Published in final edited form as:
Cancer Causes Control. 2011 April ; 22(4): 631–647. doi:10.1007/s10552-011-9736-5.

The California Neighborhoods Data System: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations

Scarlett Lin Gomez,
Cancer Prevention Institute of California, 2201 Walnut Avenue, Suite 300, Fremont, CA 94536, USA. Department of Health Research and Policy, School of Medicine, Stanford University, Stanford, CA, USA

Sally L. Glaser,
Cancer Prevention Institute of California, 2201 Walnut Avenue, Suite 300, Fremont, CA 94536, USA. Department of Health Research and Policy, School of Medicine, Stanford University, Stanford, CA, USA

Laura A. McClure,
Cancer Prevention Institute of California, 2201 Walnut Avenue, Suite 300, Fremont, CA 94536, USA

Sarah J. Shema,
Cancer Prevention Institute of California, 2201 Walnut Avenue, Suite 300, Fremont, CA 94536, USA

Melissa Kealey,
Division of Community Health and Human Development and Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA

Theresa H. M. Keegan, and
Cancer Prevention Institute of California, 2201 Walnut Avenue, Suite 300, Fremont, CA 94536, USA. Department of Health Research and Policy, School of Medicine, Stanford University, Stanford, CA, USA

William A. Satariano
Division of Community Health and Human Development and Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA

Scarlett Lin Gomez: scarlett@cpic.org

Abstract

Research on neighborhoods and health has been growing. However, studies have not investigated the association of specific neighborhood measures, including socioeconomic and built environments, with cancer incidence or outcomes. We developed the California Neighborhoods Data System (CNDS), an integrated system of small area-level measures of socioeconomic and built environments for California, which can be readily linked to individual-level geocoded records. The CNDS includes measures such as socioeconomic status, population density, racial residential segregation, ethnic enclaves, distance to hospitals, walkable destinations, and street connectivity. Linking the CNDS to geocoded cancer patient information from the California Cancer Registry, we demonstrate the variability of CNDS measures by neighborhood

Correspondence to: Scarlett Lin Gomez, scarlett@cpic.org.

socioeconomic status and predominant race/ethnicity for the 7,049 California census tracts, as well as by patient race/ethnicity. The CNDS represents an efficient and cost-effective resource for cancer epidemiology and control. It expands our ability to understand the role of neighborhoods with regard to cancer incidence and outcomes. Used in conjunction with cancer registry data, these additional contextual measures enable the type of transdisciplinary, "cells-to-society" research that is now being recognized as necessary for addressing population disparities in cancer incidence and outcomes.

Keywords

Neighborhood; Socioeconomic environment; Built environment; Immigration; Contextual factors; GIS

Introduction

> We now know that like real estate, health is location, location, location. Where you live makes an enormous difference in terms of the air you breathe, the schools you go to, the work, transportation, housing, streets, violence levels, etcetera, that you live with on a day-to-day basis. So unless we create some innovative strategies to fundamentally change the nature of disadvantaged neighborhoods, we're in trouble. George Kaplan [1].

A growing body of research [2] is demonstrating that health in human populations [3] is affected by the environment in which we live, including their social [4], socioeconomic, and built [2, 5] aspects (i.e., man-made surroundings that provide the setting for human activity [6]). Initially, this work involved area-based socioeconomic measures [7] although only as proxies for individual-level socioeconomic status (SES) [8–11]. However, studies have shown consistently that area-based socioeconomic measures and related neighborhood socioeconomic factors are associated with health outcomes independent of individual-level factors [12–14], including SES [15–19]. Haan et al. found higher mortality for California residents of a federally designated poverty area than for those living in non-poverty areas, even after adjusting for characteristics of the individual residents [15]. Yen and Kaplan reported that contextual socioeconomic and built environments (e.g., commercial stores, SES, population and size of census tract, housing (renters and single-family dwellings)) were strongly associated with mortality after adjusting for the same individual-level social and behavioral factors as Haan et al. [19]. These observations led to a new generation of research demonstrating that something about the neighborhood itself impacts a variety of health outcomes [3, 20–27]. More important, these studies suggest that interventions and policies that focus only on individuals and ignore the neighborhood environment may be less effective in improving public health than interventions that focus on both individuals and their neighborhoods [17].

To date, little attention has been paid to the effects of the socioeconomic and built environments on the occurrence of cancer or on the quality and duration of life following a cancer diagnosis. Yet, the socioeconomic and built environments have potential to impact cancer incidence and/or outcomes through a number of mechanisms, including those that involve behavioral risk factors, stress, access to health care and resources, social support, social capital, and collective efficacy (i.e., shared beliefs among neighbors about their shared capability for action [28]) [29–31]. Research into these possibilities should be feasible, given that the United States has strong systems of population-based cancer registration, with information routinely collected on cancer patient demographics, diagnosis, treatment and survival through state and federal programs (the National Cancer Institute

Surveillance, Epidemiology, and End Results (SEER) program, Centers for Disease Control and Prevention National Program for Cancer Registries, and state departments of health). These cancer surveillance data have been vital to the monitoring of detailed cancer patterns [32, 33], and their research value has been further enhanced through linkages to complementary databases, such as claims data for characterizing treatments and comorbidities from Medicare [34, 35], sociodemographic data from the National Longitudinal Mortality Study (NLMS) [36, 37], insurance status data from Medicaid [38], birth characteristics data from birth certificates [39, 40], and environmental exposure data from various sources [41–46]. Supplementation of cancer registry data through such linkages has enabled better understanding of the impact of these additional factors on cancer incidence and outcomes at the population-level. For example, the addition of area-based socioeconomic measures to cancer registry data led to profound advancements in public health knowledge about the joint effects of SES and race/ethnicity on cancer population patterns [11, 47–50].

Because population-based cancer registry data are routinely geocoded, opportunities also exist to integrate these data with existing area-based census and other geographic information systems (GIS) data on patient neighborhood environments such as land-use patterns, housing density, and street patterns. The joining of these two types of geocoded data would permit surveillance of cancer patterns enhanced by including patient neighborhood in addition to individual characteristics in analyses. This richer data resource would have significant potential for enabling much-needed breakthroughs into the prevention of the occurrence of cancer as well as its adverse outcomes.

Toward this end, we compiled the California Neighborhoods Data System (CNDS), an extensive set of geospatial data to characterize the socioeconomic and built environment in California. Because of California's size and diversity, this data set is both large and diverse in geographic, racial/ethnic, and socioeconomic characteristics. Further, to explore associations of the socioeconomic and built environment with cancer incidence, diagnosis, treatment, and survival on the population level, we linked the CNDS to patient records in the California Cancer Registry (CCR), the statewide population-based cancer surveillance system that also comprises three SEER registries.

We describe here the methods used to compile the CNDS and present descriptive statistics to illustrate the variability of the socioeconomic and built environment measures across California neighborhoods by SES and racial/ethnic composition. We also present these measures for cancer patients by race/ethnicity, using colorectal cancer as an example, and discuss the potential usefulness of the CNDS in cancer surveillance research.

Methods

The California Neighborhoods Data System (CNDS)

To establish the CNDS, we identified sources of geospatial data for characterizing the neighborhood socioeconomic and built environments of California, focusing on publicly available databases (e.g., national data sources such as the Census) obtained at no or minimal cost (with the exception of business listing data) for cost effectiveness. After downloading the data, we geocoded records, when necessary, and used either new or existing measures to operationalize the variables. We compiled only data available at a census tract (CT) or finer level, as prior research has shown that census tracts are a reasonable level of geography for estimating neighborhood effects [51–53]. The CNDS thus contains information on socioeconomic and built environment measures for all 7,049 CTs in California, except for 47 CTs for which information was unavailable due to inadequate population size. Table 1 shows the specific socioeconomic and built environment data in the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

CNDS and cites the prior studies evaluating the validated or otherwise conceptualized measures used. For consistency, this paper presents data for all measures at the CT-level.

To illustrate how the socioeconomic and built environment measures in the CNDS vary by common sociodemographic factors at the neighborhood-level, we determined distributions of the socioeconomic and built environment measures for California CTs overall and then stratified them by: (1) statewide quintiles of a commonly used composite index of SES [54] (which includes Census data on education, income, occupation, and housing costs) and (2) the predominant race/ethnicity of the CT (defined as >50% non-Hispanic White (called “White”), non-His-panic Black (called “Black”), Hispanic, non-Hispanic Asian/Pacific Islander (called “API”), or mixed, if not any of the previous).

CNDS-CCR

To employ the CNDS for cancer-related research, we obtained cancer information from the CCR, which is the largest, contiguous-area, population-based cancer registry system in the country, and has been collecting data under state mandate using uniform, high-quality reporting standards since 1988. We appended all CCR cancer patient records to the CNDS using geocodes based on residential address at diagnosis from the CCR. In addition to a range of patient demographic and tumor information, the CCR also includes data on the reporting (usually diagnosing) hospital for each diagnosis and has been linked to hospital utilization data from the California Office of Statewide Health Planning and Development (OSHPD) to include information on hospital characteristics, such as the facility’s number of beds, ownership status, and Medicaid enrollment [55–60]. In addition, for each California hospital, we have determined academic or teaching status, and any affiliation with an NCI-designated cancer center, and computed measures of racial/ethnic cancer patient composition for each hospital. Thus, the CNDS-CCR also includes data on institutional characteristics, including distance to and characteristics of hospitals.

To describe variations in socioeconomic and built environment measures for a population-based sample of CCR cancer patients, we linked each measure at the CT-level to in situ and invasive colorectal cancer cases diagnosed in California during the period 1996 through 2004. We selected colorectal cancer as an example because it is a commonly diagnosed cancer across both genders, and all racial/ethnic, and socioeconomic groups [61], and this time frame because it is centered around and/or includes the time period when most of the contextual data were collected. All patient addresses at diagnosis had been geocoded and assigned a CT. Cases were excluded from the analyses based on having missing or invalid CT number (n = 710), missing or invalid latitude and/or longitude (n = 312), or geocoding based on something other than a complete and valid address (n = 6,438). This resulted in a sample size of 129,172 colorectal cancer patients for these analyses.

In addition to the socioeconomic and built environment measures for the contextual analysis, we included other geocoded data for an individual-level analysis. These data include business listings, used for characterizing socioeconomic and built neighborhood resources, and the cases’ hospital of diagnosis. For the business data, we computed the numbers and types of businesses within a 400 or 1,600 m zone around each patient’s residence at diagnosis. We also computed the distance between the patient’s residence at diagnosis and the diagnosing hospital with the “greater circles” distance calculation, which computes the shortest distance between two points along a sphere, such as the earth, using a SAS macro developed by the North American Association of Central Cancer Registries [62]. The colorectal cancer patient-level dataset was then linked to the CT-level socioeconomic and built environment measures in the CNDS. The distributions of these measures among cancer patients were stratified by race/ethnicity (Table 4).

Results

Distributions of CNDS socioeconomic and built environment measures by neighborhood SES

Table 2 presents the distributions of most of the CNDS measures for California CTs overall and stratified by neighborhood SES. Higher SES neighborhoods tend to be less "built" (i.e., comprising evidence of more human designs), as measured by population density; percent of households that are crowded (≥1 occupant/room); percent non-single family housing units and housing structures with ≥10 units; percent traveling to work by public transportation, foot or bicycle; percent working at home; percent traveling more than 45 min to work; number of streets and intersections; and size and length of street blocks. Neighborhood SES also varies with racial/ethnic and immigration composition, with higher SES neighborhoods having fewer Blacks and Hispanics, more Whites and APIs, fewer immigrants and residents with limited English proficiency, and fewer Hispanic ethnic enclaves (In anthropology and sociology, an "ethnic enclave" is a complex and dynamic construct [63, 64]. In general, ethnic enclaves refer to geographic units that tend to maintain more cultural mores and distinction from the larger surrounding area. Here, our conceptualization of ethnic enclave is based on an index that includes census variables (see Table 1) on the composition of the neighborhood residents in terms of its race/ethnicity, language, nativity, and recency of immigration.) The association of SES with the enclave index is less pronounced for APIs than Hispanics because APIs have a high percent of foreign-born but also tend to be of higher SES.

Distributions of CNDS socioeconomic and built environment measures by predominant neighborhood race/ethnicity

CTs that are predominantly White or API generally have a more favorable distribution of socioeconomic environment measures (i.e., higher neighborhood SES) than predominantly Black or Hispanic CTs, while CTs of mixed race/ethnicity often display intermediate values (Table 3). Overall neighborhood SES is higher among predominantly White and API CTs than among predominantly Black and Hispanic CTs; however, predominantly Black CTs are more evenly distributed among the lowest three quintiles of SES whereas the vast majority of Hispanic CTs are in the lowest SES quintile.

Despite being similar to White CTs for most measures, predominantly API CTs more closely resemble predominantly Hispanic and, for some measures, Black CTs, on factors such as percent crowding, percent with limited English proficiency, percent foreign-born (having the highest of any race/ethnicity), percent age ≥25 years without a high school diploma, and population density (Table 3). However, these API neighborhoods have higher neighborhood SES measures than White CTs for measures such as median household income, median gross rent, and median value of owner-occupied houses. Predominately, API CTs are also predominantly Asian ethnic enclaves. Predominantly, Hispanic CTs are more evenly distributed among the top two quintiles of the Hispanic enclave index.

Built environment measures do not appear to vary widely by predominant neighborhood race/ethnicity except for street connectivity measures. The size and length of street blocks are greater (indicating less walkability) in predominantly White CTs than in CTs of any other predominant race/ethnicity.

Distributions of socioeconomic and built environment measures by among colorectal cancer cases

Table 4 presents the distributions of socioeconomic and built environment measures among colorectal cancer cases in California stratified by patient race/ethnicity. While these results

NIH-PA Author Manuscript

NIH-PA Author Manuscript

are largely similar to the statewide CT-level distributions by race/ethnicity, they show that Hispanic colorectal cancer cases tend to live in CTs with slightly more favorable socioeconomic environments than the predominantly Hispanic CTs at the state level. For example, compared to the predominantly Hispanic CTs, Hispanic cancer cases live in CTs with a lower population density; lower percent crowding; lower percent of foreign-born residents and residents with limited English proficiency; greater percent in higher neighborhood SES quintiles; lower percent in the most unacculturated quintile of the enclave index; lower percent age ≥ 25 years without a high school diploma and percent below 200% federal poverty level; and greater median household income, median gross rent, and median house value.

Black colorectal cancer patients tended to live in neighborhoods with higher population density, shorter block length and size, and more walkable destinations. API colorectal cancer patients were also more likely to live near more walkable destinations and in neighborhoods with higher population density.

Black patients were more often diagnosed at larger, private or health maintenance organization (HMO) hospitals than the other races/ethnicities. Distance to diagnosing hospital was similar among all races/ethnicities. Patients were generally more likely to be diagnosed at hospitals with relatively higher proportions of patients of their same race/ethnicity.

In summary, California CTs of higher SES also tended to be less built, and of lower Hispanic enclave status. The association between neighborhood SES and API ethnic enclave was less pronounced and may be due to the heterogeneity in SES of the combined API population. Predominantly White and API CTs tend to be of higher SES for most socioeconomic environment measures, while predominantly Black and Hispanic CTs tend to be of lower SES for most measures. The only built environment measure that seemed to vary with neighborhood racial/ethnic composition was street connectivity, with predominantly White CTs having larger size and longer length of street blocks. Patient-level distributions of socioeconomic and built environment measures by race/ethnicity were generally similar to CT-level distributions. Hospital-level factors varied across racial/ethnic groups of patients.

Discussion

The CNDS represents a comprehensive and geographically extensive resource, based entirely on secondary data and designed to include data that are easily accessible to most researchers. Our initial, primary objective was to create a resource that could be used in combination with an existing statewide cancer surveillance system to extend cancer surveillance to neighborhood factors. However, this innovative resource of detailed population-level information has been useful for expanding our research to understand how the surrounding environment can be addressed to reduce cancer burdens across the California population. This resource serves as a model for similar unique databases that could be created for other geographic regions, including those with cancer registries.

Most studies of neighborhoods and cancer have been restricted to area-based socioeconomic measures. The few studies that have included both individual-based and area-based measures of SES have demonstrated independent effects of both individual-based and area-based SES on cancer incidence [65–67] and screening [68]. These findings strongly suggest that it may be useful to move beyond these area-based measures to identify the specific neighborhood elements that appear to have a profound effect on cancer risk [69]. The development of the CNDS, with its linkage to data from a large population-based cancer registry, makes this more detailed and sophisticated research possible.

Neighborhood environments can be measured in a variety of ways, including self-report, detailed audits, and GIS-based sources. As one goal for the CNDS was to be able to link it to the geocoded CCR database to enhance population-based cancer surveillance in California, we created this resource from secondary data available for the entire state. We also focused on socioeconomic and built environment measures to characterize those aspects of an individual's neighborhood environment anticipated to affect population variations in cancer incidence and outcomes. In this paper, we have used descriptive data to illustrate that the CNDS socioeconomic and built environment measures do indeed vary across neighborhoods, by SES and racial/ethnic composition, and among individual colorectal cancer patients, by race/ethnicity. The extent of the variability suggests that these measures may be sensitive enough for further characterizing variations in cancer incidence and outcomes, defining at-risk patient subgroups, and explaining racial/ethnic and socioeconomic disparities. However, the effects of these socioeconomic and built environment measures can vary by other neighborhood and patient factors such as SES, race/ethnicity, gender, and age, as we and others have shown [17, 70]. Consequently, we have found it important to consider interactions among these factors when examining associations with cancer incidence and outcomes. For example, in prior work, we found that for Black breast cancer patients, living in predominantly Black census block groups within highly segregated areas was associated with better survival than living in census block groups with fewer Blacks within less segregated areas. However, among white breast cancer patients, living in predominantly Black census block groups within highly segregated areas was associated with poorer survival than living in census block groups with fewer Blacks within less segregated areas [70]. Such results illustrate the potential for CNDS data to yield more detailed results (e.g., identify specific subgroups at risk) than cancer registry data alone and thus to be more useful in cancer control efforts.

Neighborhood-level factors have the potential to influence health through a number of mechanisms. These include access to health-promoting resources such as social and health services, affordable and nutritious foods, recreational facilities, decent housing, and transportation; walkability; shared attitudes among the community toward health and health behaviors; stress; social support; and environmental exposure to toxins [1, 2, 29–31, 71–75]. As identifying underlying mechanisms for associations is certainly complicated [22, 27, 76], researchers should be mindful of this complexity, defining particular mechanisms to be tested, identifying relevant intermediate outcomes such as physical activity and body mass index [77], if appropriate, and selecting the most relevant neighborhood measures accordingly. In general, research on neighborhood factors also comes with its own set of methodological considerations, such as the scale of geography; the definition of "neighborhood" [78]; confounding by proximity, or geospatial autocorrelation; the appropriate multilevel models and methods to use [79]; sources of data; and policy implications [3]. Kawachi and Berkman [29] provide a comprehensive discussion of some of these methodological considerations. The use of cancer registry data also is subject to some limitations, including the lack of information on individual-level SES and residential history. However, these considerations are similar to those impacting research using measures from the CNDS in general, whether in conjunction with individual-level data from the CCR or from other sources. For cancer, contextual factors may be limited, especially for studies of incidence, as they often are measured at a single point in time, while cancer latency is long, and for most cancer sites, the most relevant exposure period is still unknown. In such studies, it may be worthwhile to collect residential history data from individual patients and geocode and link these to historical contextual data [3]. Individual-level factors also may interact with neighborhood-level factors to impact health [17, 69]; findings from recent studies suggest that individual SES (usually measured by education and/or income) and neighborhood SES are both important to breast and prostate cancer risk [65–67]. Future studies also should consider the interactions between individuals' self-reports of their neighborhoods and neighborhood measures based on secondary data, as these may indicate differences in how individuals perceive and use the same neighborhoods. Although individual-level SES information is not currently obtained by US population-based cancer registries, including the CCR, the measures in the CNDS can be linked to other datasets with individual-level SES data, such as from case–control or cohort studies, thus expanding the scope of studies to include multilevel analyses.

The components of any data resource on neighborhood characteristics are dynamic, as neighborhoods can change, and new data can become available, particularly with the increasingly widespread use of GIS. For example, we are currently evaluating the utility of using location data on parks and other points of interest data used in geographic positioning systems (GPS). Accordingly, the measures in the CNDS should be perceived to be dynamic and likely to evolve as new measures are available; for example, we will add to the CNDS measures from the 2010 Census when they become available. Some measures cannot be included because they are not widely available; in the CNDS, we were unable to incorporate a measure of criminal activity, as these data currently only exist at the jurisdiction-, city-, or sometimes even county-level [80, 81], which encompass large and heterogeneous catchment regions for many areas in California.

The CNDS represents an important example of how existing, geographically based data sources collected for varying purposes can be manipulated and combined to any geocoded, individual-level health dataset to address the role of contextual factors on population patterns in disease outcomes. In addition, as we have demonstrated, it can be linked to large-sample population-based cancer surveillance data in order to conduct in-depth studies vital to cancer monitoring, control and prevention. Researchers interested in collaborative work using the CNDS-CCR resource, or in access to the CNDS for research in California, should contact the authors for further information. The use of GIS-based strategies as well as mapping techniques will further improve the use of contextual measures and continue to add value to the disease surveillance systems they complement in future studies. For cancer, these contextual measures, used in conjunction with the routinely collected cancer registry data on patient and tumor characteristics, and treatment data, enable the type of "cells-to-society" research now recognized as important for addressing population disparities in cancer [13].

Acknowledgments

The authors thank Ms. Jane Pham, Ms. June Kristine Winters, Mr. Andrew Hertz, and Dr. Myles Cockburn for their help with aspects of the work related to this manuscript. This study was supported by a grant from the National Cancer Institute (R03 CA117324), and by a Rapid Response Surveillance Study from the SEER program under a modification to contract N01-PC-35136 awarded to the Cancer Prevention Institute of California. The collection of cancer incidence data used in this study was supported by the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute Surveillance, Epidemiology and End Results Program under contract N01-PC-35136 awarded to the Cancer Prevention Institute of California, contract N01-PC-35139 awarded to the University of Southern California, and contract N02-PC-15105 awarded to the Public Health Institute; and the Centers for Disease Control and Prevention National Program of Cancer Registries, under agreement #U55/CCR921930-02 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the authors and endorsement by the State of California, Department of Health Services, the National Cancer Institute, and the Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred.

References

1. Kaplan, GA. School of Public Health. Findings magazine. University of Michigan; 2009. Hourglass Nation. Fall
2. Jackson RJ. The impact of the built environment on health: an emerging field. Am J Public Health. 2003; 93(9):1382–1384. [PubMed: 12948946]
3. Diez Roux, AV. The examination of neighborhood effects on health: conceptual and methodological issues related to the presence of multiple levels of organization. In: Kawachi, I.; Berkman, LF., editors. Neighborhoods and Health. New York: Oxford University Press; 2003. p. 45-64.
4. Barnett E, Caspar M. A definition of “social environment”. Am J Public Health. 2001; 91(3):465. [PubMed: 11249033]
5. Srinivasan S, O’Fallon LR, Dearry A. Creating healthy communities, healthy homes, healthy people: initiating a research agenda on the built environment and public health. Am J Public Health. 2003; 93(9):1446–1450. [PubMed: 12948961]
6. Wikipedia. Built environment. 2010.
7. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: the public health disparities geocoding project (US). J Epidemiol Community Health. 2003; 57(3):186–199. [PubMed: 12594195]
8. Berkman, LF.; Macintyre, S. The measurement of social class in health studies: old measures and new formulations. In: Kogevinas, M.; Pearce, N.; Susser, M.; Boffetta, P., editors. Social inequalities and cancer. International Agency for Research on Cancer; Lyon: 1997. p. 51-64.
9. Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. Am J Epidemiol. 1998; 148(5):475–486. [PubMed: 9737560]
10. 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(5):703–710. [PubMed: 1566949]
11. Krieger N. Socioeconomic data in cancer registries. Am J Public Health. 2001; 91(1):156–157. [PubMed: 11189815]
12. Gehlert, S.; Rebbeck, T.; Lurie, N.; Warnecke, RB.; Paskett, E.; Goodwin, J. Institute NC. Cells to society: overcoming health disparities. Washington: 2007.
13. Warnecke RB, Oh A, Breen N, Gehlert S, Paskett E, Tucker KL, Lurie N, Rebbeck T, Goodwin J, Flack J, Srinivasan S, Kerner J, Heurtin-Roberts S, Abeles R, Tyson FL, Patmios G, Hiatt RA. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. Am J Public Health. 2008; 98(9):1608–1615. [PubMed: 18633099]
14. Krieger N, Fee E. Social class: the missing link in US health data. Int J Health Serv. 1994; 24(1): 25–44. [PubMed: 8150566]
15. Haan M, Kaplan GA, Camacho T. Poverty and health. Prospective evidence from the Alameda County study. Am J Epidemiol. 1987; 125(6):989–998. [PubMed: 3578257]
16. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: a multilevel analysis of Massachusetts births, 1989–1991. Am J Epidemiol. 2006; 164(9):823–834. [PubMed: 16968866]
17. Winkleby M, Cubbin C, Ahn D. Effect of cross-level interaction between individual and neighborhood socioeconomic status on adult mortality rates. Am J Public Health. 2006; 96(12): 2145–2153. [PubMed: 17077398]
18. Winkleby MA, Cubbin C. Influence of individual and neighbourhood socioeconomic status on mortality among black, Mexican-American, and white women and men in the United States. J Epidemiol Community Health. 2003; 57(6):444–452. [PubMed: 12775792]
19. Yen IH, Kaplan GA. Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study. Am J Epidemiol. 1999; 149(10):898–907. [PubMed: 10342798]

20. Diez Roux AV. Investigating neighborhood and area effects on health. Am J Public Health. 2001; 91(11):1783–1789. [PubMed: 11684601]
21. Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. Epidemiol Rev. 2004; 26:104–111. [PubMed: 15234951]
22. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. Soc Sci Med. 2004; 58(10):1953–1960. [PubMed: 15020010]
23. Boardman JD. Stress and physical health: the role of neighborhoods as mediating and moderating mechanisms. Soc Sci Med. 2004; 58:2473–2483. [PubMed: 15081198]
24. Feldman PJ, Steptoe A. How neighborhoods and physical functioning are related: the roles of neighborhood socioeconomic status, perceived neighborhood strain, and individual health risk factors. Ann Behav Med. 2004; 27(2):91–99. [PubMed: 15026293]
25. Foster S, Giles-Corti B. The built environment, neighborhood crime and constrained physical activity: an exploration of inconsistent findings. Prev Med. 2008; 47(3):241–251. [PubMed: 18499242]
26. Leyden KM. Social capital and the built environment: the importance of walkable neighborhoods. Am J Public Health. 2003; 93(9):1546–1551. [PubMed: 12948978]
27. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. Soc Sci Med. 2004; 58(10):1961–1967. [PubMed: 15020011]
28. Sampson, RJ. Neighborhood-level context and health: lessons from sociology. In: KawachiI, BLe, editor. Neighborhoods and health. New York: Oxford University Press; 2003. p. 132-146.
29. Kawachi, I.; Berkman, LF. Neighborhoods and health. Oxford University Press; New York: 2003.
30. Lurie N, Dubowitz T. Health disparities and access to health. JAMA. 2007; 297(10):1118–1121. [PubMed: 17356034]
31. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Rep. 2001; 116:404–416. [PubMed: 12042604]
32. Glaser SL, Clarke CA, Gomez SL, O'Malley CD, Purdie DM, West DW. Cancer surveillance research: a vital subdiscipline of cancer epidemiology. Cancer Causes Control. 2005; 16:1009–1019. [PubMed: 16184466]
33. Howe HL, Edwards BK, Young JL, Shen T, West DW, Hutton M, Correa CN. A vision for cancer incidence surveillance in the United States. Cancer Causes Control. 2003; 14(7):663–672. [PubMed: 14575364]
34. Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare-tumor registry database. Med Care. 1993; 31(8):732–748. [PubMed: 8336512]
35. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care. 2002; 40(8 Suppl):IV-3–18.
36. Clegg LX, Reichman ME, Hankey BF, Miller BA, Lin YD, Johnson NJ, Schwartz SM, Bernstein L, Chen VW, Goodman MT, Gomez SL, Graff JJ, Lynch CF, Lin CC, Edwards BK. Quality of race, Hispanic ethnicity, and immigrant status in population-based cancer registry data: implications for health disparity studies. Cancer Causes Control. 2007; 18:177–187. [PubMed: 17219013]
37. Clegg LX, Reichman ME, Miller BA, Hankey BF, Singh GK, Lin YD, Goodman MT, Lynch CF, Schwartz SM, Chen VW, Bernstein L, Gomez SL, Graff JJ, Lin CC, Johnson NJ, Edwards BK. Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal mortality study. Cancer Causes Control. 2009; 20(4):417–435. [PubMed: 19002764]
38. Chan JK, Gomez SL, O'Malley CD, Perkins CI, Clarke CA. Validity of cancer registry medicaid status against enrollment files: implications for population-based studies of cancer outcomes. Med Care. 2006; 44(10):952–955. [PubMed: 17001267]
39. Reynolds P, Urayama KY, Von Behren J, Feusner J. Birth characteristics and hepatoblastoma risk in young children. Cancer. 2004; 100(5):1070–1076. [PubMed: 14983504]
40. Von Behren J, Reynolds P. Birth characteristics and brain cancers in young children. Int J Epidemiol. 2003; 32(2):248–256. [PubMed: 12714545]

41. Reynolds P, Von Behren J, Gunier RB, Goldberg DE, Harnly M, Hertz A. Agricultural pesticide use and childhood cancer in California. Epidemiology. 2005; 16(1):93–100. [PubMed: 15613951]

42. Reynolds P, Von Behren J, Gunier RB, Goldberg DE, Hertz A. Residential exposure to traffic in California and childhood cancer. Epidemiology. 2004; 15(1):6–12. [PubMed: 14712141]

43. Von Behren J, Reynolds P, Gunier RB, Rull RP, Hertz A, Urayama KY, Kronish D, Buffler PA. Residential traffic density and childhood leukemia risk. Cancer Epidemiol Biomarkers Prev. 2008; 17(9):2298–2301. [PubMed: 18768496]

44. Ward MH, Colt JS, Metayer C, Gunier RB, Lubin J, Crouse V, Nishioka MG, Reynolds P, Buffler PA. Residential exposure to polychlorinated biphenyls and organochlorine pesticides and risk of childhood leukemia. Environ Health Perspect. 2009; 117(6):1007–1013. [PubMed: 19590698]

45. Burns PB, Swanson GM. The occupational cancer incidence surveillance study (OCISS): risk of lung cancer by usual occupation and industry in the Detroit metropolitan area. Am J Ind Med. 1991; 19(5):655–671. [PubMed: 2053580]

46. Swanson GM, Schwartz AG, Brown KL. Population-based occupational cancer incidence surveillance. Utilization of the telephone interview. J Occup Med. 1985; 27(6):439–444. [PubMed: 4020502]

47. Clarke CA, Miller T, Chang ET, Yin D, Cockburn M, Gomez SL. Racial and social class gradients in life expectancy in contemporary California. Soc Sci Med. 2010; 70(9):1373–1380. [PubMed: 20171001]

48. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Yin R, Coull BA. Race/ethnicity and changing US socioeconomic gradients in breast cancer incidence: California and Massachusetts, 1978–2002 (United States). Cancer Causes Control. 2006; 17(2):217–226. [PubMed: 16425100]

49. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. Am J Epidemiol. 2002; 156(5):471–482. [PubMed: 12196317]

50. Liu L, Cozen W, Bernstein L, Ross RK, Deapen D. Changing relationship between socioeconomic status and prostate cancer incidence. J Natl Cancer Inst. 2001; 93(9):705–709. [PubMed: 11333293]

51. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health. 1997; 18:341–378. [PubMed: 9143723]

52. Krieger, N.; Zierler, S.; Hogan, JW.; Waterman, PD.; Chen, J.; Lemieux, K.; Gjelsvik, A. Geocoding and measurement of neighborhood socioeconomic position. Oxford University Press; New York: 2003.

53. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures–the public health disparities geocoding project. Am J Public Health. 2003; 93(10):1655–1671. [PubMed: 14534218]

54. Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. Cancer Causes Control. 2001; 12(8):703–711. [PubMed: 11562110]

55. Chang ET, Shema SJ, Wakelee HA, Clarke CA, Gomez SL. Uncovering disparities in survival after non-small-cell lung cancer among Asian/Pacific Islander ethnic populations in California. Cancer Epidemiol Biomarkers Prev. 2009; 18(8):2248–2255. [PubMed: 19622719]

56. Gomez SL, Clarke CA, Shema SJ, Chang ET, Keegan TH, Glaser SL. Disparities in breast cancer survival among Asian women by ethnicity and immigrant status: a population-based study. Am J Public Health. 2010; 100(5):861–869. [PubMed: 20299648]

57. Gomez SL, Glaser SL. Quality of cancer registry birthplace data for Hispanics living in the United States. Cancer Causes Control. 2005; 16(6):713–723. [PubMed: 16049810]

58. Gomez SL, Glaser SL. Misclassification of race/ethnicity in a population-based cancer registry. Cancer Causes Control. 2006; 17:771–781. [PubMed: 16783605]

59. Gomez SL, Glaser SL, Kelsey JL, Lee MM. Bias in completeness of birthplace data for Asian groups in a population-based cancer registry (United States). Cancer Causes Control. 2004; 15(3):243–253. [PubMed: 15090719]

60. Rhoads KF, Ackerson LK, Jha AK, Dudley RA. Quality of colon cancer outcomes in hospitals with a high percentage of Medicaid patients. J Am Coll Surg. 2008; 207(2):197–204. [PubMed: 18656047]
61. Horner, MJ.; Ries, LAG.; Krapcho, M.; Neyman, N.; Aminou, R.; Howlader, N.; Altekruse, SF.; Feuer, EJ.; Huang, L.; Mariotto, A.; Miller, BA.; Lewis, DR.; Eisner, MP.; Stinchcomb, DG.; Edwards, BK. Institute NC. SEER Cancer Statistics Review, 1975–2006. Bethesda: 2009. http://seer.cancer.gov/csr/1975_2006/, based on November 2008 SEER data submission, posted to the SEER web site
62. Great Circle Distance Calculations. Data Analysis Tools: The North American Association of Central Cancer Registries (NAACCR). 2007.
63. Takaki, R. Strangers from a different shore: a History of Asian Americans. Penguin Books; New York: 1989.
64. Chung, TL-N. Asian American Health: discrepancies, convergence, and enclave-specific trends. In: Abesamis-Mendoza, N.; Ho-Asjoe, H.; Bateman, WB., editors. Handbook of Asian American Health: taking notice and taking action. Greenwood Publishing Group; Santa Barbara: 2010.
65. Robert SA, Strombom I, Trentham-Dietz A, Hampton JM, McElroy JA, Newcomb PA, Remington PL. Socioeconomic risk factors for breast cancer: distinguishing individual-and community-level effects. Epidemiology. 2004; 15(4):442–450. [PubMed: 15232405]
66. Sanderson M, Coker AL, Perez A, Du XL, Peltz G, Fadden MK. A multilevel analysis of socioeconomic status and prostate cancer risk. Ann Epidemiol. 2006; 16(12):901–907. [PubMed: 16843007]
67. Webster TF, Hoffman K, Weinberg J, Vieira V, Aschengrau A. Community- and individual-level socioeconomic status and breast cancer risk: multilevel modeling on Cape Cod, Massachusetts. Environ Health Perspect. 2008; 116(8):1125–1129. [PubMed: 18709175]
68. Datta GD, Colditz GA, Kawachi I, Subramanian SV, Palmer JR, Rosenberg L. Individual-, neighborhood-, and state-level socioeconomic predictors of cervical carcinoma screening among US black women: a multilevel analysis. Cancer. 2006; 106(3):664–669. [PubMed: 16378349]
69. Robert SA. Socioeconomic position and health: the independent contribution of community socioeconomic context. Annu Rev Sociol. 1999; 25:489–516.
70. Warner ET, Gomez SL. Impact of neighborhood racial composition and metropolitan residential segregation on disparities in breast cancer stage at diagnosis and survival between black and white women in California. J Commun Health. 2010
71. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health. 2001; 55(2):111–122. [PubMed: 11154250]
72. Subramanian SV, Chen JT, Rehkopf DH, Waterman PD, Krieger N. Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. Am J Public Health. 2005; 95(2):260–265. [PubMed: 15671462]
73. Yen IH, Syme SL. The social environment and health: a discussion of the epidemiologic literature. Annu Rev Public Health. 1999; 20:287–308. [PubMed: 10352860]
74. MacIntyre S, MacIver S, Sooman A. Area, class, and health: should we be focusing on places or people? J Soc Polic. 1993; 22:213–234.
75. Gee GC, Payne-Sturges DC. Environmental health disparities: a framework integrating psychosocial and environmental concepts. Environ Health Perspect. 2004; 112(17):1645–1653. [PubMed: 15579407]
76. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. Soc Sci Med. 2004; 58(10):1929–1952. [PubMed: 15020009]
77. Keegan THM, Hurley S, Goldberg D, Nelson DO, Reynolds P, Bernstein L, Horn-Ross PL, Gomez SL. The association between neighborhood characteristics and body size and physical activity in the California teachers study cohort. Am J Public Health. (in press).
78. Smith G, Gidlow C, Davey R, Foster C. What is my walking neighbourhood? A pilot study of English adults' definitions of their local walking neighbourhoods. Int J Behav Nutr Phys Act. 2010; 7:34. [PubMed: 20459636]

79. Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Lippman SA, Jewell N, Bruckner T, Satariano WA. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology. 2010; 21(4):467–474. [PubMed: 20220526]
80. RAND. RAND California Community Statistics. http://ca.rand.org/stats/community/community.html
81. State of California Department of Justice. http://ag.ca.gov/crime.php
82. Census 2000 Summary File 3—California. US Census Bureau; 2002.
83. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. Am J Public Health. 2005; 95(2):312–323. [PubMed: 15671470]
84. Lee C, Moudon AV. The 3Ds + R: Quantifying land use and urban form correlates of walking. Transp Res Part D. 2006; 11:204–215.
85. Rural-Urban Commuting Area Codes. United States Department of Agriculture, Economic Research Service. 2000. Available at: http://www.ers.usda.gov/Data/RuralUrbanCommutingAreaCodes/
86. Data Core Library Collection: Segregation Indices Data Series. Rand's Center for Population Health and Health Disparities (CPHHD); 2000.
87. Acevedo-Garcia D, Lochner KA, Osypuk TL, Subramanian SV. Future directions in residential segregation and health research: a multilevel approach. Am J Public Health. 2003; 93(2):215–221. [PubMed: 12554572]
88. Gee GC, Ro A, Shariff-Marco S, Chae D. Racial discrimination and health among Asian Americans: evidence, assessment, and directions for future research. Epidemiol Rev. 2009; 31:130–151. [PubMed: 19805401]
89. Data Core Library Collection. Street Connectivity Data Series. Rand's Center for Population Health and Health Disparities (CPHHD); 2000.
90. Pendola R, Gen S. BMI, auto use, and the urban environment in San Francisco. Health Place. 2007; 13(2):551–556. [PubMed: 16621666]
91. The Annual Utilization Report of Hospitals. Office of Statewide Health Planning and Development; 2001.
92. ReferenceUSA-Business Data. University of Texas; Austin: 2004. Available at: http://www.lib.utexas.edu/indexes/titles.html?id=305
93. ArcGIS Business Analyst Online. Environmental Systems Research Institute (ESRI). Available at: http://www.esri.com/data/esri_data/business.html2004 and 2006
94. Hoehner CM, Schootman M. Concordance of commercial data sources for neighborhood-effects studies. J Urban Health. 2010; 87(4):713–725. [PubMed: 20480397]
95. Lee C, Vernez Moudon A. Correlates of walking for transportation or recreation purposes. J Phys Act Health. 2006; 3(Suppl 1):S77–S98.
96. Moudon AV, Lee C, Cheadle AD, Garvin C, Rd DB, Schmid TL, Weathers RD. Attributes of environments supporting walking. Am J Health Promot. 2007; 21(5):448–459. [PubMed: 17515010]

Table 1

Source and description of socioeconomic and built environment data, California Neighborhoods Data System

Construct	Data source(s)	Measures/indices and level of geography
Neighborhood socioeconomic status and material deprivation	2000 US Census data [82]	Previously validated composite SES measure from Yost et al. [54], created by principal component analysis of seven indicator variables at the census block-group and tract level: Liu education index (out of individuals age 25 and older, proportion with college, high school and less than high school weighted by 16, 12 and 9 respectively) [50], proportion with a blue collar job, proportion older than age 16 in the workforce without a job, median household income, percent below 200% of the poverty line, median rent, median house value. The index is categorized into quintiles with 1 being lowest and 5 the highest. SES variables selected as priorities by the Public Health Disparities Geocoding Project (Krieger et al. [83]): occupation (working class = percent of persons employed in predominantly working class occupations); income (median household income; low income = percent of households with income < 50% of the US median household; high income = percent of households with incomes> = 400% of the US median); poverty (percent of persons below federally-defined poverty line); education (less than high school = percent of persons, age 25 and older, with less than a 12th grade education; college = percent of persons, age 25 and older, with at least 4 years of college; Liu education index; crowding (% households with > = 1 person per room). Townsend Index (UK deprivation measure consisting of a standardized z score combining data on percent crowding, percent unemployment, percent no car ownership, and percent renters). Level of geography = Census block-group and tract
Population density	2000 US Census data [82]	Population counts divided by land area. Can be used as indicator of built environment [84] and crowding. Level of geography = Census block-group and tract
Rural/urban measure	US Department Agriculture-Economic Research Service (2003), based on 2000 Census data [85]	The rural–urban commuting area (RUCA) codes classify U.S. census tracts using measures of population density, urbanization, and daily commuting. The classification contains two levels. Whole numbers [1–10] delineate metropolitan, micropolitan, small town, and rural commuting areas based on the size and direction of the primary (largest) commuting flows. These 10 codes are further subdivided to permit stricter or looser delimitation of commuting areas, based on secondary (second largest) commuting flows. The approach errs in the direction of more codes, providing flexibility in combining levels to meet varying definitional needs and preferences. Categories used include metropolitan (defined as urbanized areas of 50,000 people or more with a density of 1,000 people or more per square mile as well as outlying areas where at least 5% of workers commute into an urbanized area); micropolitan (defined as large urban clusters of at least 10,000 but less than 50,000 people or outlying areas where at least 10% of workers commute into a large urban cluster); small towns (defined as small urban clusters of 2,500–9,999 people or outlying areas where at least 10% of workers commute into a small urban cluster); and rural areas (any remaining areas not considered an urbanized area or urban cluster). Level of geography = Census tract
Racial/ethnic composition	2000 US Census data [82]	Racial/ethnic distribution of Census block-group or tract. Predominant race defined as > 50% non-Hispanic (NH) White, NH Black, Hispanic, NH Asian/PI, or mixed, if not any of the previous. Level of geography = Census block-group and tract
Racial/ethnic residential segregation	Center for Population Health and Health Disparities (CPHHD) [13] A RAND Health Program [86]	As a form of institutional discrimination [31, 87, 88], measures of segregation calculated for NH blacks, NH Asians/Pacific Islanders (API), Hispanics, relative to NH Whites: Atkinson Index, Absolute Centralization Index, Absolute Clustering Index, Dissimilarity Index, Duncan's Decay Isolation Index, Distance Decay Interaction Index, Gini Index, Entropy Index, Relative Centralization Index, Relative Clustering Index, Relative Concentration Index, Spatial Proximity Index, Correlation Ratio, Isolation Index, and Interaction Index. These specific measures and possible mechanisms relating to health outcomes are defined in several sources [31, 70, 87, 88]. Level of geography = Census places with at least 10 census tracts
Housing environment	2000 US Census data [82]	Housing variables at the block group level from census data as described by: % of non-single family units; % of housing structures with more than

Construct	Data source(s)	Measures/indices and level of geography
		10 units; % crowded (defined as % of houses with ≥1 person per room); % of households renting. Level of geography = Census block-group and tract
Street connectivity	Center for Population Health and Health Disparities (CPHHD) [13] A RAND Health Program [89]	Increasingly used as a measure of walkability [84]. Street network based measures calculated per census tract, including ratio of street segments to intersections; median and average block size (sq ft); median and average length of street blocks (ft); gamma measure (ratio of actual number street segments to maximum possible given the number of intersections); and alpha measure (ratio of the actual number of complete loops to the maximum number of possible loops given the number of intersections). Level of geography = Census tract
Neighborhood immigration/acculturation	2000 US Census data [82]	Selected measures to capture immigration and acculturation characteristics among residents: % foreign-born; % recently immigrated (defined as immigration from 1995 to 2000); % with poor English language proficiency (defined as those who speak English “not well” or “not at all”); % linguistically isolated households (defined as households in which no person 14 years old and over speaks only English and no person 14 years old and over who speaks a language other than English speaks English “very well”). Composite “ethnic enclave” measure created by principal component analysis of indicator variables at the Census block-group and tract level. An ethnic enclave is geographical unit that maintains more cultural mores and is culturally and/or ethnically distinct from its surrounding area. The index is categorized into quintiles with 1 representing the least ethnic (or most acculturated) and 5 the most ethnic (or least acculturated). *Hispanic-specific index*: % foreign-born; % recent immigrants; % households that are linguistically isolated; % of Spanish language-speaking households that are linguistically isolated; % all language speakers with limited English proficiency; % of Spanish language-speakers with limited English proficiency; % Hispanic. *Asian/Pacific Islander (API)-specific index*: % recent immigrants; % of API language-speaking households that are linguistically isolated; % of API language-speakers with limited English proficiency; % API. Level of geography = Census block-group and tract
Commute time and method of commute	2000 US Census data [82]	Commute patterns: % of residents commuting various times to work (< 30 m, 30–44 m, 45–59 m, 60 m+); % traveling by car or motorcycle, public transportation, walking or biking, other, work from home. Used as measure of built environment [90]. Level of geography = Census block-group and tract
Hospitals	Office of Statewide Health and Policy Development [91]	Distance to hospital where cancer was diagnosed and hospital characteristics: number of beds (< 200 vs 200+ beds), ownership status (public, health maintenance organization (HMO), University of California/private), academic and teaching status, affiliation with NIH-designated cancer center, racial/ethnic and (neighborhood) SES distribution of cancer cases diagnosed at that hospital. Used to indicate hospital characteristics, as well as access to care. Level of geography = point-source
Neighborhood business and services	Addresses and types of businesses from ReferenceUSA (2002) [92];Addresses and types of businesses from InfoUSA database bundled with ESRI Business Analyst (2004 and 2006) [93]; time series data (1990?) also available from Dun and Bradstreet [94]	Number of businesses in selected categories previously shown or theorized to be associated with built environment (e.g., walkability) [84, 95, 96], and socioeconomic environment (i.e., desirable and undesirable social resources/amenities) [19]: food stores (supermarkets, convenience stores, and specialty food stores), eating places (restaurants, coffee shops, and bars), alcohol outlets, retail outlets (drugstores, video stores, hardware stores, and other retail outlets), services (banks, post offices, libraries, salons, and laundromats), and the total number of business or destinations (includes food stores, eating places, alcohol outlets, retail outlets, services, entertainment outlets, recreation facilities, and parks). Level of geography = point-source

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Distribution of socioeconomic and built environment measures (definitions in Table 1) across California neighborhoods (i.e., Census tracts) overall and by socioeconomic status, 2000 (n = 7,049 census tracts)

Environment measure	Statewide	SES quintiles (*n* census tracts)				
		1 (low)	2	3	4	5 (high)
		n = 1,409	*n* = 1,410	*n* = 1,410	*n* = 1,410	*n* = 1,410
Housing						
% crowded (category)						
0–4.9%	28.3	1.1	6.0	18.8	42.6	72.9
5–9%	20.3	2.4	16.7	29.2	33.4	20.2
10–19%	20.0	11.7	31.0	32.5	18.5	6.3
≥20%	31.4	84.9	46.3	19.5	5.6	0.6
% crowded (continuous)	10.6	39.6	18.5	10.5	5.8	2.8
% non-single family units of total units	28.6	40.4	34.8	28.5	23.3	12.3
% housing structures w/10+units	8.5	11.5	10.1	9.1	8.8	4.0
% renting	38.8	62.6	44.9	35.8	31.4	19.1
Population density						
Mean population density (per sq mile) (continuous)	5,905	10,221	6,820	5,754	5,481	4,322
Population density (per sq mile) (category)						
% 1st quartile (low)	25.0	17.4	28.2	27.1	23.5	27.6
% 2nd quartile	25.0	15.9	16.6	24.5	30.8	37.3
% 3rd quartile	25.0	17.0	25.9	29.4	28.9	24.4
% 4th quartile (high)	25.0	49.8	29.2	19.0	16.8	10.7
Urbanization (%)						
Metropolitan	92.6	90.6	85.9	90.8	97.2	99.2
Micropolitan	4.2	5.6	7.6	5.5	2.1	0.4
Small town	1.2	1.7	2.2	1.7	0.3	0.0
Rural	2.0	2.1	4.3	2.0	0.4	0.5
Transport and travel time to work						
% travel to work by car/motorcycle	89.6	83.9	89.5	90.8	91.0	89.9
% travel to work by public transportation	2.5	7.0	3.0	2.2	1.9	1.5
% travel to work by foot/bicycle	2.4	4.0	2.9	2.2	1.8	1.3

Environment measure	Statewide	SES quintiles (*n* census tracts)				
		1 (low)	2	3	4	5 (high)
		n = 1,409	*n* = 1,410	*n* = 1,410	*n* = 1,410	*n* = 1,410
% travel to work by other means	0.5	1.1	0.7	0.5	0.4	0.4
% work at home	3.2	2.1	2.5	3.0	3.7	5.5
% traveled < 30 min	60.0	59.8	61.2	61.6	59.9	57.2
% traveled 30–44 min	21.4	22.7	20.3	20.2	21.0	22.6
% traveled 45–59 min	7.7	6.9	7.0	7.5	8.3	9.3
% traveled 60 + mins	8.9	9.7	8.7	8.5	8.4	8.8
Street Connectivity						
Mean ratio streets/intersections	1.30	1.40	1.32	1.28	1.25	1.21
Median size of street blocks (sq ft)	269,558.2	239,312.2	261,463.5	275,530.1	279,204.3	300,579.4
Median length of street blocks (feet)	2,421.7	2,150.2	2,368.2	2,455.6	2,505.0	2,612.3
Mean size of street blocks (sq ft)	551,442.9	385,158.7	537,239.1	589,942.6	610,621.0	692,197.1
Mean length of street blocks (feet)	3,209.0	2,641.6	3,139.9	3,287.5	3,365.0	3,694.0
Mean gamma measure	0.44	0.48	0.45	0.43	0.43	0.41
Mean alpha measure	0.16	0.21	0.17	0.15	0.13	0.11
Race/ethnicity, immigration						
% foreign-born (category)						
0–9%	17.8	3.9	16.0	26.0	25.2	17.5
10–24%	38.5	16.2	35.1	38.7	45.2	57.4
25–49%	34.4	52.2	40.0	30.8	26.4	23.0
≥50%	9.3	27.6	9.0	4.6	3.3	2.1
% foreign-born (continuous)	21.8	40.2	24.5	18.1	16.4	16.7
% recent immigrants	15.9	18.7	17.0	15.1	14.5	13.4
% with limited English proficiency	6.3	24.2	10.8	5.5	3.4	2.4
% non-Hispanic White	51.5	9.3	35.8	53.7	65.3	75.6
% non-Hispanic Black	2.4	3.7	3.5	3.0	2.4	1.1
% Hispanic	21.9	67.1	37.0	23.6	13.9	7.4
% non-Hispanic Asian/Pacific Islander	6.1	2.7	5.2	5.9	8.0	9.6
% non-Hispanic Other Race	3.4	1.9	3.6	3.8	3.9	3.3
Ethnic enclave index						

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Environment measure	**Statewide**	**SES quintiles (*n* census tracts)**				
		1 (low)	**2**	**3**	**4**	**5 (high)**
		***n* = 1,409**	***n* = 1,410**	***n* = 1,410**	***n* = 1,410**	***n* = 1,410**
Hispanic						
% in 1st quintile (low enclave)	20.0	2.0	9.9	18.4	26.5	43.2
% in 2nd quintile	20.0	2.3	9.3	21.0	31.1	36.4
% in 3rd quintile	20.0	5.9	18.7	26.5	30.8	18.2
% in 4th quintile	20.0	21.4	36.4	29.2	10.8	2.1
% in 5th quintile (high enclave)	20.0	68.5	25.7	4.9	0.8	0.0
Asian/Pacific Islander						
% in 1st quintile (low enclave)	20.0	35.2	21.9	21.1	12.9	8.9
% in 2nd quintile	20.0	19.2	20.4	19.9	20.9	19.6
% in 3rd quintile	20.0	15.1	20.4	19.8	20.3	24.4
% in 4th quintile	20.0	14.7	19.6	19.8	22.3	23.6
% in 5th quintile (high enclave)	20.0	15.8	17.8	19.4	23.6	23.4

For continuous variables, the median value is presented

Table 3

Distribution of socioeconomic and built environment measures (definitions in Table 1) across California neighborhoods (i.e., Census tracts) by predominant race/ethnicity of Census tract, 2000 (n = 7,049 census tracts)

Environment measure	**Predominant race/ethnicity (*n* census tracts)**				
	NH White	**NH Black**	**Hispanic**	**NH API**	**Mixed**
	***n* = 3,628**	***n* = 135**	***n* = 1,607**	***n* = 187**	***n* = 1,492**
Socioeconomic status					
% age 25 + years without high school diploma	9.8	27.6	52.3	17.7	24.1
% age 25 + years with college degree	41.9	20.5	10.6	43.2	25.7
Education index	14.5	13.1	11.5	14.0	13.3
% below 200% poverty level	16.9	41.7	55.6	18.3	32.0
Median household income ($)	$57,750	$32,689	$32,689	$65,737	$45,035
% workers age 16 + years unemployed	2.7	5.9	5.8	2.7	4.2
% blue collar workers	9.5	10.5	17.6	10.1	13.0
Median gross rent ($)	$908	$677	$634	$1,097	$791
Median value of owner-occupied houses ($)	$234,300	$157,500	$146,900	$299,500	$173,600
Socioeconomic status index					
% in 1st quintile (low)	3.1	27.4	64.4	3.8	14.6
% in 2nd quintile	13.2	39.3	26.1	16.1	28.9
% in 3rd quintile	21.4	23.7	8.6	15.6	29.1
% in 4th quintile	28.6	8.2	0.8	25.3	20.2
% in 5th quintile (high)	33.7	1.5	0.1	39.3	7.3
Housing					
% crowded (category)					
0–4.9%	52.2	5.9	0.4	7.5	4.6
5–9%	32.1	18.5	0.5	12.9	14.1
10–19%	14.6	37.8	6.1	33.9	45.1
≥20%	1.1	37.8	93.0	45.7	36.2
% crowded (continuous)	4.8	17.4	39.7	19.1	16.9
% non-single family units of total units	23.5	36.7	35.4	17.6	36.2
% housing structures w/10 + units	5.5	8.7	11.6	4.3	13.2
% renting	29.2	48.7	56.8	28.2	46.6

Environment measure	Predominant race/ethnicity (*n* census tracts)				
	NH White	**NH Black**	**Hispanic**	**NH API**	**Mixed**
	***n* = 3,628**	***n* = 135**	***n* = 1,607**	***n* = 187**	***n* = 1,492**
Population density					
Mean population density (per sq mile) (continuous)	3,693	11,438	10,683	10,124	7,703
Population density (per sq mile) (category)					
% in 1st quartile (low)	36.4	5.2	13.3	7.0	14.1
% in 2nd quartile	32.7	8.9	12.7	16.0	22.1
% in 3rd quartile	22.5	25.2	22.0	27.8	33.9
% in 4th quartile (high)	8.4	60.7	52.1	49.2	29.9
Urbanization (%)					
Metropolitan	89.8	100.0	93.5	100.0	96.6
Micropolitan	5.6	0.0	4.4	0.0	1.5
Small town	1.7	0.0	0.9	0.0	0.7
Rural	2.9	0.0	1.2	0.0	1.2
Transport and travel time to work					
% travel to work by car/motorcycle	90.4	84.6	85.9	91.1	90.0
% travel to work by public transportation	1.4	10.4	6.1	3.5	3.5
% travel to work by foot/bicycle	2.0	1.6	3.6	1.3	2.2
% travel to work by other means	0.4	0.5	1.0	0.4	0.5
% work at home	4.5	2.2	1.8	2.2	2.5
% traveled < 30 min	62.1	51.4	58.3	53.6	59.3
% traveled 30–44 min	19.0	27.1	23.6	27.4	21.8
% traveled 45–59 min	7.5	9.8	7.4	9.5	8.2
% traveled 60 + mins	8.3	10.2	9.4	9.2	9.2
Street Connectivity					
Mean ratio streets/intersections	1.2	1.4	1.4	1.3	1.3
Median size of street blocks (sq ft)	292,025.8	223,925.3	244,971.5	262,068.7	265,069.1
Median length of street blocks (feet)	2,561.5	2,058.8	2,214.0	2,380.5	2,398.9
Mean size of street blocks (sq ft)	747,479.6	304,009.9	407,609.5	443,957.2	492,676.4
Mean length of street blocks (feet)	3,638.2	2,406.5	2,742.5	3,006.0	3,084.2
Mean gamma measure	0.42	0.50	0.47	0.43	0.44

Environment measure	Predominant race/ethnicity (*n* census tracts)				
	NH White	**NH Black**	**Hispanic**	**NH API**	**Mixed**
	***n* = 3,628**	***n* = 135**	***n* = 1,607**	***n* = 187**	***n* = 1,492**
Mean alpha measure	0.13	0.24	0.20	0.15	0.16
Race/ethnicity, immigration					
% foreign-born (category)					
0–9%	32.6	20.0	0.4	0.0	2.4
10–24%	55.0	60.7	7.6	0.0	34.4
25–49%	12.0	19.3	65.0	42.8	56.5
≥50%	0.4	0.0	27.1	57.2	6.7
% foreign-born (continuous)	13.1	17.9	42.4	51.8	29.5
% recent immigrants	14.0	15.4	18.1	16.0	16.9
% with limited English proficiency	2.6	7.5	23.8	15.4	9.6
% non-Hispanic White	73.1	2.7	10.6	16.6	36.8
% non-Hispanic Black	1.5	62.1	2.6	2.0	6.7
% Hispanic	11.7	21.9	70.2	12.7	32.9
% non-Hispanic Asian/Pacific Islander	5.2	2.5	3.6	57.9	16.2
% non-Hispanic Other Race	3.7	3.0	1.7	2.8	4.2
Ethnic enclave index					
Hispanic					
% in 1st quintile (low enclave)	37.2	19.3	0.1	0.5	2.0
% in 2nd quintile	34.3	19.3	0.3	7.0	8.1
% in 3rd quintile	22.6	17.0	1.3	41.4	31.7
% in 4th quintile	5.8	42.2	22.1	38.7	48.2
% in 5th quintile (high enclave)	0.2	2.2	76.2	12.4	10.0
Asian/Pacific Islander					
% in 1st quintile (low enclave)	22.6	40.0	28.2	0.0	5.4
% in 2nd quintile	25.2	21.5	19.9	0.0	9.7
% in 3rd quintile	24.5	19.3	17.4	0.5	14.2
% in 4th quintile	19.5	12.6	19.2	0.5	25.2
% in 5th quintile (high enclave)	8.1	6.7	15.4	98.9	45.4

For continuous variables, the median value is presented

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Distribution of socioeconomic and built environment measures among colorectal cancer patients by race/ethnicity, 1996–2004 diagnoses, California (*n* = 128,184)

Environment measure	**Race/ethnicity (*n* patients)**			
	NH White	**NH Black**	**Hispanic**	**NH API**
	***n* = 90,638**	***n* = 8,913**	***n* = 16,399**	***n* = 12,234**
Socioeconomic status				
% age 25 + years without high school diploma	13.0	29.9	30.1	18.2
% age 25 + years with college degree	35.5	19.2	19.7	34.7
Education index	14.2	12.9	12.8	13.9
% below 200% poverty level	20.9	41.4	36.4	22.9
Median household income ($)	$51,538	$36,431	$42,333	$53,090
% workers age 16 + years unemployed	3.0	5.4	4.4	3.1
% blue collar workers	10.5	12.7	14.1	11.0
Median gross rent ($)	$852	$686	$740	$899
Median value of owner-occupied houses ($)	$205,800	$153,500	$163,025	$228,150
Socioeconomic status index				
% in 1st quintile (low)	8.3	33.8	30.6	12.5
% in 2nd quintile	17.1	25.7	25.3	18.2
% in 3rd quintile	23.0	20.0	19.7	19.5
% in 4th quintile	24.9	13.7	14.7	23.6
% in 5th quintile (high)	26.7	6.9	9.7	26.2
Housing				
% crowded (category)				
0–4.9%	41.0	11.2	13.4	19.5
5–9%	24.9	15.9	14.1	19.4
10–19%	19.6	24.5	21.1	24.2
≥20%	14.5	48.5	51.5	36.9
% crowded (continuous)	6.3	19.3	20.8	14.2
% non-single family units of total units	27.1	33.3	28.6	28.7
% housing structures w/10 + units	8.0	9.5	9.5	10.9
% renting	32.6	49.1	42.6	38.7

Environment measure	Race/ethnicity (*n* patients)			
	NH White	**NH Black**	**Hispanic**	**NH API**
	***n* = 90,638**	***n* = 8,913**	***n* = 16,399**	***n* = 12,234**
Population Density				
Population density (per sq mile) (category)				
% in 1st quartile (low)	29.2	9.7	17.3	10.9
% in 2nd quartile	30.8	17.3	21.5	21.2
% in 3rd quartile	25.2	25.0	29.5	28.8
% in 4th quartile (high)	14.8	48.0	31.7	39.1
Urbanization (%)				
Metropolitan	91.1	98.8	93.8	98.6
Micropolitan	5.8	0.9	4.5	1.2
Small town	1.6	0.3	1.0	0.1
Rural	1.4	0.0	0.7	0.2
Transport and travel time to work				
% travel to work by car/motorcycle	90.3	86.6	89.8	89.4
% travel to work by public transportation	1.8	6.8	3.2	3.9
% travel to work by foot/bicycle	2.1	2.1	2.6	2.1
% travel to work by other means	0.5	0.6	0.7	0.5
% work at home	3.9	2.5	2.5	2.8
% traveled < 30 min	61.3	54.2	59.4	57.1
% traveled 30–44 min	20.0	24.7	22.0	23.8
% traveled 45–59 min	7.7	8.9	7.8	8.7
% traveled 60 + min	8.7	10.3	9.2	9.1
Street connectivity				
Mean Total street segments	203	130	161	145
Mean Total intersections	156	92	120	110
Mean Ratio streets/intersections	1.3	1.4	1.3	1.3
Median size of street blocks (sq ft)	281,855.1	239,088.2	258,836.9	254,985.6
Median length of street blocks (feet)	2,495.4	2,217.9	2,344.7	2,340.5
Mean size of street blocks (sq ft)	650,980.8	366,243.7	486,710.4	457,651.4
Mean length of street blocks (feet)	3,394.6	2,638.5	3,020.0	2,981.4

Environment measure	Race/ethnicity (*n* patients)			
	NH White	**NH Black**	**Hispanic**	**NH API**
	***n* = 90,638**	***n* = 8,913**	***n* = 16,399**	***n* = 12,234**
Mean gamma measure	0.43	0.48	0.45	0.45
Mean alpha measure	0.14	0.21	0.18	0.16
Businesses and destinations (mean)				
Number of food stores w/in 1,600 m[a]	11	26	19	20
Number of food stores w/in 400 m[a]	0	1	1	1
Number of eating places w/in 1,600 m[b]	27	39	38	46
Number of eating places w/in 400 m[b]	0	1	1	1
Number of alcohol outlets w/in 1,600 m[c]	3	7	5	5
Number of alcohol outlets w/in 400 m[c]	0	0	0	0
Number of retail outlets w/in 1,600 m[d]	63	98	89	102
Number of retail outlets w/in 400 m[d]	2	3	3	3
Number of services w/in 1,600 m[e]	26	45	33	40
Number of services w/in 400 m[e]	0	2	1	1
Total number of businesses or destinations w/in 1,600 m[f]	140	228	192	226
Total number of businesses or destinations w/in 400 m[f]	3	9	6	8
Race/ethnicity, immigration				
% foreign-born (category)				
0–9%	24.7	10.8	8.5	4.7
10–24%	46.1	37.8	30.1	25.2
25–49%	25.4	46.3	48.0	49.4
≥50%	3.8	5.1	13.4	20.7
% foreign-born (continuous)	16.5	25.3	30.6	35.6
% recent immigrants	14.7	16.0	15.9	15.9
% with limited English proficiency	3.8	10.4	12.2	10.4
% non-Hispanic White	67.6	11.5	29.1	34.2
% non-Hispanic Black	1.7	23.7	2.5	2.7
% Hispanic	14.4	30.5	44.0	19.6

Environment measure	Race/ethnicity (*n* patients)			
	NH White	**NH Black**	**Hispanic**	**NH API**
	***n* = 90,638**	***n* = 8,913**	***n* = 16,399**	***n* = 12,234**
% non-Hispanic Asian/Pacific Islander	5.4	4.9	5.9	21.9
% non-Hispanic Other Race	3.6	3.2	2.9	3.4
Ethnic enclave index				
Hispanic				
% in 1st quintile (low enclave)	29.2	10.9	8.0	9.0
% in 2nd quintile	25.9	15.4	12.2	16.8
% in 3rd quintile	22.2	18.0	17.1	27.7
% in 4th quintile	16.0	32.7	27.7	28.0
% in 5th quintile (high enclave)	6.8	23.0	34.6	18.7
Asian/Pacific Islander				
% in 1st quintile (low enclave)	20.0	28.8	18.8	2.9
% in 2nd quintile	24.1	17.5	20.3	7.7
% in 3rd quintile	22.2	18.0	19.9	11.8
% in 4th quintile	18.9	17.1	20.5	20.8
% in 5th quintile (high enclave)	14.9	18.7	20.5	56.8
Hospital Characteristics				
Size				
% < 200 beds	25.3	12.9	23.1	21.1
% ≥200 beds	65.8	78.1	69.2	70.7
% Unknown	8.9	9.0	7.8	8.2
Ownership Status				
% Private	27.8	26.2	26.6	29.8
% Public	47.8	36.5	48.8	46.2
% HMO	15.6	28.3	16.8	15.9
% Unknown	8.9	9.0	7.8	8.2
Racial/Ethnic Distribution				
% White	78.5	56.5	68.4	62.3
% Black	2.4	12.6	4.3	4.6
% Hispanic	9.3	10.9	15.3	13.0

Environment measure	Race/ethnicity (*n* patients)			
	NH White	**NH Black**	**Hispanic**	**NH API**
	***n* = 90,638**	***n* = 8,913**	***n* = 16,399**	***n* = 12,234**
% Asian/Pacific Islander	4.3	6.2	6.2	12.7
% Other/Unknown	0.3	0.4	0.3	0.3
Mean distance to reporting hospital (greater circle distance miles)	4.4	4.7	4.7	4.5

For continuous variables, the median value is presented

[a]Includes supermarkets, specialty food stores, convenience stores, and liquor stores

[b]Includes restaurants, coffee shops, and bars

[c]Includes liquors stores and bars

[d]Includes drugstores, video stores, hardware stores, and other retail outlets

[e]Includes banks, post offices, libraries, salons, and laundry mats

[f]Includes food stores, eating places, alcohol outlets, retail outlets, services, entertainment outlets, recreation facilities, and parks