

# Sociodemographic Characteristics of Members of a Large, Integrated Health Care System: Comparison with US Census Bureau Data

Corinna Koebnick, PhD, MS; Annette M Langer-Gould, MD, PhD, MS; Michael K Gould, MD, MS; Chun R Chao, PhD, MS; Rajan L Iyer, MPH; Ning Smith, PhD; Wansu Chen, MS; Steven J Jacobsen, MD, PhD

Perm J 2012 Summer;16(3):37-41

## Abstract

**Background:** Data from the memberships of large, integrated health care systems can be valuable for clinical, epidemiologic, and health services research, but a potential selection bias may threaten the inference to the population of interest.

**Methods:** We reviewed administrative records of members of Kaiser Permanente Southern California (KPSC) in 2000 and 2010, and we compared their sociodemographic characteristics with those of the underlying population in the coverage area on the basis of US Census Bureau data.

**Results:** We identified 3,328,579 KPSC members in 2000 and 3,357,959 KPSC members in 2010, representing approximately 16% of the population in the coverage area. The distribution of sex and age of KPSC members appeared to be similar to the census reference population in 2000 and 2010 except with a slightly higher proportion of 40 to 64 year olds. The proportion of Hispanics/Latinos was comparable between KPSC and the census reference population (37.5% vs 38.2%, respectively, in 2000 and 45.2% vs 43.3% in 2010). However, KPSC members included more blacks (14.9% vs 7.0% in 2000 and 10.8% vs 6.5% in 2010). Neighborhood educational levels and neighborhood household incomes were generally similar between KPSC members and the census reference population, but with a marginal underrepresentation of individuals with extremely low income and high education.

**Conclusions:** The membership of KPSC reflects the socioeconomic diversity of the Southern California census population, suggesting that findings from this setting may provide valid inference for clinical, epidemiologic, and health services research.

## Introduction

Data from the memberships of integrated health care organizations offer several advantages for health researchers, including large samples and availability of electronic health records (EHR) that provide diagnostic codes, pharmacy records, vaccination records, and membership characteristics.<sup>1-9</sup> In some cases, these data may be augmented by comprehensive inpatient and outpatient progress notes, radiologic images, and reports.<sup>10-12</sup> These features facilitate researchers in performing studies of health disparities, long-term patient outcomes, and comparative effectiveness in a timely and cost-efficient manner.

Most US health plan members, however, receive health insurance through the employer of at least one family member. This covered individual may be healthier and may have other advantages, such as more years of education than the general population, thus raising concern that findings from studies performed in integrated health care settings may not be generalizable to younger or disadvantaged portions of the US population. Furthermore, because low socioeconomic status may be associated with poor health outcomes,<sup>13-15</sup> a healthy worker effect may bias findings from studies in these settings by underestimating the magnitude of the effect of important

predictors for poor health outcomes that are also associated with low socioeconomic status or by failing to identify such predictors in entirety.

The purpose of this study was to compare the sociodemographic characteristics of the members of a large integrated health care organization, Kaiser Permanente Southern California (KPSC), with the census population of the Southern California coverage area.

## Methods

### Setting and Design

An integrated health care system, KPSC provides comprehensive health care for more than 3.4 million of the 23 million residents of Southern California. Members receive medical care in 14 hospitals and more than 197 medical offices in 10 counties of Southern California: Imperial, Kern, Los Angeles, Orange, Riverside, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, and Ventura. Medical information is captured in complete EHR that include all inpatient and outpatient progress notes; pharmacy records; radiology reports and images; and membership characteristics, including race/ethnicity and language preference, both written and spoken. Members can obtain KPSC insurance coverage through employer-based plans, individual plans, and Medicare or state-subsidized health care for the indigent.

For this study, we identified all individuals who were members of KPSC at any time in the years 2000 and 2010. Sociodemographic information was collected at the time of Health Plan enrollment, and

Corinna Koebnick, PhD, MS, is a Research Scientist for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: corinna.koebnick@kp.org. Annette M Langer-Gould, MD, PhD, MS, is a Research Scientist for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: annette.m.langer-gould@kp.org. Michael K Gould, MD, MS, is a Research Scientist for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: michael.k.gould@kp.org. Chun R Chao, PhD, MS, is a Research Scientist for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: chun.r.chao@kp.org. Rajan L Iyer, MPH, is a Research Associate for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: rajan.l.iyer@kp.org. Ning Smith, PhD, is a Biostatistician for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: ning.x.smith@kp.org. Wansu Chen, MS, is the Group Leader of Biostatistics, Programming & Database Development for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: wansu.chen@kp.org. Steven J Jacobsen, MD, PhD, is the Director of Research for Kaiser Permanente Research and Evaluation in Pasadena, CA. E-mail: steven.j.jacobsen@kp.org.

missing or incorrect information may have been updated during inpatient and outpatient medical visits. The institutional review board of KPSC reviewed and approved the study protocol.

### Race and Ethnicity

We categorized race as white, black, American Indian/Alaskan Native, Asian/Pacific Islander, multiple races, and other

rac. Ethnicity was classified as Hispanic or non-Hispanic. Race and ethnicity information for KPSC members was extracted from administrative records, a method previously validated against birth certificate information.<sup>16</sup>

### Socioeconomic Status

As indicators of socioeconomic status, we used three different measures: neigh-

borhood education, neighborhood income, and participation in Medi-Cal (Medicaid) or other state-subsidized health care coverage programs. Neighborhood education and neighborhood income were estimated on the basis of the linkage of Health Plan members' addresses via geocoding (Geospatial Entity Object Coding) with US Census block data.<sup>17</sup>

### Reference Populations

The reference populations included all residents of the 10 counties of Southern California who were included in the 2000 and 2010 censuses. Information about the Southern California census populations was retrieved from the US Census Bureau files using the full data set through the Web-based query portal ([www.census.gov](http://www.census.gov)). Census information on sex, race, ethnicity, education, household income, households with income below the poverty level, and public assistance income were extracted from demographic profile summary files. To match Health Plan administrative records, we collapsed the available race categories from the census questionnaire to the following categories: white, black, American Indian/Alaskan Native, Asian/Pacific Islander, multiple races, and other race.

### Statistical Analysis

We report descriptive statistics for variables of interest in the KPSC population and the Southern California reference population. We report similar descriptive statistics stratified by age group only for the year 2000, because these data were not available for the census population in 2010. We did not perform formal statistical tests to identify differences between the two populations. Because of the large population size, even small—but not necessarily relevant—differences between populations would result in a significant test result.

### Results

Members of KPSC in 2000 and 2010 represented approximately 16.1% of the census reference population in the KPSC coverage area (Table 1). The overall distribution of gender and age of KPSC members appeared to be similar to the census reference population in 2000 and 2010, with the exception that the 40- to 64-year-old age group was marginally

**Table 1. Demographic characteristics<sup>a</sup>**

Demographic characteristic	2000		2010	
	Census	KPSC	Census	KPSC
Total population (N)	20,637,512	3,328,579	22,680,010	3,657,959
Sex				
Male	49.8	48.9	49.7	48.4
Female	50.2	51.1	50.3	51.6
Age group, years				
0 to 9	16.0	15.3	13.6	12.9
10 to 14	7.8	8.0	7.2	7.5
15 to 19	7.3	7.5	7.9	8.1
20 to 39	31.1	29.1	29.1	26.2
40 to 64	27.6	30.8	31.3	34.1
≥65	10.9	9.2	10.9	11.2
Race <sup>b</sup>				
Non-Hispanic white	42.3	46.3	36.4	34.0
Hispanic white	15.3	22.7	20.5	36.9
Black	7.0	14.9	6.5	10.8
American Indian/Alaska Native	0.9	0.2	0.9	0.3
Asian/Pacific Islander	9.9	8.6	11.8	10.1
Other races	19.9	7.1	19.3	7.5
Multiple races	4.7	0.3	4.6	0.4
Hispanic or Latino <sup>b</sup>	38.2	37.5	43.3	45.2
Neighborhood education <sup>c</sup>				
Less than high school	25.8	25.2	21.3	21.4
High school graduate	20.0	20.9	20.9	23.0
Some college or associate degree	29.5	30.6	29.1	29.9
Bachelor's degree	16.0	15.2	18.5	17.0
Graduate or professional degree	8.7	7.9	10.1	8.7
Neighborhood household income <sup>c</sup>				
<\$10,000	8.8	8.1	6.1	5.9
\$10,000 to \$14,999	5.9	5.6	5.3	3.4
\$15,000 to \$24,999	12.1	11.8	10.3	8.4
\$25,000 to \$34,999	11.9	11.9	9.5	9.1
\$35,000 to \$49,999	15.4	15.9	13.4	13.2
\$50,000 to \$74,999	19.0	20.3	17.7	19.0
\$75,000 to \$99,999	11.1	11.8	12.4	14.4
\$100,000 to \$149,999	9.6	9.6	14.0	15.4
≥\$150,000	6.1	5.0	11.2	11.2

<sup>a</sup>Data are percentages of subjects unless otherwise indicated. Some data do not total to 100% because of rounding.

<sup>b</sup>For KPSC, information about race and ethnicity was based on administrative records among those with known race ethnicity (members with unknown race: 43.7% in 2000 and 24.9% in 2010).

<sup>c</sup>Neighborhood income and education are not reported income and education but are estimated on the basis of members' addresses using neighborhood income and education from US Census tract information.

KPSC = Kaiser Permanente Southern California.

overrepresented among KPSC members (30.8% vs 27.6% in 2000 and 34.1% vs 31.3% in 2010; Table 1).

The proportion of Hispanics/Latinos was comparable between KPSC and the census reference population in 2000 (37.5% vs 38.2%) and 2010 (45.2% vs 43.3%). However, KPSC members included more blacks in both 2000 and 2010 (14.9% vs 7.0% in 2000 and 10.8% vs 6.5% in 2010). Non-Hispanic whites were slightly over-represented among KPSC members in 2000, but in 2010 this group was somewhat underrepresented (46.3% vs 42.3% in 2000 and 34.0% vs 36.4% in 2010).

Whereas the KPSC membership and the census reference population had similar proportions of Hispanics in both 2000 and 2010, the census population included fewer self-reported Hispanic whites and more individuals who classified themselves as “other race” in these years.

Neighborhood educational level and neighborhood household income were generally similar between KPSC members and the census reference population (Table 1). However, slightly fewer KPSC members in 2010 resided in neighborhoods with household incomes below \$25,000 (17.7% vs 21.6%, respectively), or in neighborhoods with a higher percentage of college graduates (25.7% vs 28.6%).

Approximately 1.7% of KPSC members received services paid by Medi-Cal, California’s state-subsidized health care program (Figure 1). The proportion of

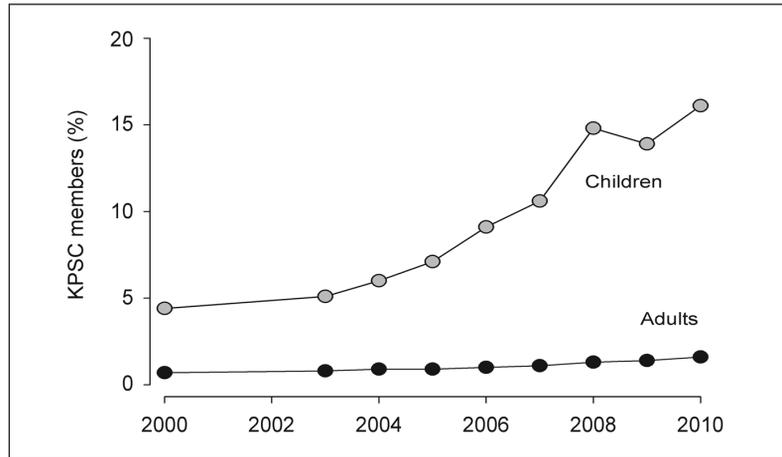


Figure 1. Proportion of Kaiser Permanente Southern California (KPSC) members who receive health care coverage by Medi-Cal (Medicaid) and other state-subsidized programs, by age group.

Adults were defined as 18 years of age or older, and children were defined as younger than 18 years of age. (Cut off age determined by Medicaid/Medi-Cal eligibility.)

KPSC members who received health care coverage by Medi-Cal and other state-subsidized programs increased from 0.7% to 1.6% among adults and from 4.4% to 16.1% among youths between 2000 and 2010. In the coverage area of Southern California, an estimated 11.6% had an income below the poverty level, and 5.1% received public assistance in 2000, whereas in 2010 an estimated 16.2% had an income below the poverty level and 4.0% received public assistance.

Members of KPSC between 0 and 19 years of age were generally similar

in demographic characteristics to the census reference population in 2000<sup>a</sup> (Table 2). Members of KPSC represented 15.2% of 0 to 9 year olds, 16.5% of 10 to 14 year olds, and 16.5% of 15 to 19 year olds in the Southern California coverage area. Differences in racial/ethnic groups between KPSC youth and Southern California census youth were similar to the differences observed in the overall populations of all ages, although the higher proportion of blacks seen in KPSC was even more pronounced among 10 to 19 year olds.

Demographic characteristic	0 to 9 years		10 to 14 years		15 to 19 years	
	KPSC	Census	KPSC	Census	KPSC	Census
Total population (N)	510,477	3,310,416	267,431	1,596,627	248,709	1,514,947
Sex (%)						
Male	51.0	51.2	50.8	51.2	50.7	51.7
Female	48.0	48.8	49.2	48.8	49.3	48.3
Race (%) <sup>a</sup>						
Non-Hispanic white	28.0	28.6	27.5	32.8	36.4	33.2
Hispanic white	42.0	20.7	37.0	18.1	34.3	17.0
Black	13.0	7.5	19.0	8.4	18.9	7.8
American Indian/Alaska Native	0.2	1.0	0.2	1.1	0.3	1.1
Asian/Pacific Islander	7.4	7.4	6.9	8.6	7.0	9.9
Other races	9.0	27.5	8.9	24.9	8.2	25.5
Multiple races	0.4	7.3	0.5	6.3	0.6	5.7
Hispanic or Latino (%)	50.4	53.1	47.2	47.2	43.2	46.2

<sup>a</sup> Some data do not total to 100% because of rounding. KPSC = Kaiser Permanente Southern California.

... the KPSC population may be particularly useful for examining the comparative effectiveness of interventions across sociodemographic subgroups.

Adult KPSC members were generally similar to the census reference population in 2000 (Table 3). Members of KPSC represented 15.1% of 20 to 39 year olds, 19.9% of 40 to 64 year olds, and 14.6% of people aged 65 years and older in the Southern California coverage area. Differences in racial/ethnic groups between KPSC adults and Southern California census adults were similar to the differences observed in the overall populations of all ages, although in both KPSC and census reference populations the proportion of Hispanics was significantly lower in adults 40 years and older.

### Discussion

The main finding of this study is that the KPSC population appeared to be similar to the Southern California census reference population in 2000 and 2010. All ages and all racial/ethnic and socioeconomic groups were represented in the KPSC population. Adults aged 40 to 64 years, who likely represent a stable working population, were only marginally over-represented among KPSC members, and the extremely poor and highly educated were only marginally underrepresented among KPSC members in 2010. In general, there were no grossly apparent differences in education or income level between KPSC and the reference population, as would be expected with a healthy insured

effect or healthy worker bias. The similar proportions of low-income individuals in KPSC and the reference population likely reflect the large number of Medi-Cal recipients who are KPSC members. Despite small differences in the proportion of demographic groups, we demonstrated large numbers of KPSC members in all subgroups across the spectrum of age, race and ethnicity, and socioeconomic groups, including a large number of individuals under the poverty threshold and enrolled in subsidized programs to cover health insurance. Our findings suggest that results from studies conducted in the KPSC population may be generalizable to the Southern California population.

The healthy worker bias is an example of a selection bias that can lead to an underestimation of morbidity because of a better health status of the workforce compared with the general population (which also includes people who are too sick to work). Comparably, an insured population may be healthier than the general population because health insurance is often employer sponsored. On the other hand, about 83% of individuals in California had health insurance coverage in 2009.<sup>18</sup> Managed care organizations provide care for a wide range of individuals receiving care through different channels, including employer-based care, family members, and programs subsidized by the state. This diversity makes healthy worker bias and gross differences in socioeconomic characteristics between

the insured and the underlying population less likely to occur.

Although we did not find strong evidence for a healthy worker bias, we cannot exclude the possibility of a mixture of healthy insured effect through attractive KP benefit plans masked by an overrepresentation of members with chronic illnesses because competitor plans are more expensive or do not cover expensive drug costs. If a strong healthy worker bias were present, one would expect an overrepresentation of the stable working population manifested by more men aged 40 to 65 years, and with a higher socioeconomic status compared with the geographic reference population.

Beyond healthy worker bias, health insurance benefit structures also influence the health of its members by discouraging chronically ill members through caps, high copays, and/or deductibles, and by attracting the healthiest of the healthy by offering very low premiums. However, it is possible that competitor plans, by offering high copays for medications and restricting access to specialists, for instance, are more expensive than KPSC and less convenient for those with chronic illnesses. It is not possible to determine how such factors influence the health of the KPSC membership by examining demographic characteristics alone.

On a national level, our findings indicate that the KPSC population may be particularly useful for examining the comparative effectiveness of interven-

**Table 3. Demographic characteristics of adults in 2000, by age group**

Demographic characteristic	20 to 39 years		40 to 64 years		≥65 years	
	KPSC	Census	KPSC	Census	KPSC	Census
Total population (N)	969,395	6,403,335	1,024,723	5,708,965	307,844	2,103,222
Sex (%)						
Male	48.6	51.5	48.1	49.0	46.0	42.0
Female	51.4	48.5	51.9	51.0	54.0	58.0
Race (%) <sup>a</sup>						
Non-Hispanic white	35.7	36.2	48.4	52.8	67.9	67.9
Hispanic white	32.4	16.6	20.6	11.6	12.0	9.3
Black or African American	14.6	6.9	14.7	7.0	11.2	5.5
American Indian/Alaska Native	0.3	1.0	0.2	0.8	0.1	0.5
Asian/Pacific Islander	8.8	10.5	9.7	11.4	6.1	9.1
Other races	7.8	24.1	6.4	13.1	2.6	5.6
Multiple races	0.5	4.7	0.1	3.3	0.1	2.2
Hispanic or Latino (%)	48.0	44.0	29.8	26.6	15.0	15.8

<sup>a</sup> Some data do not total to 100% because of rounding.  
KPSC = Kaiser Permanente Southern California.

tions across sociodemographic subgroups. The diversity and large number of KPSC members make it possible to conduct subgroup analyses aimed at identifying sources of heterogeneity on the basis of demographic factors and estimating risks within such subgroups. In this way, studies conducted in KPSC could help to accomplish this important objective of comparative effectiveness research.<sup>19</sup> Risk estimates generated from such subgroups and general trends are likely to be generalizable in most instances. However, findings from such studies, particularly absolute rates, may not always be generalizable on a national level. On the other hand, the spectrum of illness and conditions seen in this setting are more likely to mirror the general population than studies conducted in tertiary care centers or referral clinics.

Health disparities have previously been attributed to the lack of health insurance.<sup>20</sup> The ethnic and racial diversity of the KPSC population and the large size of these racial and ethnic groups make KPSC an ideal setting to investigate health disparities that persist despite equal access to care.

Limitations of these data include the well-known limitations of the US Census, including undercounting certain minority groups and misclassification of Hispanic whites as “other.” Another issue is missing race and ethnicity information among KPSC members, particularly in 2000. We cannot exclude that differences in the proportion of missing values may partially explain the observed differences between KPSC members in 2000 and 2010 or differences between KPSC members and the census population. This may be especially true for the higher proportion of blacks among KP members. Previous research investigating the quality of race and ethnicity information in KPSC children has shown that missing race is mostly at random with the exception of black children, who have a slightly higher chance of having race information in their EHRs.<sup>16</sup>

Another potential limitation is the reliance on geocoding to obtain a KPSC member's neighborhood education and income instead of self-reported education and income. Neighborhood education and income may or may not exactly reflect an individual's education or income living in that neighborhood. However, it will

accurately reflect the distribution of the population when used for studies that include very large populations, as seen here. In addition, we were unable to compare education, income, and demographics by age group strata with the 2010 US Census because these data are not available. Finally, because our goal was to evaluate overall comparability, we did not perform formal statistical tests to identify differences between the two populations. Given the very large samples, we would expect that differences between groups would be highly significant even when trivial in magnitude or importance.

Strengths of the KPSC population include its similarity to the geographic reference population from which it is drawn, resulting in relatively large Hispanic, black, and Asian populations among children and adults.

In conclusion, the diversity of the KPSC membership along with the comprehensive medical records make this an ideal population to address clinical, epidemiologic, and health services-related questions where race or ethnicity, age, and all but the extreme ends of the income spectrum play key roles. ♦

<sup>a</sup> Cut off for Eligibility for Medicaid/Medi-Cal is age 18 years (as seen in Figure 1); census data, however, came from aggregated tables using age 19 years as the cut off. To have comparable groups, age 19 was used for our characteristics data.

#### Disclosure Statement

*The author(s) have no conflicts of interest to disclose.*

#### Acknowledgments

*This research was supported by Kaiser Permanente Direct Community Benefit Funds. Kathleen Loudon, ELS, of Loudon Health Communications provided editorial assistance.*

#### References

- Selby J. Why research at KP? Perm J 2005 Winter;9(1):10.
- Selby JV. Linking automated databases for research in managed care settings. Ann Intern Med 1997 Oct 15;127(8 Pt 2):719-24.
- Beaverson JM, Ryu J. Quality at Kaiser Permanente: using the population care model. Md Med 2011;12(2):15,17.
- Carroll NM, Ellis JL, Lockett CF, Raebel MA. Improving the validity of determining medication adherence from electronic health record medications orders. J Am Med Inform Assoc 2011 Sep-Oct;18(5):717-20.
- Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009 Mar-Apr;28(2):323-33.
- Wu JJ, Black MH, Smith N, Porter AH, Jacobsen SJ, Koebnick C. Low prevalence of psoriasis among children and adolescents in a large multiethnic cohort in southern California. J Am Acad Dermatol 2011 Nov;65(5):957-64.
- Koebnick C, Smith N, Coleman KJ, et al. Prevalence of extreme obesity in a multiethnic cohort of children and adolescents. J Pediatr 2010 Jul;157(1):26-31.
- Smith N, Iyer RL, Langer-Gould A, et al. Health plan administrative records versus birth certificate records: quality of race and ethnicity information in children. BMC Health Serv Res 2010 Nov 23;10:316.
- Sy LS, Liu IL, Solano Z, et al. Accuracy of influenza vaccination status in a computer-based immunization tracking system of a managed care organization. Vaccine 2010 Jul 19;28(32):5254-9.
- Getahun D, Fasset MJ, Jacobsen SJ. Gestational diabetes: risk of recurrence in subsequent pregnancies. Am J Obstet Gynecol 2010 Nov;203(5):467.
- Langer-Gould A, Albers KB, Van Den Eeden SK, Nelson LM. Autoimmune diseases prior to the diagnosis of multiple sclerosis: a population-based case-control study. Mult Scler 2010 Jul;16(7):855-61.
- Raebel MA, Smith ML, Saylor G, et al. The positive predictive value of a hyperkalemia diagnosis in automated health care data. Pharmacoeconom Drug Saf 2010 Nov;19(11):1204-8.
- Chen E, Martin AD, Matthews KA. Understanding health disparities: the role of race and socioeconomic status in children's health. Am J Public Health 2006 Apr;96(4):702-8.
- 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 Oct;93(10):1655-71.
- 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 Nov 1;164(9):823-34.
- Smith N, Iyer RL, Langer-Gould AM, et al. Health plan administrative records versus birth certificate records: quality of race and ethnicity information in children. BMC Health Serv Res 2010 Nov 23;10:316.
- Chen W, Petitti DB, Enger S. Limitations and potential uses of census-based data on ethnicity in a diverse community. Ann Epidemiol 2004 May;14(5):339-45.
- Li C, Balluz LS, Okoro CA, et al; Centers for Disease Control and Prevention (CDC). Surveillance of certain health behaviors and conditions among states and selected local areas—Behavioral Risk Factor Surveillance System, United States, 2009. MMWR Surveill Summ 2011 Aug 19;60(9):1-250.
- Committee on Comparative Effectiveness Research Prioritization. Initial national priorities for comparative effectiveness research. Washington, DC: The National Academies Press, 2009.
- Raphael JL, Beal AC. A review of the evidence for disparities in child vs adult health care: a disparity in disparities. J Natl Med Assoc 2010 Aug;102(8):684-91.