BMJ Open

Cross-sectional analysis of place-based and racial disparities in hospitalisation rates by disease category in California in 2001 and 2011

Eva Raphael ,[1] R Gaynes,[2] Rita Hamad[1,3]

To cite: Raphael E, Gaynes R, Hamad R. Cross-sectional analysis of place-based and racial disparities in hospitalisation rates by disease category in California in 2001 and 2011. *BMJ Open* 2019;**9**:e031556. doi:10.1136/bmjopen-2019-031556

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-031556).

Received 09 May 2019
Revised 09 September 2019
Accepted 24 September 2019

[1]Family and Community Medicine, University of California San Francisco, San Francisco, California, USA
[2]Department of Medicine, Division of Infectious Diseases, Emory University, Atlanta, Georgia, USA
[3]Philip R Lee Institute for Health Policy Studies, University of California, San Francisco, San Francisco, California, USA

Correspondence to
Dr Eva Raphael;
eva.raphael@ucsf.edu

ABSTRACT

Objectives To study the association of place-based socioeconomic factors with disease distribution by comparing hospitalisation rates in California in 2001 and 2011 by zip code median household income.

Design Serial cross-sectional study testing the association between hospitalisation rates and zip code-level median income, with subgroup analyses by zip code income and race.

Participants/setting Our study included all hospitalised adults over 18 years old living in California in 2001 and 2011 who were not pregnant or incarcerated. This included all acute-care hospitalisations in California including 1632 zip codes in 2001 and 1672 zip codes in 2011.

Primary and secondary outcomes We compared age-standardised hospitalisations per 100 000 persons, overall and for several disease categories.

Results There were 1.58 and 1.78 million hospitalisations in California in 2001 and 2011, respectively. Spatial analysis showed the highest hospitalisation rates in urban inner cities and rural areas, with more than 5000 hospitalisations per 100 000 persons. Hospitalisations per 100 000 persons were consistently highest in the lowest zip code income quintile and particularly among black patients.

Conclusion Hospitalisation rates rose from 2001 to 2011 among Californians living in low-income and middle-income zip codes. Integrating spatially defined state hospital discharge and federal zip code income data provided a granular description of disease burden. This method may help identify high-risk areas and evaluate public health interventions targeting health disparities.

Strengths and limitations of this study

- ► This is among the first statewide assessment of disease burden based on linking state hospitalisation data with federal zip code income data.
- ► This observational cross-sectional study is based on over 3 million hospitalisations in 2001 and 2011 combined.
- ► The study was limited by missing demographic data on race and gender, as much as 20% in some zip code income quintiles.

INTRODUCTION

Neighbourhood-level poverty has been associated with poor health outcomes.[1–3] Most studies focus on the association of neighbourhood-level poverty with specific diseases or mortality,[4] but few explore its effect on healthcare use.[5] Hospitalisation rates, a commonly used measure of morbidity, have been shown to be proxies for poor health, as opposed to outpatient clinic visits which may indicate appropriate treatment of preventable health conditions.[6] Hospitalisations represent costly encounters, both high in financial expense and in morbidity. Hospital-based data are routinely collected as measures of healthcare quality, use and expenditure, and are an underused data resource in studies of the social determinants of health.[6]

Multiple place-based factors have been identified as possible mediators of health outcomes.[7] These include environmental exposures, the physical environment, walkability, access to healthy foods and the social environment.[7] Low area-level income may also impact health, with one prior experimental study showing decreased rates of obesity and diabetes in low-income individuals moving from high-poverty to low-poverty neighbourhoods.[8] While factors that influence trends in hospitalisations include technological advances and reimbursement policies, the association of neighbourhood-level income and hospitalisations may fluctuate over time due to economic downturns and uneven implementation of public health policies that may render certain groups more vulnerable.

Understanding place-based differences in health—including healthcare use—is key to designing and implementing public health interventions. In this study, we therefore assessed population-level burden of disease based on hospitalisation rates and income in California, the most populous state in

the USA. We linked state-level hospitalisation data with income data from the US Census Bureau to compare temporal and geographic differences in hospitalisation rates and zip code income in 2001 and 2011. We hypothesised that lower zip code-level income is associated with increased hospitalisation rates and that they differed between 2001 and 2011. This study provides the granularity of data needed to identify at-risk subpopulations and areas in California. It also demonstrates the use of hospital discharge data in place-based population research.

METHODS

Outcomes

In this serial cross-sectional study, our primary outcome measures were hospitalisation rates (per 100 000) for specific disease conditions. These were constructed based on data from the California Office of Statewide Health Planning and Development (OSHPD), which releases individual-level, deidentified hospital discharge data annually by county and zip code. OSHPD data include demographic information (age, race, gender), up to 25 diagnosis codes following the International Classification of Diseases (ICD) and duration of hospitalisation. Total charges accrued are also reported; they encompass all service charges during the hospitalisation, including daily hospital services and ancillary services, excluding physician charges. We compared OSHPD data for 2001 and 2011. The study was conducted with anonymised data and involved no contact with patients.

All acute care hospitalisations of persons aged 18 years or older whose residential five-digit zip codes were in California were included. Non-acute admissions to skilled nursing facilities or inpatient hospice were excluded as were hospitalisations of pregnant women and incarcerated adults. These data were tabulated by federal zip code tabulation areas, which are geographical representations of zip codes.[9] For simplicity, we refer to these as zip codes throughout the manuscript. Zip code-level aggregation has been used in prior studies, with comparable heterogeneity as census tract.[10 11]

California adult population estimates by income category and race were obtained from the US Census Bureau. For each disease category and zip code income group, we obtained age-standardised hospitalisations by summing hospitalisations for each age group per 100 000 persons within each racial category within that age group adjusted by the percentage of California population in each age group (box 1).

Exposure

Our primary exposure variable was median household income at the zip code level, divided into quintiles. Population and zip code median household income estimates were obtained from the 2000 US Census and the 2011 5-year American Community Survey (2007–2011) from the US Census Bureau.[12]

Box 1 Formula for age-standardised hospitalisation rate calculation

Age-standardised hospitalisation rate=(number of hospitalisations in age group 1*×%CA population in age group 1×100 000)/CA population of age group 1* +(number of hospitalisations in age group 2*×%CA population in age group 2×100 000)/CA population of age group 2*+(number of hospitalisations in age group 3*×%CA population in age group 3×100000)/CA population of age group 3*

*When calculated for racial categories, number of hospitalisations and California population for each age group were obtained for each racial category separately. Note: age group 1 includes individuals 18–34 years old, age group 2 includes individuals 35–64 years old, and age group 3 includes individuals 65 and older. CA, California.

Data analysis

First, we compared hospitalisation statistics for 2001 and 2011 (table 1). We included total number of hospitalisations and age-standardised hospitalisations per 100 000 person-year, mean length of stay (days) and cost of those hospitalisations per population within individual zip codes. We then reported demographic characteristics for all hospitalisations.

Next, we created income quintiles based on the distribution of annual median household income of all California zip codes from the US Census Bureau for 2000 and 2011. For 2000, the income quintiles were the following: first quintile, US$0–US$31 002; second quintile, US$31 003–US$38 697; third quintile, US$38 698–US$47 810; fourth quintile, US$47 811–US$61 766 and fifth quintile, above US$61 767. For 2011, the income quintiles were: first quintile, US$0–US$39 442; second quintile, US$39 443–US$51 644; third quintile, US$51 645–US$64 250; fourth quintile, US$64 251–US$82 633 and fifth quintile, above US$82 634. We stratified hospitalisation rates by zip code income quintile and race for both 2001 and 2011 (table 2).

Secondary analysis

In a secondary analysis, we also examined differences in hospitalisation rates by disease categories (see online supplementary file 1). The diagnostic codes for each hospitalisation were classified into the 19 disease categories used in the 2010 Global Burden of Disease Survey based on a mapping of ICD-9-CM codes (see online supplementary file 2).[13–15] We mapped hospitalisation rates for the most common disease categories (see online supplementary file 3) and reported hospitalisation rates by zip code income quintile including race for all disease categories (see online supplementary files 4-6).

Missing data

Race data were missing in about 20% of hospitalisations. Sex data were missing in about 10% of hospitalisations. Observations with missing values for race or sex were omitted from rate calculations, as were the hospitalisations which were not matched to a California zip code. Out of the 1 581 543 and 1 783 145 hospitalisations in 2001 and

Table 1 Demographic characteristics for all hospitalisations, California 2001–2011

	2001 N (%)	2011 N (%)
Total hospitalisations	1 581 543	1 783 145
Age-standardised hospitalisations per 100 000 person-year	6339	6315
Mean length of stay (days)	4.9	4.7
Hospital charge per population (2011 US dollars)	2143	3828
Gender		
Male	640 924 (40.53)	742 833 (41.66)
Female	755 189 (47.75)	846 160 (47.45)
Unknown gender	185 430 (11.72)	194 152 (10.89)
Race		
White	890 060 (56.28)	865 693 (48.55)
Black	104 152 (6.59)	126 714 (7.11)
Latino	194 002 (12.27)	303 554 (17.02)
Asian	62 253 (3.94)	90 878 (5.10)
Native American	1343 (0.08)	1808 (0.10)
Other race	15 103 (0.95)	32 218 (1.81)
Unknown race	314 630 (19.89)	362 280 (20.32)
Age		
18–34	139 697 (8.83)	165 211 (9.27)
35–64	604 262 (38.21)	733 250 (41.12)
65 and over	837 584 (52.96)	884 684 (49.61)

Data from OSHPD. Age-standardised hospitalisations rates were obtained by adjusting the crude hospitalisation rate for age (using age group California coefficients) and population (using populations within each age group). Hospital charges were inflation adjusted to year 2011 US dollars and divided by California population over 18 years old. Hospitalisations by demographic characteristics are crude numbers.
OSHPD, Office of Statewide Health Planning and Development.

2011, respectively, a total of 1 242 297 (78.6%) and 1 400 467 (78.5%) hospitalisations were included in our analyses. Given the limited number of variables available in the OSHPD datasets, we were not able to impute missing values, which may introduce bias if data are not missing at random.

Our data were comprised of 3719 zip codes in 2001 and 3574 zip codes in 2011. Of these, we excluded 2087 zip codes in 2001 and 1902 zip codes in 2011 that were outside California (figure 1). Of these, we only mapped 1448 (39%) zip codes from 2001 and 1504 (42%) zip codes in 2011, as zip codes with less than 500 inhabitants were excluded to protect privacy. The number of California zip codes vary given changes to postal routes over the 10-year period.

Patient and public involvement

There was no patient or public involvement in the design or planning of the study.

RESULTS

Geographical distribution of hospitalisations

Figures 2 and 3 illustrate the geographical distribution at the zip code level of median household income quintiles and hospitalisations by 100 000 persons in California, respectively. The highest zip code income quintiles in California clustered around the large metropolitan areas including the San Francisco Bay area, the Sacramento area, Los Angeles and San Diego. The bottom zip code income quintiles were found in more rural areas, as well as in isolated sections of large metropolitan areas. Between 2001 and 2011, median household income of zip codes around metropolitan areas increased, but it remained unchanged in the lower income zip codes. Higher total age-standardised hospitalisation rates (figure 3) and disease-specific hospitalisation rates (see online supplementary file 2) were found in rural and low-income zip codes.

Demographic characteristics and hospitalisation rates

There were 1 581 543 hospitalisations in 2001 and 1 783 145 hospitalisations in 2011 in California (table 1). Overall, there were 6339 and 6315 hospitalisations per 100 000 persons for 2001 and 2011, respectively. Between 2001 and 2011, there was no substantial difference in hospitalisations distribution by sex, whereas hospitalisations in Latinos increased by half and in individuals in the 'other race' category doubled (table 1).

Mean length of stay decreased and hospital charges increased from 2001 to 2011 in all zip code income quintiles. However, this masked heterogeneity by zip code-level income for expenditures. Hospital expenditure per zip code population decreased with increasing income quintiles, with the lowest zip code income quintile having US$2561 per zip code population in 2001 and US$4473 in 2011. The highest income quintile had the lowest hospital charges, with US$1654 in 2001 and US$2990 in 2011 (table 2).

There was a higher percentage of hospitalisations per zip code population over 18 years old in the lower zip code income quintiles in 2011 compared with 2001, from 7.49% of the population to 7.86% in the first (lowest) zip code income quintile and 7.23%–7.31% in the second zip code income quintile, respectively (table 2). In comparison, hospitalisations decreased slightly in the higher income quintiles from 2001 to 2011, from 5.12% to 4.87% in the fifth (highest) income quintile (table 2). There were more age-standardised hospitalisations in lower income zip code quintiles both in 2001 and 2011 compared with higher zip code income quintiles, with 8236 and 8421 hospitalisations per 100 000 persons in the first zip code income quintile and 5093 and 4617 hospitalisations per 100 000 persons in the fifth zip code income

Table 2 Demographic characteristics for all hospitalisations by zip code income, California 2001–2011

	Median household income zip code quintiles N (%)									
	First quintile (lowest)		Second quintile (low)		Third quintile (middle)		Fourth quintile (high)		Fifth quintile (highest)	
	2001	2011	2001	2011	2001	2011	2001	2011	2001	2011
Population over 18	3 027 857	3 199 202	4 403 545	5 630 115	5 449 073	6 577 564	5 880 284	6 422 577	5 839 227	5 812 685
Total hospitalisations	226 717 (7.49)	251 305 (7.86)	318 316 (7.23)	411 388 (7.31)	347 443 (6.38)	428 404 (6.51)	352 610 (6.00)	375 309 (5.84)	298 810 (5.12)	282 847 (4.87)
Age- standardised hospitalisations per 100 000	8236	8421	7401	7558	6503	6698	5863	5577	5093	4617
Mean length of stay (days)	5.2	4.9	5.0	4.7	4.8	4.7	4.7	4.6	4.7	4.6
Hospital charge per population (2011 US dollars)	2561	4473	2508	4277	2150	3866	1899	3513	1654	2990
Gender										
Male	92 890 (41.0)	104 697 (41.7)	130 336 (41.0)	172 908 (42.0)	139 266 (40.1)	179 475 (41.9)	142 052 (40.3)	155 718 (41.5)	119 561 (40.0)	114 920 (40.6)
Female	105 518 (46.5)	117 273 (46.7)	156 036 (49.0)	200 035 (48.6)	171 837 (49.5)	208 337 (48.6)	169 985 (48.2)	180 246 (48.0)	141 294 (47.3)	132 205 (46.7)
Unknown gender	28 309 (12.5)	29 335 (11.7)	31 944 (10.0)	38 445 (9.35)	36 340 (10.5)	40 592 (9.48)	40 573 (11.5)	39 345 (10.5)	37 955 (12.7)	35 722 (11.6)
Race										
White	81 361 (35.9)	75 730 (30.1)	162 089 (50.9)	179 269 (43.6)	201 823 (58.1)	203 698 (47.6)	221 156 (62.7)	215 856 (57.5)	206 304 (69.0)	176 914 (62.6)
Black	33 185 (14.6)	38 507 (15.3)	35 545 (11.2)	40 564 (9.86)	14 980 (4.31)	24 621 (5.75)	12 507 (3.55)	14 705 (3.92)	5237 (1.75)	6267 (2.22)
Latino	51 736 (22.8)	69 820 (27.8)	48 406 (15.2)	91 202 (22.2)	51 000 (14.7)	86 581 (20.2)	29 235 (8.29)	38 493 (10.3)	10 161 (3.40)	14 242 (5.04)
Asian	7820 (3.45)	9996 (3.98)	11 064 (3.48)	16 802 (4.08)	12 360 (3.56)	22 789 (5.32)	16 755 (4.75)	23 903 (6.37)	13 758 (4.60)	16 975 (6.00)
Native American	421 (0.19)	460 (0.18)	356 (0.11)	497 (0.12)	193 (0.06)	329 (0.08)	181 (0.05)	333 (0.09)	95 (0.03)	142 (0.05)
Other race	1888 (0.83)	3340 (1.33)	2986 (0.94)	8179 (1.99)	3648 (1.05)	8801 (2.05)	3735 (1.06)	6680 (1.78)	2512 (0.84)	4772 (1.69)
Unknown race	50 306 (22.2)	53 452 (21.3)	57 870 (18.2)	74 875 (18.2)	63 439 (18.3)	81 585 (19.0)	69 041 (19.6)	75 339 (20.1)	60 743 (20.3)	63 535 (22.5)
Age										
18–34	24 464 (10.8)	27 968 (11.1)	31 173 (9.79)	42 828 (10.4)	31 235 (8.99)	42 064 (9.82)	27 434 (7.78)	29 157 (7.77)	20 749 (6.94)	20 059 (7.09)
35–64	95 510 (42.1)	115 778 (46.1)	127 090 (39.9)	181 770 (44.2)	129 002 (37.1)	177 441 (41.4)	125 434 (35.6)	139 230 (37.1)	107 602 (36.0)	99 431 (35.2)
65 and over	106 743 (47.1	107 559 (42.8)	160 053 (50.3)	186 790 (45.4)	187 206 (53.9)	208 899 (48.8)	199 742 (56.7)	206 922 (55.1)	170 459 (57.1)	163 357 (57.8)

Data from OSHPD and US Census. Age-standardised hospitalisations rates were obtained by adjusting the crude hospitalisation rate for age (using age group California coefficients) and population (using populations within each age group and zip code). Hospital charges inflation adjusted to year 2011 US dollars and divided by California population over 18 years old. Hospitalisations by demographic characteristics are crude numbers.
OSHPD, Office of Statewide Health Planning and Development.

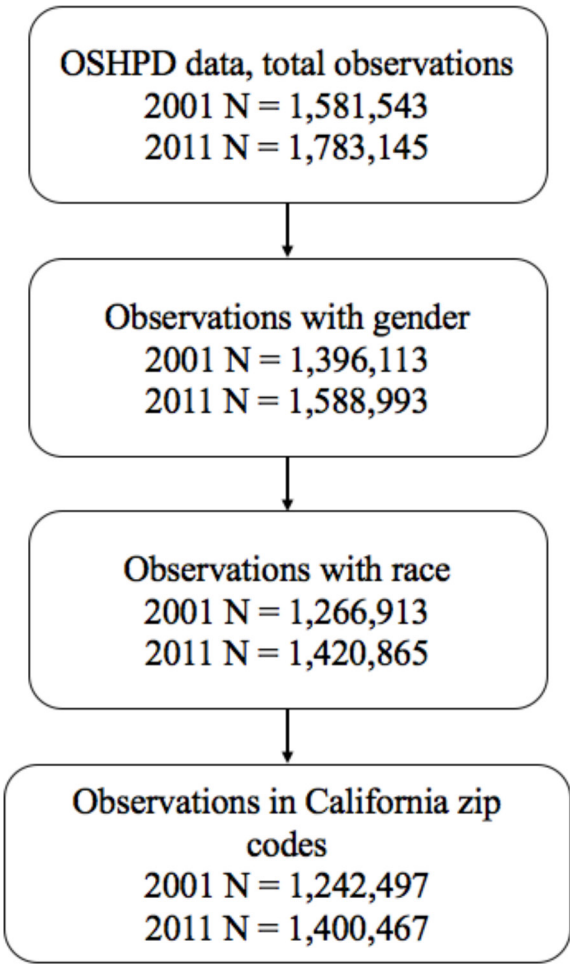

Figure 1 Sample flowchart. Note: data from US Census and OSHPD. OSHPD, Office of Statewide Health Planning and Development.

quintile. Age-standardised hospitalisation rates increased in all income quintiles except the fourth and fifth zip code income quintile (figure 4).

Age-standardised hospitalisation rates were highest in blacks for both years in all zip code income categories compared with all other racial groups (figure 5). Age-standardised hospitalisation rates decreased from 2001 to 2011 in all racial groups except for whites in the first and fifth zip code income groups, blacks in the fifth zip code income group and Latinos in the fourth and fifth zip code income groups (figure 5 and online supplementary file 1).

DISCUSSION

This study, based on over 3 million hospitalisations in 2001 and 2011 in California, highlights important disparities in healthcare use and cost according to zip code-level median household income and hospitalisation rates, which may reflect disparities in disease burden. While total age-standardised hospitalisation rates decreased from 2001 to 2011, in concordance with national hospitalisation trends, they impacted zip code income groups and racial groups differently.[16] We found hospitalisation rates

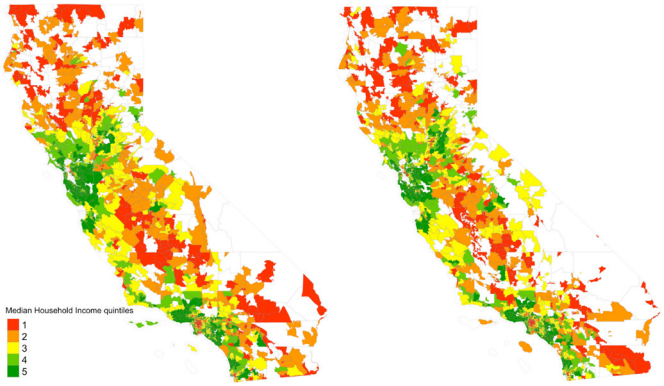

Figure 2 Zip code median household income distribution in California, 2001 and 2011. Note: For 2001, the income quintiles were the following: first quintile (lowest), US$0–US$31 002; second quintile, US$31 003–US$38 697; third quintile, US$38 698–US$47 810; fourth quintile, US$47 811–US$61 766 and fifth quintile (highest), above US$61 767. For 2011, the income quintiles were: first quintile (lowest), US$0–US$39 442; second quintile, US$39 443–US$51 644; third quintile, US$51 645–US$64 250; fourth quintile, US$64 251–US$82 633 and fifth quintile (highest), above US$82 634. Data from US Census and OSHPD, GIS mapping files obtained as TIGER/Line Shapefiles from US Census (https://www.census.gov/geographies.html) and as shapefiles from Data.gov (https://www.data.gov/), both open data sites. Hospitalisations rates were obtained each zip code income quintile by adjusting the crude hospitalisation rate for age (using age group California coefficients) and population (using populations within each age group and zip code). Zip codes in white represent areas where populations are either under 500 inhabitants or areas where we do not have data. OSHPD, Office of Statewide Health Planning and Development.

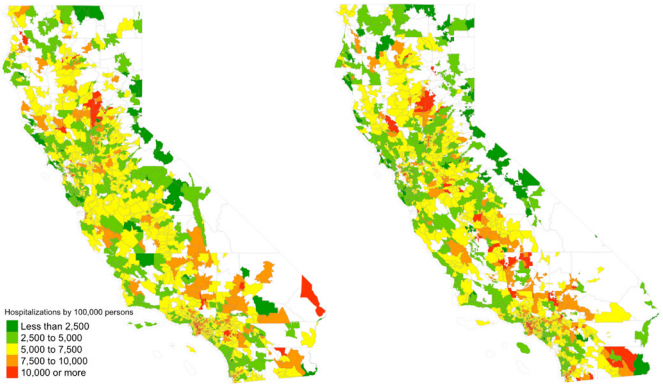

Figure 3 Distribution of all hospitalisations in California, 2001 and 2011. Note: data from US Census and OSHPD, geographic information system (GIS) mapping files obtained as Topologically Integrated Geographic Encoding and Referencing (TIGER/Line) shapefiles from US Census (https://www.census.gov/geographies.html) and as shapefiles from Data.gov (https://www.data.gov/), both open data sites. Hospitalisations rates were obtained each zip code income quintile by adjusting the crude hospitalisation rate for age (using age group California coefficients) and population (using populations within each age group and zip code). Zip codes in white represent areas where populations are either under 500 inhabitants or areas where we do not have data. OSHPD, Office of Statewide Health Planning and Development.

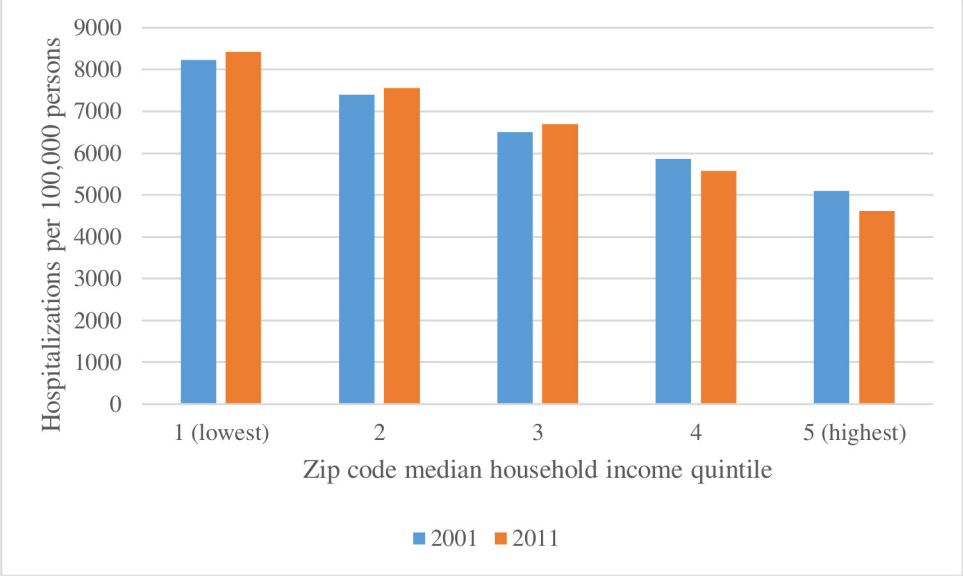

Figure 4 Total hospitalisation rates by income, California. Note: data from US Census and OSHPD, GIS mapping files obtained as TIGER/Line Shapefiles from US Census (https://www.census.gov/geographies.html) and as shapefiles from Data.gov (https://www.data.gov/), both open data sites. Hospitalisations rates were obtained for each zip code income quintile by adjusting the crude hospitalisation rate for age (using age group California coefficients) and population (using populations within each age group and zip code). OSHPD, Office of Statewide Health Planning and Development.

decreased with increasing zip code income. We also noted that an overall increase in hospital expenditures masked subgroup differences, in that hospital charges were both the highest and increased the most in the lowest income quintile, highlighting a clear disparity between zip code income quintiles.

These findings are consistent with prior work, which found that young and middle-aged residents of poor US urban neighbourhoods were at higher risk of early mortality due to chronic diseases.[17 18] A recent study linking Veterans Health Administration data with US Census data showed higher hospitalisation rates in white Veterans and in Veterans livings in low-income census tracts.[19] Another study mapped hospital days and zip code income for California urban areas, emphasising the importance of disaggregating county-level data by showing a strong association between low zip code income and higher percentage of disability and greater use of hospitals.[20] Other studies have associated neighbourhood poverty and disadvantage with adverse cardiovascular outcomes, particularly in black women.[21 22] This income-health gradient is also described in a more recent large national study which reported increasing mortality associated with less affluent local geographies.[4] Higher rates of hospitalisations in children for preventable conditions were reported in zip codes with higher income inequality.[3] These patterns may be explained by differences in incidence of chronic disease as well as treatment of those conditions. Our study adds to this literature by providing more granular data at zip code and at diagnostic disease code levels.

In our study, blacks made up 7% of total hospitalisations, similar to the per cent of black adults in the California population.[12] Yet, our calculation of age-standardised hospitalisation rates per racial group exposed a disparity: blacks consistently had higher age-standardised hospitalisation rates in all zip code income groups and disease categories. Assessing overall hospitalisation rates only would have masked this disparity. California trends are similar to those at the national level, where blacks have more hospital stays than other racial groups.[16] It has been postulated that racial disparities in the USA are driven by social disadvantages (ie, unemployment, incarceration, lower socioeconomic status), institutional racism, differential usage of healthcare services, distrust of the medical system due to past medical abuses and cumulative adverse life experiences.[23] While neighbourhood socioeconomic status has been cited as a possible driver for increased disease burden in blacks, in our study, we see greater rates of hospitalisations in blacks regardless of zip code income as has been noted in other studies.[24]

Some successful neighbourhood-level interventions have focused on alleviating the experience of poverty.[25] These include community development, zoning and demolition of distressed housing.[26] Expansion of rental vouchers programme has been supported by robust evidence of health improvements as a result of those programme.[26] Successful neighbourhood-level interventions depend on multiple factors such as involving communities and designing multidisciplinary solutions. However, the first step to any such intervention is the identification of vulnerable populations, their geographic distribution and the presence of inequities, much like our study has done.[25] The areas of high hospitalisation rates we identified in 2001 and 2011 still had a shortage of primary and specialty care in 2015.[27] Our study brings

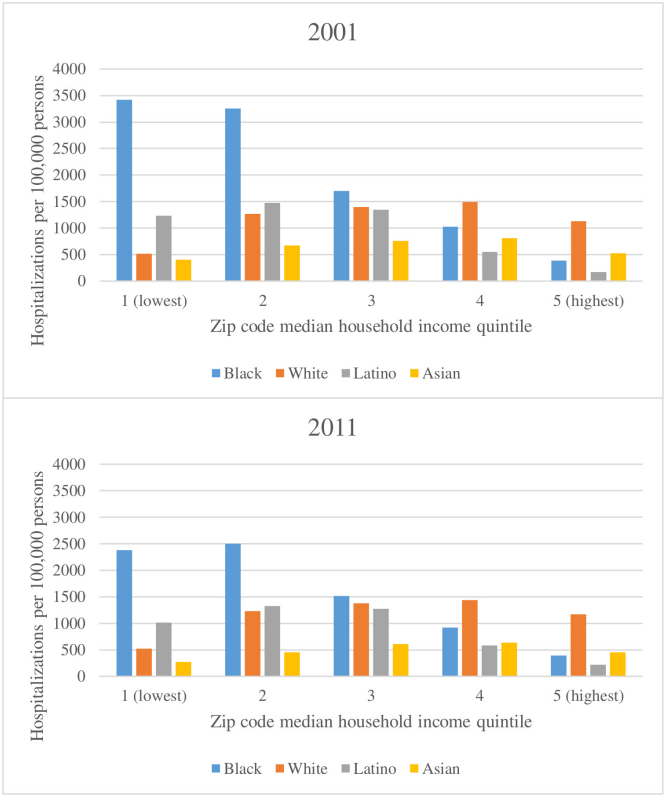

Figure 5 Total hospitalisations by race and income, California. Note: data from US Census and OSHPD, GIS mapping files obtained as TIGER/Line Shapefiles from US Census (https://www.census.gov/geographies.html) and as shapefiles from Data.gov (https://www.data.gov/), both open data sites. Hospitalisations rates were obtained for each zip code income quintile and racial group by adjusting the crude hospitalisation rate for age (using age group California coefficients) and population (using populations within each age group, racial group and zip code). OSHPD, Office of Statewide Health Planning and Development.

attention to health inequities in vulnerable populations, such as blacks, but also in vulnerable areas, notably low-income zip codes with populations with high rates of hospitalisations. Major efforts, such as the Neighborhood Atlas which maps the Area Deprivation Index, have facilitated studies on census block group-level socioeconomic deprivation and health outcomes.[28 29] Future interventions could involve using such data in identifying at-risk areas that would benefit from increasing access to primary care, as studies have shown the substantial impact primary care has on mortality.[30] In addition, policies could be tested to increase that access by expanding health insurance coverage.

There are several limitations to our study. One is the use of deidentified hospital discharge data instead of individual-level data, preventing calculations of incidence, prevalence or disability-adjusted life-years. Relatedly, we were not able to assess for readmissions. Also, OSHPD data do not include information on individual income or employment, and OSHPD sociodemographic data were missing for 20% of observations. Given the few demographic variables available in the OSHPD dataset, we were unable to impute missing data. Of note, it is likely that in some cases, race data were reported by hospital staff rather than patients themselves, which is likely to result in misclassification. Additionally, as we did not have access to individual addresses, we relied on zip codes; these present some limitations, since they are designed based on postal routes and may not correspond closely with neighbourhoods, and they are also larger and more heterogeneous than smaller area measures such as census tracts.[31] Zip codes may represent aggregates of varying access to healthcare. Estimates may also have been biased because we had to exclude zip codes outside of California and zip codes with small numbers of residents (approximately 40%). Because we used quintiles of income aggregated at the state level, this analysis therefore does not fully capture the effects of individual zip codes; future studies could address this, while also including other zip code-level predictors like unemployment or housing that might be used to test the effects of specific local interventions. Also, certain diagnostic codes might be over-represented in hospital discharge data. As admission to a hospital usually represents a severe clinical presentation, diagnoses codes for less severe presentations may be under-represented. Additionally, although this study uses data from several years ago, it nevertheless provides insight into differences over a decade; future work should replicate this analysis with more recent data to capture the latest economic and political trends, especially after the passage and implementation of the Affordable Care Act. Finally, the standardisation method that we used would not be appropriate if we were adjusting for additional variables beyond age. Future studies could instead use different methods for multivariable adjustment.

Our study uncovers clear patterns in disparities in hospitalisation rates by zip code income at two time points 10 years apart. Future studies could extend this work to examine small area effects, which could aid the development of interventions to address these costly disparities at the zip code level. The many factors that lead to changes in hospitalisation rates—prevalence of a certain health condition, events that trigger acute presentations, access to primary care for preventable conditions—may be affected by both generalised trends, like increase in insurance coverage, and local settings, like air quality due to fires. Hospitalisation data linked to socioeconomic data offer insight on temporal and geographical trends and could also be used to estimate the effects of local, state or federal policies. Hospitalisation data can be used to measure disease burden in communities but also to assess the effect of interventions such as increasing access to primary care.

CONCLUSION

By combining spatially defined state hospitalisation data with spatially defined area-level income data, we uncovered important differences in healthcare use in

California. The lower the zip code-level income, the greater the rates of hospitalisation. We also found that blacks have disproportionally higher rates of hospitalisation than any other racial group, regardless of zip code income. Such methods can be used by public health departments to identify 'hot spots' areas that would benefit from community-based interventions.[32] By quantifying and mapping health outcomes with finer units of social demographic information, studies such as this one may help to provide guidance for allocations of limited public health resources.

Acknowledgements The authors would like to thank Dr Jayant Rajan for his involvement in the beginning steps of this study and the UCSF Primary Care Research Fellowship for their valuable review of the manuscript.Publication made possible in part by support from the UCSF Open Access Publishing Fund.

Contributors ER conceived of the study, analysed and interpreted the data, and drafted the manuscript. RG and RH contributed to study design and data interpretation, and reviewed the manuscript critically for intellectual content. All authors read and approved of the final manuscript.

Funding ER was supported by the following funding: T32HP19025.

Map disclaimer The depiction of boundaries on the map(s) in this article do not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. The map(s) are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval University of California, San Francisco, Human Research Protection Program and IRB 14-14049.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data from the US Census Bureau are available in a public, open-access repository. Data from OSHPD may be obtained from a third party and are not publicly available.

ORCID iD

Eva Raphael http://orcid.org/0000-0003-2854-0247

REFERENCES

1 Pickett KE, Wilkinson RG. Income inequality and health: a causal review. *Soc Sci Med* 2015;128:316–26.
2 Ross CE, Mirowsky J. Neighborhood disadvantage, disorder, and health. *J Health Soc Behav* 2001;42:258–76.
3 Bettenhausen JL, Colvin JD, Berry JG, *et al*. Association of income inequality with pediatric hospitalizations for ambulatory Care-Sensitive conditions. *JAMA Pediatr* 2017;171:e170322.
4 Chetty R, Stepner M, Abraham S, *et al*. The association between income and life expectancy in the United States, 2001-2014. *JAMA* 2016;315:1750–66.
5 Hastings KG, Boothroyd DB, Kapphahn K, *et al*. Socioeconomic differences in the epidemiologic transition from heart disease to cancer as the leading cause of death in the United States, 2003 to 2015: an observational study. *Ann Intern Med* 2018;169:836.
6 Thacker SB, Stroup DF, Carande-Kulis V, *et al*. Measuring the public's health. *Public Health Rep* 2006;121:14–22.
7 Diez Roux AV, Neighborhoods DRAV. Neighborhoods and health: what do we know? what should we do? *Am J Public Health* 2016;106:430–1.
8 Ludwig J, Sanbonmatsu L, Gennetian L, *et al*. Neighborhoods, obesity, and diabetes--a randomized social experiment. *N Engl J Med* 2011;365:1509–19.
9 Zip code Tabulation areas. *U.S. Census Bureau*.
10 Browning CR, Cagney KA. Moving beyond poverty: neighborhood structure, social processes, and health. *J Health Soc Behav* 2003;44:552–71.
11 Geronimus AT, Bound J. Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples. *Am J Epidemiol* 1998;148:475–86.
12 American FactFinder. U.S. census bureau
13 Murray CJL, Ezzati M, Flaxman AD, *et al*. GBD 2010: design, definitions, and metrics. *The Lancet* 2012;380:2063–6.
14 International Classification of Diseases,Ninth Revision (ICD-9). Available: https://www.cdc.gov/nchs/icd/icd9.htm
15 Murray CJL, Vos T, Lozano R, *et al*. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The Lancet* 2012;380:2197–223.
16 Table 81 (page 1 of 4). persons with Hospital stays in the past year, by selected characteristics: United States, selected years 1997–2015*Center for Disease Control and Prevention*2016
17 Geronimus AT, Bound J, Waidmann TA. Poverty, time, and place: variation in excess mortality across selected us populations, 1980-1990. *J Epidemiol Community Health* 1999;53:325–34.
18 Geronimus AT. To mitigate, resist, or undo: addressing structural influences on the health of urban populations. *Am J Public Health* 2000;90:867–72.
19 Hatef E, Searle KM, Predmore Z, *et al*. The impact of social determinants of health on hospitalization in the Veterans health administration. *Am J Prev Med* 2019;56:811–8.
20 Cooper RA, Cooper MA, McGinley EL, *et al*. Poverty, wealth, and health care utilization: a geographic assessment. *J Urban Health* 2012;89:828–47.
21 Barber S, Hickson DA, Wang X, *et al*. Neighborhood disadvantage, poor social conditions, and cardiovascular disease incidence among African American adults in the Jackson heart study. *Am J Public Health* 2016;106:2219–26.
22 Spatz ES, Beckman AL, Wang Y, *et al*. Geographic variation in trends and disparities in acute myocardial infarction hospitalization and mortality by income levels, 1999-2013. *JAMA Cardiol* 2016;1:255–65.
23 Brown TH, Hargrove TW, Griffith DM. Racial/Ethnic disparities in men's health: examining psychosocial mechanisms. *Fam Community Health* 2015;38:307–18.
24 Wilson KB, Thorpe RJ, LaVeist TA. Dollar for dollar: racial and ethnic inequalities in health and health-related outcomes among persons with very high income. *Prev Med* 2017;96:149–53.
25 Sandel M, Faugno E, Mingo A, *et al*. Neighborhood-Level interventions to improve childhood opportunity and lift children out of poverty. *Acad Pediatr* 2016;16:S128–35.
26 Lindberg RA, Shenassa ED, Acevedo-Garcia D, *et al*. Housing interventions at the neighborhood level and health: a review of the evidence. *J Public Health Manag Pract* 2010;16:S44–52.
27 Gaines R. California maps: how many primary care and specialist physicians are in your County? Available: https://www.chcf.org/publication/california-maps-how-many-primary-care-and-specialist-physicians-are-in-your-county/ [Accessed 8 May 2019].
28 Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. *N Engl J Med* 2018;378:2456–8.
29 Hu J, Kind AJH, Nerenz D. Area deprivation index predicts readmission risk at an urban teaching hospital. *Am J Med Qual* 2018;33:493–501.
30 Basu S, Berkowitz SA, Phillips RL, *et al*. Association of primary care physician supply with population mortality in the United States, 2005-2015. *JAMA Intern Med* 2019:2005–15.
31 Krieger N, Chen JT, Waterman PD, *et al*. 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;156:471–82.
32 Basu S, Siddiqi A. Geographic disparities in US mortality: "hot-spotting" large databases. *Epidemiology* 2014;25:468–70.
33 Vos T, Flaxman AD, Naghavi M, *et al*. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of disease study 2010. *The Lancet* 2012;380:2163–96.