 1.81)	1.89 (1.69 to 2.11)
	Wealth (categorical)	5.6	5.7	5.7	3.0	3.7	4.5	1.93 (1.55 to 2.41)	1.55 (1.28 to 1.87)	1.27 (1.08 to 1.49)	2.06 (1.68 to 2.51)	2.17 (1.82 to 2.58)	1.75 (1.51 to 2.03)
	Crowding (quintile)	7.0	7.3	6.5	4.4	3.9	3.2	1.63 (1.52 to 1.74)	1.95 (1.76 to 2.16)	2.15 (1.76 to 2.62)	2.20 (1.97 to 2.47)	2.42 (2.17 to 2.70)	2.08 (1.86 to 2.33)
	Low education (categorical)	7.1	7.3	6.8	3.9	3.7	4.0	1.91 (1.73 to 2.10)	2.08 (1.85 to 2.33)	1.74 (1.56 to 1.94)	2.50 (2.24 to 2.80)	2.47 (2.21 to 2.76)	1.99 (1.78 to 2.22)
	High education (categorical)	6.3	6.4	6.3	3.8	3.8	4.2	1.70 (1.48 to 1.96)	1.75 (1.53 to 2.01)	1.53 (1.31 to 1.77)	2.11 (1.87 to 2.37)	2.26 (2.01 to 2.54)	1.75 (1.56 to 1.96)
	Townsend index (quintile)	7.2	7.5	6.6	3.8	3.7	3.3	1.98 (1.78 to 2.20)	2.11 (1.91 to 2.33)	2.05 (1.68 to 2.50)	2.57 (2.30 to 2.87)	2.71 (2.43 to 3.02)	2.42 (2.16 to 2.72)
	Index of Local Economic Resources (quintile)	7.4	7.4	6.8	3.7	3.7	3.9	2.07 (1.87 to 2.29)	2.09 (1.90 to 2.31)	1.80 (1.64 to 1.98)	2.46 (2.20 to 2.75)	2.53 (2.27 to 2.82)	2.26 (2.02 to 2.53)
	Factor 1 (quintile)	7.3	7.5	6.4	3.7	3.7	3.5	2.04 (1.85 to 2.26)	2.07 (1.87 to 2.30)	1.91 (1.61 to 2.26)	2.57 (2.30 to 2.87)	2.63 (2.36 to 2.93)	2.34 (2.08 to 2.63)
	Factor 2 (quintile)	7.0	7.1	6.3	3.7	4.2	4.0	1.96 (1.76 to 2.18)	1.76 (1.59 to 1.95)	1.61 (1.45 to 1.80)	2.40 (2.15 to 2.68)	2.22 (1.99 to 2.47)	1.87 (1.68 to 2.09)
	SEP Index (quintile)	7.5	7.5	6.7	3.5	3.5	4.0	2.21 (1.98 to 2.46)	2.21 (1.99 to 2.45)	1.74 (1.54 to 1.97)	2.71 (2.42 to 3.03)	2.65 (2.38 to 2.96)	2.23 (1.99 to 2.50)
	SEP1 (categorical)	7.8	7.9	7.7	3.3	3.3	5.1	2.51 (2.03 to 3.10)	2.49 (1.99 to 3.12)	1.55 (1.28 to 1.88)	2.42 (2.16 to 2.71)	2.48 (2.22 to 2.78)	2.04 (1.81 to 2.30)
	Median value	7.1	7.3	6.7	3.7	3.7	4.0	1.96	2.06	1.74	2.42	2.47	2.08

*See appendix for cut off points used for each variable.

Table 4 Increased blood lead concentrations (≥10 µg/dl), stratified by the census block group (BG), census tract (CT), and zip code (ZC) area-based socioeconomic measures for cases in areas with the least and most resources: proportion, odds ratio (OR), and relative index of inequality (RII), Rhode Island (1994–1996; n=62514)*

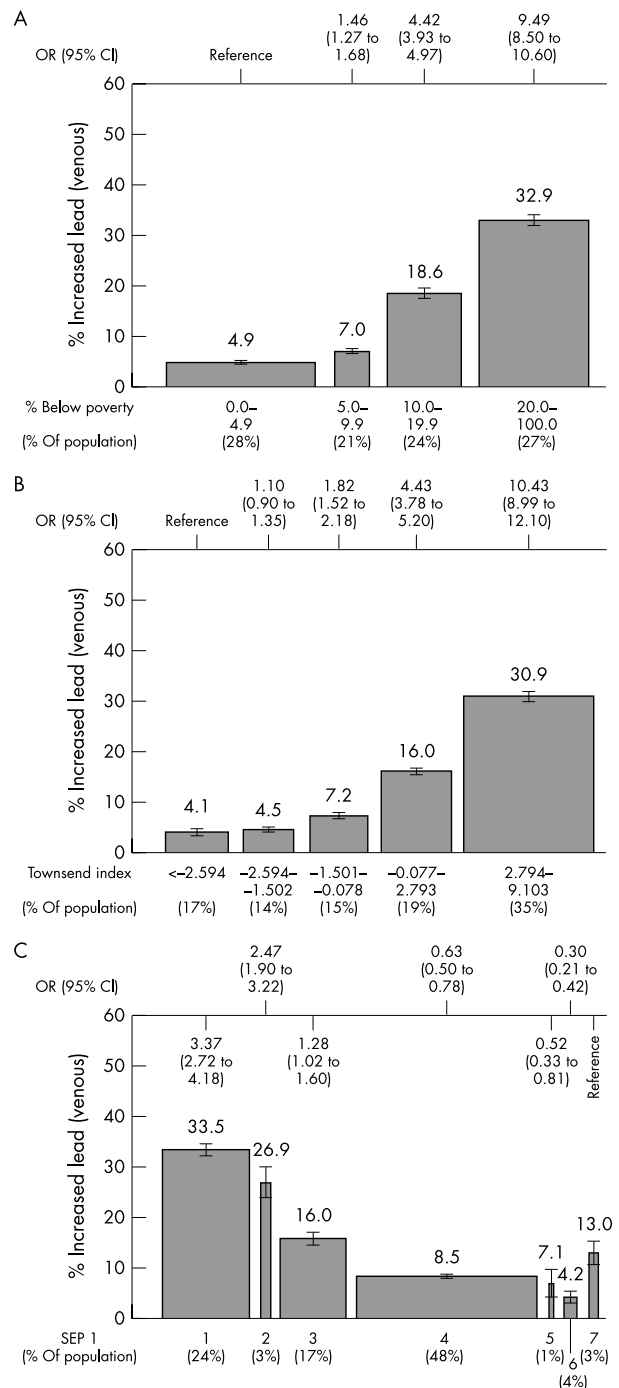| Health outcome	Area-based socioeconomic measure	Proportion %: least resources			Proportion %: most resources			OR: least/most (95% CI)			RII (95% CI)		
		BG	CT	ZC	BG	CT	ZC	BG	CT	ZC	BG	CT	ZC
% high lead (≥10 µg/dl): venous	Working class (categorical)	24.1	26.2	28.8	9.1	11.2	19.9	3.18 (2.70 to 3.75)	2.82 (2.34 to 3.41)	1.63 (1.37 to 1.95)	8.64 (7.58 to 9.86)	11.63 (10.18 to 13.28)	11.05 (9.72 to 12.57)
	Unemployed (quintile)	27.9	28.4	28.0	10.7	5.4	3.7	3.23 (2.90 to 3.60)	6.90 (6.03 to 7.89)	10.12 (8.23 to 12.44)	7.17 (6.35 to 8.09)	16.47 (14.42 to 18.80)	18.55 (16.14 to 21.32)
	Median household income (quintile)	29.5	30.5	26.4	4.5	3.5	3.4	8.87 (7.68 to 10.24)	12.19 (10.30 to 14.42)	10.29 (7.80 to 13.56)	20.33 (17.75 to 23.28)	28.32 (24.58 to 32.64)	31.93 (27.11 to 37.60)
	Poverty (categorical)	31.5	32.9	32.4	5.5	4.9	4.0	7.94 (7.17 to 8.80)	9.49 (8.50 to 10.60)	11.58 (9.88 to 13.58)	22.79 (19.90 to 26.09)	27.95 (24.36 to 32.07)	24.44 (21.35 to 27.97)
	Gini (quintile)	22.7	26.7	26.9	8.3	5.0	2.8	3.24 (2.87 to 3.67)	6.97 (5.99 to 8.12)	12.59 (8.30 to 19.09)	4.39 (3.90 to 4.93)	8.51 (7.53 to 9.61)	21.25 (18.33 to 24.65)
	Wealth (categorical)	16.9	17.7	18.0	8.7	13.6	15.6	2.13 (1.67 to 2.71)	1.37 (1.12 to 1.67)	1.19 (1.01 to 1.40)	5.06 (3.91 to 6.54)	6.77 (5.33 to 8.60)	4.46 (3.65 to 5.45)
	Crowding (quintile)	28.5	29.8	25.9	8.3	6.5	4.6	4.39 (4.07 to 4.73)	6.14 (5.37 to 7.03)	7.29 (5.36 to 9.93)	12.83 (11.28 to 14.59)	20.47 (17.82 to 23.51)	23.00 (19.71 to 26.85)
	Low education (categorical)	27.5	28.5	27.3	6.7	6.4	7.7	5.28 (4.63 to 6.03)	5.87 (4.97 to 6.94)	4.49 (3.89 to 5.18)	14.10 (12.34 to 16.12)	16.18 (14.14 to 18.51)	10.41 (9.15 to 11.84)
	High education (categorical)	21.6	22.5	23.4	9.3	10.8	13.2	2.70 (2.29 to 3.18)	2.39 (2.05 to 2.78)	2.00 (1.70 to 2.36)	6.39 (5.57 to 7.32)	9.16 (7.94 to 10.58)	5.49 (4.83 to 6.22)
	Townsend index (quintile)	29.4	30.9	26.0	3.7	4.1	2.8	10.90 (9.22 to 12.87)	10.43 (8.99 to 12.10)	12.27 (8.29 to 18.15)	26.85 (23.25 to 31.00)	33.91 (29.31 to 39.24)	41.14 (34.49 to 49.07)
	Index of Local Economic Resources (quintile)	29.7	30.5	27.1	5.4	5.6	6.6	7.46 (6.52 to 8.53)	7.36 (6.48 to 8.37)	5.26 (4.66 to 5.94)	16.92 (14.81 to 19.32)	23.62 (20.54 to 27.17)	22.84 (19.63 to 26.59)
	Factor 1 (quintile)	29.5	31.0	25.0	3.6	3.4	3.7	11.29 (9.64 to 13.23)	12.64 (10.64 to 15.01)	8.59 (6.39 to 11.56)	28.22 (24.39 to 32.65)	33.22 (28.76 to 38.38)	38.37 (32.01 to 46.00)
	Factor 2 (quintile)	26.8	27.4	23.6	6.7	9.8	7.7	5.15 (4.50 to 5.89)	3.46 (3.07 to 3.90)	3.70 (3.20 to 4.27)	10.60 (9.31 to 12.06)	8.82 (7.79 to 9.98)	7.70 (6.77 to 8.75)
	SEP Index (quintile)	30.0	30.0	26.4	6.0	4.8	9.4	6.69 (5.82 to 7.67)	8.46 (7.28 to 9.82)	3.46 (2.98 to 4.01)	25.49 (22.06 to 29.40)	24.67 (21.41 to 29.46)	22.72 (19.44 to 26.54)
	SEP1 (categorical)	31.9	33.5	32.4	8.1	13.0	20.0	5.31 (4.15 to 6.80)	3.37 (2.72 to 4.18)	1.92 (1.60 to 2.30)	12.28 (10.81 to 13.95)	16.06 (14.15 to 18.23)	10.78 (9.50 to 12.23)
	Median value	28.5	29.8	26.4	6.7	5.6	6.6	5.28	6.90	5.26	12.83	16.47	21.25
% high lead (≥10 µg/dl): capillary	Working class (categorical)	30.5	33.8	37.0	13.8	15.7	24.8	2.75 (2.45 to 3.09)	2.74 (2.39 to 3.14)	1.78 (1.52 to 2.09)	4.86 (4.37 to 5.41)	5.60 (5.04 to 6.22)	5.21 (4.68 to 5.79)
	Unemployed (quintile)	34.4	36.0	35.1	16.5	12.7	11.4	2.64 (2.40 to 2.90)	3.86 (3.51 to 4.25)	4.21 (3.78 to 4.69)	3.70 (3.35 to 4.10)	6.11 (5.51 to 6.77)	6.86 (6.17 to 7.63)
	Median household income (quintile)	35.6	37.4	33.5	10.7	11.0	10.5	4.62 (4.19 to 5.09)	4.83 (4.39 to 5.33)	4.29 (3.77 to 4.89)	7.17 (6.45 to 7.96)	8.03 (7.23 to 8.93)	7.21 (6.46 to 8.04)
	Poverty (categorical)	37.9	38.2	46.8	12.9	12.4	11.2	4.14 (3.83 to 4.47)	4.34 (3.99 to 4.72)	6.97 (6.24 to 7.79)	7.11 (6.40 to 7.90)	7.33 (6.60 to 8.13)	8.13 (7.30 to 9.07)
	Gini (quintile)	26.4	29.9	34.2	16.1	14.5	14.2	1.86 (1.69 to 2.06)	2.52 (2.28 to 2.78)	3.14 (2.62 to 3.75)	2.23 (2.02 to 2.46)	2.67 (2.42 to 2.94)	4.11 (3.70 to 4.57)
	Wealth (categorical)	23.6	24.4	25.8	14.4	16.5	20.8	1.84 (1.58 to 2.13)	1.63 (1.42 to 1.87)	1.32 (1.15 to 1.53)	3.53 (2.99 to 4.17)	3.82 (3.30 to 4.42)	3.58 (3.16 to 4.07)
	Crowding (quintile)	34.5	37.5	33.2	16.5	14.2	11.6	2.66 (2.50 to 2.84)	3.63 (3.31 to 3.96)	3.80 (3.26 to 4.43)	4.54 (4.09 to 5.05)	5.62 (5.07 to 6.23)	5.90 (5.30 to 6.57)
	Low education (categorical)	33.9	35.3	33.6	13.4	14.5	13.8	3.31 (3.03 to 3.61)	3.23 (2.94 to 3.55)	3.16 (2.88 to 3.46)	5.72 (5.15 to 6.35)	5.62 (5.06 to 6.24)	4.58 (4.13 to 5.09)
	High education (categorical)	28.8	29.7	30.5	13.2	15.8	15.0	2.66 (2.38 to 2.98)	2.25 (2.02 to 2.50)	2.49 (2.18 to 2.84)	4.49 (4.02 to 5.01)	4.61 (4.13 to 5.14)	3.82 (3.43 to 4.25)
	Townsend index (quintile)	36.6	37.8	32.1	11.2	10.0	7.5	4.57 (4.13 to 5.06)	5.48 (4.95 to 6.07)	5.83 (4.81 to 7.05)	8.18 (7.35 to 9.10)	9.06 (8.14 to 10.07)	7.42 (6.63 to 8.30)
	Index of Local Economic Resources (quintile)	36.4	37.7	34.5	11.4	12.4	13.4	4.45 (4.04 to 4.89)	4.29 (3.92 to 4.69)	3.41 (3.14 to 3.70)	6.86 (6.18 to 7.61)	7.25 (6.53 to 8.05)	6.45 (5.79 to 7.18)
	Factor 1 (quintile)	36.1	37.3	31.7	9.4	9.4	8.4	5.44 (4.91 to 6.03)	5.70 (5.12 to 6.35)	5.10 (4.30 to 6.06)	8.59 (7.72 to 9.57)	8.93 (8.03 to 9.94)	7.42 (6.62 to 8.31)
	Factor 2 (quintile)	33.3	35.1	30.9	13.4	15.8	13.8	3.24 (2.97 to 3.54)	2.88 (2.65 to 3.13)	2.79 (2.55 to 3.04)	5.45 (4.91 to 6.05)	4.98 (4.50 to 5.52)	4.38 (3.95 to 4.87)
	SEP Index (quintile)	36.7	37.4	33.6	12.1	12.7	14.2	4.23 (3.86 to 4.63)	4.10 (3.75 to 4.48)	3.06 (2.77 to 3.37)	7.97 (7.16 to 8.87)	7.46 (6.72 to 8.29)	5.86 (5.26 to 6.52)
	SEP1 (categorical)	38.0	41.2	46.9	12.2	15.8	24.9	4.40 (3.73 to 5.19)	3.74 (3.17 to 4.42)	2.66 (2.25 to 3.15)	5.93 (5.33 to 6.59)	6.03 (5.43 to 6.71)	5.93 (5.28 to 6.66)
	Median value	34.5	37.3	33.6	13.2	14.2	13.8	3.31	3.74	3.16	5.72	6.03	5.90

*See appendix for cut off points employed for each variable.
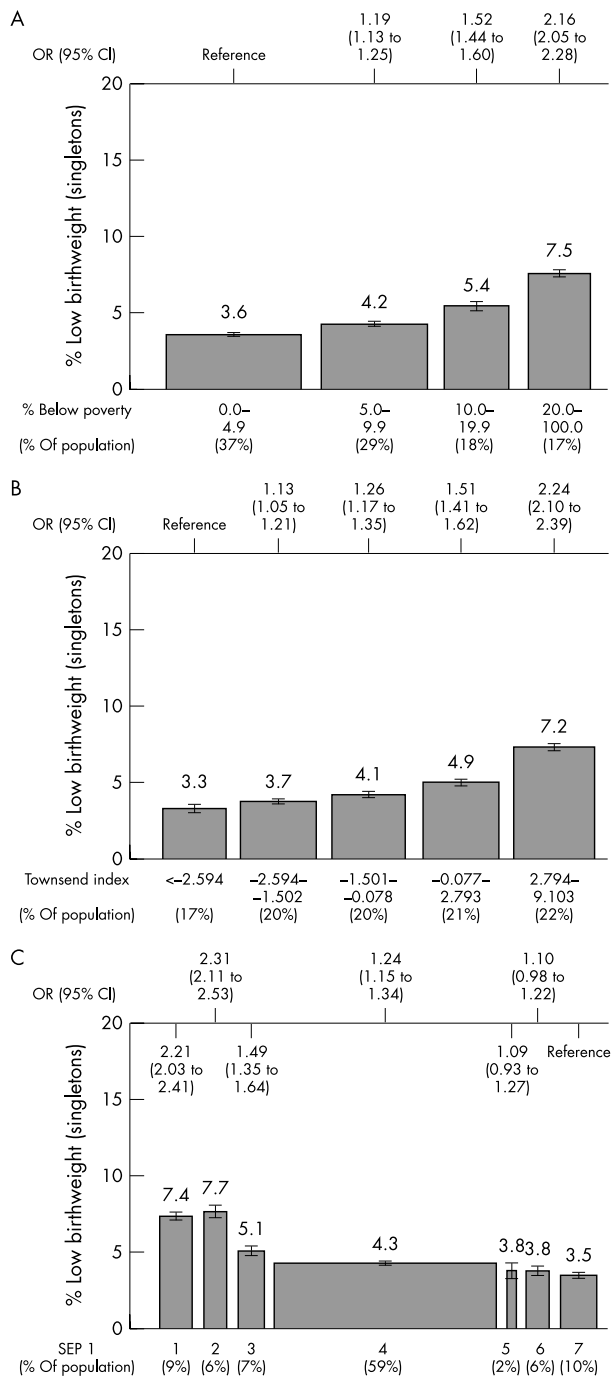
Figure 1 Proportion of low birthweight births (singleton births), stratified by the census tract measures of: (a) poverty, (b) Townsend index, and (c) SEP1, Massachusetts, 1989–1991.

Figure 2 Proportion of children with increased lead concentrations (venous; ≥10 µg/dl), stratified by the census tract measures of: (a) poverty, (b) Townsend index, and (c) SEP1, Rhode Island, 1994–1996.

concentrations (venous) in Rhode Island using the census tract versions of the three measures that most consistently detected socioeconomic gradients in health while differently delimiting the population at risk: poverty (single variable, categorical), Townsend index (composite, quintile), and SEP1 (composite, categorical).

Finally, in light of CDC recommendations that lead screening be targeted to areas where ≥27% of housing was built before 1950,[4][22] table 5 provides data on the odds of children having increased blood lead concentrations in relation to both this housing measure and the poverty measure, for all three levels of geography. Of note, both the poverty and housing measures detected populations at increased risk, and their combined effect was at least or more than additive (OR for poor area with old housing: 11 to 14 for venous, and 5 to 6 for capillary).

DISCUSSION

This study, a component of the first systematic US investigation evaluating diverse area based socioeconomic measures within and across multiple levels of geography for outcomes spanning from birth to death,[25] provides evidence that, in the case of childhood health, both choice of measure and level of geography matter. Specifically, analysing data pertaining to

Table 5 Poverty and old housing in relation to risk of elevated blood lead level (≥10 μg/dl) among children 1–5 years old, RI, 1994–1996 (venous n=31076; capillary n=31438)

Blood sample	Area based characteristic	BG Children screened N (%)	BG % elevated lead	BG OR (95% CI)	CT Children screened N (%)	CT % Increased lead	CT OR (95% CI)	ZC Children screened N (%)	ZC % Increased lead	ZC OR (95% CI)
Venous	Poverty:									
	Poor (≥20% < poverty)	8690 (30.1)	31.5	4.25 (3.99 to 4.54)	7940 (27.3)	32.9	4.41 (4.14 to 4.71)	6496 (24.3)	32.4	3.75 (3.51 to 4.00)
	Not poor (<20% < poverty)	20163 (69.9)	9.7	1.00	21152 (72.7)	10.0	1.00	21617 (75.7)	11.3	1.00
	Housing:									
	Old (≥27% pre-1950 housing)	21578 (74.8)	20.4	6.02 (5.34 to 6.80)	22892 (79.7)	3.5	6.70 (5.83 to 7.69)	25012 (87.6)	18.2	5.01 (4.25 to 5.92)
	Not old (<27% pre-1950 housing)	7275 (25.2)	4.1	1.00	6200 (21.3)	0.0	1.00	3551 (12.4)	4.3	1.00
	Poverty and housing:									
	Poor AND old housing	8436 (29.2)	31.9	12.67 (11.07 to 14.49)	7850 (27.0)	12.7	14.19 (12.25 to 16.43)	6946 (24.3)	32.4	10.80 (9.11 to 12.81)
	Poor AND NOT old housing	254 (0.9)	18.5	6.15 (4.37 to 8.65)	90 (0.3)	3.4	4.84 (2.65 to 8.85)	0 (0.0)	–*	–*
	NOT poor AND old housing	13142 (45.5)	13.0	4.06 (3.55 to 4.66)	15042 (51.7)	14.4	4.17 (3.60 to 4.83)	18066 (63.2)	12.7	3.29 (2.78 to 3.89)
	NOT poor and NOT old housing	7021 (24.3)	3.6	1.00	6110 (21.0)	0.0	1.00	3551 (12.4)	4.3	1.00
Capillary	Poverty:									
	Poor (≥20% < poverty)	5748 (19.4)	37.9	2.75 (2.58 to 2.93)	4859 (16.2)	38.2	2.66 (2.49 to 2.85)	2953 (10.2)	46.8	3.65 (3.37 to 3.94)
	Not poor (<20% < poverty)	23860 (80.6)	18.2	1.00	25044 (83.8)	18.8	1.00	26122 (89.8)	19.5	1.00
	Housing:									
	Old (≥27% pre-1950 housing)	20789 (70.2)	26.7	3.01 (2.80 to 3.24)	21914 (73.3)	25.9	2.82 (2.62 to 3.05)	24896 (85.6)	23.8	2.13 (1.94 to 2.34)
	Not old (<27% pre-1950 housing)	8819 (29.8)	10.8	1.00	7989 (26.7)	11.0	1.00	4179 (14.4)	12.8	1.00
	Poverty and housing:									
	Poor AND old housing	5584 (18.9)	38.4	5.25 (4.81 to 5.73)	4799 (16.0)	38.4	5.06 (4.62 to 5.55)	2953 (10.2)	46.8	6.00 (5.34 to 6.74)
	Poor AND NOT old housing	164 (0.6)	22.6	2.46 (1.69 to 3.57)	60 (0.2)	21.7	2.25 (1.21 to 4.18)	0 (0.0)	–*	–*
	NOT poor AND old housing	15205 (51.4)	22.5	2.45 (2.26 to 2.65)	17115 (57.2)	22.4	2.35 (2.17 to 2.55)	21943 (75.5)	20.7	1.78 (1.62 to 1.96)
	NOT poor and NOT old housing	8655 (29.2)	10.6	1.00	7929 (26.5)	10.9	1.00	4179 (14.4)	12.8	1.00

*No population in these strata.

low birth weight and childhood lead poisoning for two New England states in the period around the 1990 census, we found that area based measures of economic deprivation typically detected larger socioeconomic gradients than area based measures of education, occupation, or wealth. Moreover, census tract and block group level area based socioeconomic measures consistently detected equivalent and typically stronger socioeconomic gradients than their ZIP code level counterparts. Lastly, categories based on quintiles and a priori cut off points detected similar socioeconomic gradients; only the latter, however, were identically delimited across levels of geography within and across states.

Methodological considerations
Of note, diverse sources of error and bias could have affected our findings, albeit in ways unlikely to yield overestimates of socioeconomic gradients in health. If, for example, under-registration or misclassification of cases (with respect to outcomes) were non-differential with respect to poverty, estimates of effect would have been less precise. Alternatively, if these errors were positively associated with poverty or if persons subject to socioeconomic deprivation were less likely to have a geocodable address (for example, a PO box), as may have occurred among small number of children with very high lead concentrations (≥20 μg/dl) detected by the venous test, the ensuing conservative bias would have led to underestimation of socioeconomic gradients in health. These types of error, however, would have affected analyses at each level of geography and thus would not invalidate comparison of socioeconomic gradients across socioeconomic measures and across levels of geography. Additionally, we minimised geocoding error by using a firm whose accuracy we had previously validated[28] and further note that the extremely low proportion of areas without data on the area based socioeconomic measures would have little impact on our analyses.

Two additional concerns are relevant to delineating area based socioeconomic measures. As in the case of individual level measures of socioeconomic position, debate exists over the benefits and drawbacks of using: (a) single variable indicators versus composite area based indicators,[7 8 38 41–43] and (b) continuous versus categorical data and, if categorical, what cut off points should be used.[7 8 41–43] In our study, we tackled these issues empirically, by using diverse single variable and composite socioeconomic measures, with cut off points based on both centile distribution and a priori considerations. Of note, we found that estimates of effects detected using the single variable measure of poverty were similar to those based on composite measures, whether modelled based on a priori categorical cut off points or as quintiles.

Additional issues involving temporal and spatial scale and analytical level of our study also merit consideration. From an aetiological perspective, likelihood of temporal misclassification of SEP in relation to the birth and lead outcomes under study was likely to have been minimised because these outcomes occurred during the same time period as the census from which we derived our area based socioeconomic measures. Also of temporal significance, all of our study's area based socioeconomic measures can meaningfully be compared across decennial censuses, a necessary attribute for

monitoring socioeconomic trends over time.[54] Regarding spatial considerations, "ecologic fallacy", an often raised concern,[7 8 14 55] is not relevant to our study design. A form of aggregation bias, this fallacy arises when solely aggregate data are analysed and confounding is introduced by the grouping process generating the aggregate data. By contrast, in our study, 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). Thus, in this approach, validity of using area based socioeconomic measures depends on the extent to which areas constitute meaningful geographical units,[12 56] as is more likely to be the case for census tracts and block groups compared with ZIP codes.[7 22 31]

Analyses conducted for this first phase of our project did not, however, investigate spatial correlation (for example, nesting of block groups within tracts), issues of adjacency (for example, effects of living in a poor block group adjacent to chiefly poor versus more affluent block groups), or—in the case of the birth data—the combined impact of individual level education and area based socioeconomic position on risk of low birth weight. Extant evidence, despite some inconsistencies, nevertheless indicates that use of standard techniques (for example, random effects models) to adjust variance estimates for spatial correlation would likely have yielded more conservative estimates of statistical significance while not substantively changing the estimates or patterns of associations themselves.[14 57–59] Had analyses taken into account issues of adjacency however, different and additional effect estimates might have been obtained.[57–59] Moreover, multilevel analyses of birth outcomes using individual level and area based socioeconomic data have shown evidence of both independent and interactive effects,[15 18 19 60] implying that analyses only using area based measures are unlikely to capture the full impact of socioeconomic position on health.

Interpretation and implication of findings

Adding plausibility to our findings are the similar results for our *Public Health Disparities Geocoding Project's* analyses of mortality rates and cancer incidence—whereby census tract and block group measures of economic deprivation consistently detected the strongest socioeconomic gradients, whereas ZIP code level measures yielded attenuated or sometimes even contrary effect estimates.[25] The small number of US epidemiological studies investigating type and level of area based socioeconomic measures, albeit for fewer outcomes and fewer levels of geography, have likewise reported analogous results.[12 18 24 61–69]

Moreover, in the case of low birth weight and childhood lead poisoning, among eight recent US studies using area based socioeconomic measures, the two presenting data on geocoding success rate[5 18] reported geocoding a comparably high proportion of cases to the census tract and block group level (none of the eight, however, provided no information on the accuracy of the geocoding). Additionally, both US and European studies of low birthweight using area based socioeconomic measures have detected socioeconomic gradients of a similar magnitude (about a twofold increase in risk).[5 11 15–19 21 60 70–73] Effect estimates on the order of a 1.5-fold

to twofold increased risk of low birth weight have likewise been reported by studies using individual level or household level data on social class, education, and related socioeconomic measures.[3 9 10 60 73] Similarly, the magnitude of increased risk detected in our study for raised blood lead concentrations is similar to that reported in previous studies using both area based[11 21] and individual level socioeconomic measures.[74]

Further bolstering use of area based socioeconomic measures at the tract and block group, compared to ZIP code, level is the recent decision of the US Census Bureau to no longer include ZIP codes in the year 2000 census.[75–77] Prompting this decision were difficulties in defining ZIP codes' physical boundaries plus the temporal instability of these boundaries. Instead, the US Census Bureau opted to create a new geographical entity, the ZIP Code Tabulation Area (ZCTA), with boundaries coterminous with census blocks.[75 76] Importantly, individuals' postal (mailing) ZIP code may not be the same as their ZCTA, meaning that postal ZIP codes recorded in many public health surveillance systems cannot validly be linked to the ZCTA data.[75–77]

In summary, then, drawing upon our a priori criteria regarding external validity, robustness, completeness, and user friendliness, and buttressed by our similar findings for mortality and cancer incidence,[25] we offer a preliminary recommendation pertaining to geocoding of US public health surveillance systems, pending our analyses of tuberculosis, sexually transmitted infections, and violence. Specifically, we suggest that records should be geocoded to the tract or block group level and be linked to easily understood poverty related measures, demarcated by meaningful a priori categorical cut off points.

. . . .

Authors' affiliations
N Krieger, J T Chen, P D Waterman, M-J Soobader,
S V Subramanian, R Carson, Harvard School of Public Health, Boston, USA

Appendix: Cut off points for each variable in tables 3 and 4 were as follows (where "C" stands for category, "Q" for quintile, "L" for lower bound, and "U" for upper bound)

			Q1	Q2	Q3	Q4	Q5
	L, U	area	L, U	L, U	L, U	L, U	L, U
Single variable							
Working class (categorical; %)	C1: (0.0, 49.9)						
	C2: (50.0, 65.9)						
	C3: (66.0, 74.9)						
	C4: (75.0, 100.0)						
Unemployed (quintile; %)		MA BG	(0.0, 2.9)	(3.0, 4.9)	(5.0, 7.0)	(7.1, 10.2)	(10.3, 100)
		MA CT	(0.0, 4.2)	(4.3, 5.6)	(5.7, 6.9)	(7.0, 9.4)	(9.5, 100)
		MA ZC	(0.0, 4.3)	(4.4, 5.2)	(5.3, 6.3)	(6.4, 7.6)	(7.7, 25.3)
		RI BG	(0.0, 3.0)	(3.1, 5.1)	(5.2, 7.1)	(7.2, 10.1)	(10.2, 56.8)
		RI CT	(0.0, 4.5)	(4.6, 5.6)	(5.7, 6.9)	(7.0, 9.3)	(9.4, 22.8)
		RI ZC	(0.0, 4.4)	(4.5, 5.3)	(5.4, 6.3)	(6.4, 7.8)	(7.9, 13.5)
Median household income (quintile; $)		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)
		MA ZC	(9,762, 30,624)	(30,625, 36,246)	(36,247, 41,396)	(41,397, 48,841)	(48,842, 94,898)
		RI BG	(4,999, 22,088)	(22,089, 30,293)	(30,294, 35,567)	(35,568, 41,204)	(41,205, 150,001)
		RI CT	(6,462, 23,667)	(23,668, 31,032)	(31,033, 35,300)	(35,301, 40,606)	(40,607, 78,666)
		RI ZC	(8,787, 29,548)	(29,549, 33,614)	(33,615, 36,921)	(36,922, 41,356)	(41,357, 60,705)
Poverty (categorical; %)	C1: (0.0, 4.9)						
	C2: (5.0, 9.9)						
	C3: (10.0, 19.9)						
	C4: (20.0, 100.0)						
Gini (quintile)		MA BG	(0.009, 0.314)	(0.315, 0.350)	(0.351, 0.379)	(0.380, 0.421)	(0.422, 0.688)
		MA CT	(0.009, 0.348)	(0.349, 0.371)	(0.372, 0.395)	(0.396, 0.428)	(0.429, 0.650)
		MA ZC	(0.208, 0.344)	(0.345, 0.369)	(0.370, 0.387)	(0.388, 0.414)	(0.415, 0.614)
		RI BG	(0.014, 0.318)	(0.319, 0.351)	(0.352, 0.381)	(0.382, 0.422)	(0.423, 0.650)
		RI CT	(0.050, 0.349)	(0.350, 0.373)	(0.374, 0.395)	(0.396, 0.426)	(0.427, 0.595)
		RI ZC	(0.186, 0.352)	(0.353, 0.364)	(0.365, 0.394)	(0.395, 0.417)	(0.418, 0.551)
Wealth (categorical; %)	C1: (0.0, 4.9)						
	C2: (5.0, 9.9)						
	C3: (10.0, 19.9)						
	C4: (20.0, 100.0)						
Crowding* (quintile; %)		MA BG	(0.0, 0.0)	(0.0, 0.0)	(0.0, 1.4)	(1.5, 3.8)	(3.9, 60.9)
		MA CT	(0.0, 0.4)	(0.5, 0.9)	(1.0, 1.7)	(1.8, 3.7)	(3.8, 42.0)
		MA ZC	(0.0, 0.3)	(0.4, 0.8)	(0.9, 1.3)	(1.4, 2.2)	(2.3, 29.7)
		RI BG	(0.0, 0.0)	(0.0, 0.0)	(0.0, 1.5)	(1.6, 3.8)	(3.8, 62.5)
		RI CT	(0.0, 0.4)	(0.5, 0.9)	(1.0, 1.6)	(1.7, 3.0)	(3.1, 27.7)
		RI ZC	(0.0, 0.4)	(0.5, 0.9)	(1.0, 1.2)	(1.3, 2.2)	(2.3, 15.3)
Low education (categorical; %)	C1: (0.0, 14.9)						
	C2: (15.0, 24.9)						
	C3 (25.0, 39.9)						
	C4: (40.0, 100.0)						
High education (categorical; %)	C1: (0.0, 14.9)						
	C2: (15.0, 24.9)						
	C3: (25.0, 39.9)						
	C4: (40.0, 100.0)						

Appendix: continued

	L, U	area	Q1 L, U	Q2 L, U	Q3 L, U	Q4 L, U	Q5 L, U
Composite variable							
Townsend index (quintile)		MA BG	(−5.531, −2.468)	(−2.467, −1.331)	(−1.330, 0.094)	(0.095, 2.425)	(2.426, 11.804)
		MA CT	(−8.123, −2.797)	(−2.796, −1.596)	(−1.595, −0.051)	(−0.050, 2.860)	(2.861, 11.223)
		MA ZC	(−7.864, −2.388)	(−2.387, −1.411)	(−1.410, −0.165)	(−0.164, 1.645)	(1.646, 13.626)
		RI BG	(−5.811, −2.410)	(−2.409, −1.250)	(−1.249, 0.162)	(0.163, 2.293)	(2.294, 9.832)
		RI CT	(−5.572, −2.595)	(−2.594, −1.502)	(−1.501, −0.078)	(−0.077, 2.793)	(2.794, 9.103)
		RI ZC	(−8.003, −1.905)	(−1.904, −0.929)	(−0.928, −0.246)	(−0.245, 2.301)	(2.302, 10.060)
Index of Local Economic Resources (quintile)		MA BG	(0, 6)	(7, 11)	(12, 15)	(16, 20)	(21, 27)
		MA CT	(0, 5)	(6, 10)	(11, 15)	(16, 19)	(20, 26)
		MA ZC	(0, 8)	(9, 12)	(13, 15)	(16, 19)	(20, 26)
		RI BG	(0, 4)	(5, 8)	(9, 12)	(13, 17)	(18, 27)
		RI CT	(0, 4)	(5, 8)	(9, 12)	(13, 16)	(17, 26)
		RI ZC	(0, 8)	(9, 10)	(11, 13)	(14, 15)	(16, 27)
Factor 1 (quintile)		MA BG	(−1.861, −0.980)	(−0.979, −0.370)	(−0.369, 0.288)	(0.289, 0.971)	(0.972, 2.110)
		MA CT	(−1.878, −0.995)	(−0.994, −0.358)	(−0.357, 0.313)	(0.314, 1.022)	(1.203, 1.854)
		MA ZC	(−1.691, −0.857)	(−0.856, −0.395)	(−0.394, 0.058)	(0.059, 0.806)	(0.807, 2.538)
		RI BG	(−1.831, −0.936)	(−0.935, −0.398)	(−0.397, 0.296)	(0.297, 0.984)	(0.985, 1.916)
		RI CT	(−2.084, −0.890)	(−0.889, −0.313)	(−0.312, 0.241)	(0.242, 1.027)	(1.028, 1.851)
		RI ZC	(−1.937, −0.713)	(−0.712, −0.293)	(−0.292, 0.152)	(0.153, 0.683)	(0.684, 2.324)
Factor 2 (quintile)		MA BG	(−1.831, −0.975)	(−0.974, −0.334)	(−0.333, 0.308)	(0.309, 0.980)	(0.981, 1.921)
		MA CT	(−1.847, −1.004)	(−1.003, −0.238)	(−0.327, 0.308)	(0.309, 1.013)	(1.014, 1.874)
		MA ZC	(−1.869, −0.996)	(−0.995, −0.363)	(−0.362, 0.261)	(0.262, 0.939)	(0.940, 2.234)
		RI BG	(−2.264, −0.978)	(−0.977, −0.214)	(−0.213, 0.470)	(0.471, 0.963)	(0.964, 1.528)
		RI CT	(−2.243, −1.082)	(−1.081, −0.195)	(−0.194, 0.448)	(0.449, 0.990)	(0.991, 1.557)
		RI ZC	(−2.223, −0.886)	(−0.885, −0.100)	(−0.099, 0.348)	(0.349, 0.810)	(0.811, 2.070)
SEP index (quintile)		MA BG	(−16.524, −2.975)	(−2.974, −1.099)	(−1.098, 0.479)	(0.480, 2.701)	(2.702, 22.208)
		MA CT	(−13.768, −3.265)	(−3.264, −1.153)	(−1.152, 0.396)	(0.397, 3.006)	(3.007, 20.605)
		MA ZC	(−14.165, −3.122)	(−3.121, −0.956)	(−0.955, 0.794)	(0.795, 2.744)	(2.745, 18.943)
		RI BG	(−16.457, −3.001)	(−3.000, −1.282)	(−1.281, 0.630)	(0.631, 2.966)	(2.967, 17.356)
		RI CT	(−15.883, −3.452)	(−3.451, −1.684)	(−1.683, 0.388)	(0.389, 3.767)	(3.768, 12.140)
		RI ZC	(−8.976, −3.221)	(−3.220, −1.079)	(−1.078, 0.511)	(0.512, 2.834)	(2.835, 13.497)

			% < poverty†	% working class	% expensive homes†
SEP1 (categorical)	C1:		≥20%	≥75%	{any value}
	C2:		≥20%	51–74%	<10%
	C3:		<20%	≥75%	{any value}
	C4:		<20%	51–74%	<10%
	C5:		{any value}	<50%	<10%
	C6:		{any value}	51–74%	≥10%
	C7:		{any value}	<50%	≥10%

*Note: for MA BG and RI BG, the referent group is C3, since C1 and C2 are empty. †Note: for % poverty, C5–C7 is effectively <20%; for % expensive homes, C1 and C3 are effectively <10%.

REFERENCES

1 **National Center for Health Statistics**. *Health, United States, 1998 with socioeconomic status and health chartbook.* Hyattsville, MD: US Dept of Health and Human Services, 1998.

2 **Krieger N**, Chen JT, Ebel G. Can we monitor socioeconomic inequalities in health? A survey of US Health Departments' data collection and reporting practices. *Public Health Rep* 1997;**112**:481–91.

3 **Dubay L**, Joyce T, Kaestner R, *et al*. Changes in prenatal care timing and low birth weight by race and socioeconomic status: implications for the Medicaid expansions for pregnant women. *Health Serv Res* 2001;**36**:373–98.

4 **Centers for Disease Control and Prevention (CDC)**. *Screening young children for lead poisoning: guidance for state and local public health officials.* Atlanta, GA: US Department of Health and Human Services, Public Health Service, CDC, 1997.

5 **Reissman DB**, Staley F, Curtis GB, *et al*. Use of geographic information system technology to aid Health Department decision making about childhood lead poisoning prevention activities. *Environ Health Perspect* 2001;**109**:89–94.

6 **Barker DJ**. *Mothers, babies and health in later life.* 2nd edn. Camden Town, UK: Churchill Livingston, 1998.

7 **Krieger N**, Williams D, Moss N. Measuring social class in US public health research: concepts, methodologies and guidelines. *Annu Rev Public Health* 1997;**18**:341–78.

8 **Lynch J**, Kaplan G. Socioeconomic position. In: Berkman L, Kawachi I, eds. *Social epidemiology.* Oxford: Oxford University Press, 2000:13–35.

9 **Parker JD**, Schoendorf KC, Kiely JL. Associations between measures of socioeconomic status and low birth weight, small for gestational age, and premature delivery in the United States. *Ann Epidemiol* 1994; **4**:271–8.

10 **Cramer JC**. Racial and ethnic differences in birthweight: the role of income and financial assistance. *Demography* 1995;**32**:231–47.

11 **Sargent JD**, Brown MJ, Freeman JL, *et al*. Childhood lead poisoning in Massachusetts communities: its association with sociodemographic and housing characteristics. *Am J Public Health* 1995;**85**:528–34.

12 **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**:703–10.

13 **Krieger N**, Zierler S, Hogan JW, *et al*. Geocoding and measurement of neighborhood socioeconomic position. In: Kawachi I, Berkman LF, eds. *Neighborhoods and health.* New York, NY: Oxford University Press (in press).

14 **Diez Roux AV**. Investigating neighborhood and area effects on health. *Am J Public Health* 2001;**91**:1783–9.

15 **O'Campo P**, Xue X, Wang MC, *et al*. Neighborhood risk factors for low birthwive in Baltimore: a multilevel analysis. *Am J Public Health* 1997;**87**:1113–18.

16 **Roberts EM**. Neighborhood social environments and the distribution of low birthweight in Chicago. *Am J Public Health* 1997;**87**:597–603.

17 **Fang J**, Madhavan S, Alderman MH. Low birth weight: race and maternal nativity—impact of community income. *Pediatrics* 1999;**103**:E5.

18 **Pearl M**, Braveman P, Abrams B. The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *Am J Public Health* 2001;**91**:1808–14.

19 **Rauh VA**, Andrews HF, Garfinkel RS. The contribution of maternal age to racial disparities in birthweight: a multilevel perspective. *Am J Public Health* 2001;**91**:1815–24.

20 **Kirby D**, Coyle K, Gould JB. Manifestations of poverty and birthrates among young teenagers in California zip code areas. *Fam Plann Perspect* 2001;**33**:63–9.

21 **Rhode Island Department of Health**. Blood lead levels among children—Rhode Island, 1993–1995. *MMWR* 1995;**44**:788–91.

22 **Hanchette CL**. GIS and decision making for public health agencies: childhood lead poisoning and welfare reform. *J Public Health Management Practice* 1999;**5**:41–7.

23 **Krieger N**. Theories for social epidemiology for the 21st century: an ecosocial perspective. *Int J Epidemiol* 2001;**30**:668–77.

24 **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.

25 **Krieger N**, Chen J, Waterman PD, *et al*. 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. Invited presentation to the National Cancer Institute, Bethesda, MD, October 11, 2001. (Available from first author).

26 **US Census Bureau**. *American fact finder.* Massachusetts. (http://factfinder.census.gov/bf/ _lang=en_vt_name=DEC_2000_SF1_U_DP1_geo_id=04000 US25.html) (last accessed 30 August 2001).

27 **US Census Bureau**. American fact finder. Rhode Island. (http://factfinder.census.gov/servlet/ BasicFactsTable?_lang=en&_vt_name=DEC_1990_STF1 _DP1&_geo_id=04000US44) (last accessed 30 August 2001).

28 **Krieger N**, Waterman P, Lemieux K, *et al*. On the wrong side of the tracts? Evaluating accuracy of geocoding for public health research. *Am J Public Health* 2001;**91**:1114–16.

29 **National Committee on Clinical Laboratory Standards (NCCLS)**. *Analytical procedures for the determination of lead in blood and urine; approved guideline.* NCCLS document C40-A. Wayne, PA: NCCLS, 2001.

30 **US Bureau of Census**. *Census of population and housing, 1990: Summary Tape File 3 technical documentation.* Washington, DC: Bureau of the Census, 1991.

31 **US Bureau of the Census**. *Geographical areas reference manual.* Washington, DC: US Dept of Commerce, 1994. (http:// www.census.gov/geo/www/garm.html) (last accessed 29 November 2001).

32 **Shaw M**, Dorling D, Gordon D, *et al. The widening gap: health inequalities and policy in Britain.* Bristol, UK: Policy Press, 1999.

33 **Evans T**, Whitehead M, Diderichsen F, *et al. Challenging inequities in health: from ethics to action.* New York, NY: Oxford University Press, 2001.

34 **Leon D**, Walt G, eds. *Poverty, inequality, and health: an international perspective.* Oxford: Oxford University Press, 2001.

35 **Berkman L**, Kawachi I, eds. *Social epidemiology.* Oxford, UK: Oxford University Press, 2000.

36 **Council of Economic Advisors for the President's Initiative on Race**. *Changing America: indicators of social and economic well-being by race and Hispanic origin.* September 1988. (http://www.whitehouse.gov/ WH/EOP/CEA/html/publications.html) (last accessed 15 May 2000).

37 **Carstairs V**, Morris R. Deprivation and mortality: an alternative to social class? *Community Med* 1989;**11**:210–19.

38 **Lee P**, Murie A, Gordon D. *Area measures of deprivation: a study of current methods and best practices in the identification of poor areas in Great Britain.* Birmingham, UK: Centre for Urban and Regional Studies, University of Birmingham, 1995.

39 **Gordon D**, Spicker P, eds. *The international glossary of poverty.* CROP International Series on Poverty. London: Zed Books, 1999.

40 **Cowell FA**. *Measuring inequality.* 2nd edn. LSE Handbooks in Economics Series. London: Prentice Hall, 1995.

41 **Gordon D**. Census based deprivation indices: their weighting and validation. *J Epidemiol Community Health* 1995;**49** (suppl 2):S39–44.

42 **Townsend P**, Phillimore P, Beattie A. *Health and deprivation: inequality and the North.* London: Croom Helm, 1988.

43 **Carstairs V**. Deprivation indices: their interpretation and use in relation to health. *J Epidemiol Community Health* 1995;**49** (suppl 2):S3–8.

44 **Casper ML**, Barnett E, Halverson JA, *et al. Women and heart disease: an atlas of racial and ethnic disparities in mortality.* Morgantown, WV: Office for Social Environment and Health Research, West Virginia University, 1999.

45 **Lawley DN**, Maxwell AE. *Factor analysis as a statistical method.* New York: Macmillan, 1971.

46 **Bartholomew DJ**. *Latent variable models and factor analysis.* London: Charles Griffin, 1987.

47 **US Bureau of the Census**. *Poverty areas.* (http://www.census.gov/ population/socdemo/statbriefs/povarea.html) (last accessed 8 October 2001).

48 **Jargowsky PA**. *Poverty and place: ghettos, barrios, and the American city.* New York: Russell Sage Foundation, 1997.

49 **Pamuk ER**. Social class inequality in mortality from 1921 to 1972 in England and Wales. *Population Studies* 1985;**39**:17–31.

50 **Wagstaff A**, Paci P, van Doorslaer E. On the measurement of inequalities in health. *Soc Sci Med* 1991;**33**:545–57.

51 **Davey Smith G**, Hart C, Hole D, *et al*. Education and occupational social class: which is the more important indicator of mortality risk? *J Epidemiol Community Health* 1998;**52**:153–60.

52 **Massachusetts Department of Public Health**. Impact of multiple births on low birthweight—Massachusetts, 1989–1996. *MMWR* 1999;**48**:289–92.

53 **SAS Institute**. *SAS/STAT user's guide, version 6, 4th edn, vol 1.* Cary, NC: SAS Institute, 1989.

54 **Rossi RJ**, Gilmartin KJ. *The handbook of social indicators: sources, characteristics, and analysis.* New York, NY: Garland STPM Press, 1980.

55 **Alker HR Jr**. A typology of ecologic fallacies. In: Doggan M, Rokkan S, eds. *Social ecology.* Cambridge, MA: MIT Press, 1969:69–86.

56 **Oppenshaw S**, Taylor PJ. The modifiable areal unit problem. In: Wrigley N, Bennett RJ, eds. *Quantitative geography.* London: Routledge and Kegan Paul, 1981:60–9.

57 **Subramanian SV**, Duncan C, Jones K. Multilevel perspectives on modeling census data. *Environment Planning A* 2001;**33**:399–417.

58 **Duncan C**, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998;**46**:97–117.

59 **Soobader M**, LeClere FB. Aggregation and the measurement of income inequality, effects on morbidity. *Soc Sci Med* 1999;**48**:733–44.

60 **Pattenden S**, Dolk H, Vrijheid M. Inequalities in low birth weight: parental social class, area deprivation, and "lone mother" status. *J Epidemiol Community Health* 1999;**53**:355–8.

61 **Geronimus AT**, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. *Am J Epidemiol* 1998;**48**:475–86.

62 **Geronimus AT**, Bound J, Neidert LJ. On the validity of using census geocode data to proxy individual socioeconomic characteristics. *J Am Stat Assoc* 1996;**91**:529–37.

63 **Cherkin DC**, Grothaus L, Wagner EH. Is magnitude of co-payment effect related to income? Using census data for health services research. *Soc Sci Med* 1992;**34**:33–41.

64 **Greenwald JP**, Polissar NL, Borgatta EF, *et al*. Detecting survival effects of socioeconomic status: problems in the use of aggregate data. *J Clin Epidemiol* 1994;**47**:903–9.

65 **Carr-Hill R**, Rice N. Is enumeration district level an improvement on ward level analysis in studies of deprivation and health? *J Epidemiol Community Health* 1995;**49**(suppl 2):S28–9.

66 **Mustard CA**, Derksen S, Berthelot JM, *et al.* Assessing ecologic proxies for household income: a comparison of household and neighbourhood level income measures in the study of population health status. *Health Place* 1999;**5**:157–71.

67 **Reijneveld SA**, Verheij RA, de Bakker DH. The impact of area deprivation on differences in health: does the choice of the geographical classification matter? *J Epidemiol Community Health* 2000;**54**:306–13.

68 **Hyndman JCT**, D'Arcy D, Holman J, *et al.* Misclassification of social disadvantage based on geographical areas: comparison of postcode and collector's districts analyses. *Int J Epidemiol* 1995;**24**:165–76.

69 **Diez Roux AV**, Kiefe CI, Jacobs DR, *et al.* Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Ann Epidemiol* 2001;**11**:395–405.

70 **Crosse EA**, Alder RJ, Ostbye T, *et al.* Small area variation in low birthweight: looking beyond socioeconomic predictors. *Can J Public Health* 1997;**88**:57–61.

71 **Lund R**, Modvig J, Hilden J, *et al.* Risk of low birthweight in social districts of Copenhagen. *Scand J Public Health* 1999;**27**:89–93.

72 **Bonellie SR**. Effect of maternal age, smoking and deprivation on birthweight. *Paediatr Perinat Epidemiol* 2001;**15**:19–26.

73 **Spencer N**, Bambang S, Logan S, *et al.* Socioeconomic status and birth weight: comparison of an area-based measure with the Registrar General's social class. *J Epidemiol Community Health* 1999;**53**:495–8.

74 **Pirkle JL**, Kaufmann RB, Brody DJ, *et al.* Exposure of the US population to lead, 1991–1994. *Environ Health Perspect* 1998;**106**:745–50.

75 **US Bureau of Census**. ZIP Code Tabulation Area (ZCTA) Frequently Asked Questions. Available at: http://www.census.gov/geo/ZCTA/zctafaq.html (accessed 12 February 2002).

76 **US Bureau of Census**. Census 2000 ZCTAs ZIP Code Tabulation Areas Technical Documentation. Available at: http://www.census.gov/geo/ZCTA/zcta_tech_doc.pdf (accessed 12 February 2002).

77 **Krieger N**, Waterman P, Chen JT, *et al.* 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**:1100–2.