

Socio-economic status and childhood leukemia in California

Abstract

Objective: To investigate the association between individual- and community-level socio-economic status (SES) and childhood leukemia, particularly its subtypes: acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML).

Study design: We conducted a large case-control registry-based study in California.

Methods: Information on 5788 cases and 5788 controls was obtained by linking California cancer and birth registries. We evaluated relative risk of childhood leukemia by community- (census-based) and individual-level SES measures (parental education and source of payment for delivery) using conditional logistic regression.

Conclusion: We found little evidence to support the suggestion that SES is associated with risk of childhood leukemia or either major subtype. It is likely that results of many previous studies that found an association between childhood leukemia and SES were influenced by selection or ecological bias.

Keywords: childhood leukemia, socio-economic status, census-based SES, parental education, father's education, P education, source of payment for delivery

Volume 3 Issue 4 - 2015

Oksuzyan S,¹ Crespi CM,² Cockburn M,³ Mezei G,⁴ Vergara X,⁴ Kheifets L⁵

¹Department of Public Health, Office of Health Assessment and Epidemiology Unit, USA

²Department of Preventive Medicine, University of Southern California, USA

³Department of Biostatistics, UCLA Fielding School of Public Health, USA

⁴Environment Department, Electric Power Research Institute, USA

⁵Department of Epidemiology, UCLA Fielding School of Public Health, USA

Correspondence: Sona Oksuzyan, PhD, MD, MPH, County of Los Angeles Department of Public Health, Office of Health Assessment and Epidemiology Unit, 313 N Figueroa Street, Suite 217, Los Angeles, CA 90012 USA, Tel 1-(213)-240-8088, Email soksuzyan@ph.lacounty.gov

Received: October 28, 2015 | **Published:** November 09, 2015

Abbreviations: SES, socio-economic status; CCR, California cancer registry; CBR, California birth registry; PCA, principal component analysis; AIC, akaike information criterion; AML, acute myeloid leukemia

Introduction

Leukemia accounts for the largest number of cases of childhood cancer in the United States and worldwide.¹ Risk factors for this disease remain mostly unidentified.

The role of socio-economic status (SES) in the development of childhood leukemia remains poorly understood. To elucidate this association, some studies used maternal or paternal education, occupation, or family income as proxies for individual-level SES.²⁻⁶ Other studies have used community-level SES, for example, census tract or ward-based socio-economic characteristics, e.g. income, population density, rent, employment, community-level occupational class, and social deprivation score.⁷⁻¹² A small number of studies have looked at SES at both individual and community levels.¹³⁻¹⁵ Poole et al. reviewed literature on SES and childhood leukemia and concluded that in studies using individual-level measures (family income, mother's and father's education) that required some subject contact, lower SES was linked to higher risk of leukemia.¹⁶ In contrast, in record-based studies using father's occupational class and in ecologic studies using community-level occupational class, higher SES was associated with higher risk of childhood leukemia. The authors suggested that these apparent contradictions may have resulted from differences in design and/or SES measure, but could not distinguish between these two hypotheses. One study by McMahon (1962), used medical payment status as a proxy measure for SES to examine mortality data from 1947 to 1960.¹⁷ Two other studies used the source of payment for prenatal care as a proxy for SES in research on preterm birth.^{18,19}

In our study, using population-based registry data from California not involving contact with subjects, we explore relationships between individual-level SES (father's and mother's education and payment source for medical services) and community-level SES (census block-based SES) with childhood leukemia and examine whether the relationships between individual- and community-level SES and childhood leukemia risk are different from each other, as hypothesized in the review by Poole et al.

Materials and methods

The population-based California Cancer Registry (CCR) was used to obtain information on all childhood leukemia cases diagnosed between 1988 and 2008 at younger than 16 years of age and born and residing in California at the time of diagnosis. The statewide cancer registry regularly records age, race/ethnicity, sex, and residence at the time of diagnosis as well as information on cancer types and characteristics. Controls were selected randomly from the California Birth Registry (CBR) for years 1986-2007 and matched to cases (1 to 1) on the basis of date of birth (± 6 months) and sex. Controls were children younger than 16 years old at the time of "pseudo-diagnosis," who had not been diagnosed in California with any type of cancer.

Information on gender, parental age (in years), perinatal characteristics (e.g. birth weight, birth order), maternal and paternal years of education, sources of payment for prenatal care and delivery, and race/ethnicity were extracted from California birth records.

Measures of socio-economic status (SES)

Individual-level measures: We used parental education and source of payment for delivery as proxies for SES at the individual level. We categorized parental years of education into four levels: <12 years (less than high school), 12 years (high school), 13-16 years (some college,

college), and 17 years and more (graduate school). Since maternal and paternal years of education were correlated, we used them in different models. We also combined fathers and mother's years of education into one variable using principal component analysis (PCA) and then used the resulting principal component score in some models.

There were two variables for source of payment available from birth records: one for prenatal care and second for delivery. The source of payment for delivery was considered a better predictor of SES than source of payment for prenatal care (personal communication, Gerald Kominski, Professor, Department of Health Services, UCLA School of Public Health, Associate Director, UCLA Center for Health Policy Research). In addition, we compared estimates and 95% confidence intervals for sources of payment for prenatal care and delivery for similarity. Because they were nearly identical, the latter was used in all analyses. We categorized source of payment for delivery using two approaches. In one approach, we created three categories for source of payment for delivery: low SES (including governmental programs such as Medicare, Medi-Cal, Title V, other governmental programs, Indian Health Services, "No care" and "Medically indigent"), middle SES (military and worker's compensations, self-pay, CHAMPUS/TRICARE, no charge and other sources of payment) and high SES (Blue Cross/Blue Shield, private insurance, and health maintenance organization/prepaid health plan). In a second classification, the middle and high categories were combined to yield a total of two categories: low (same as before) and middle-high (combining the highest two in the previous classification) SES.

Community-level measures: A measure of census-based SES was provided by our collaborators from USC. The variable was derived from U.S. Census data using principal components analysis based on seven indicator variables at a census block group level: education index developed by Liu et al. [20] that weights the proportion of people with a given level of education by a number of years need to attain that level, proportion with a blue-collar job, proportion older than 16 without a job, median household income, proportion below 200% of the federal poverty level, median rent, median home value. Available to us were the quintiles of the PCA scores.²¹

Census-based SES was assessed using the address at birth obtained from CBR and the address at diagnosis from CCR. Since individual-level SES was available only at birth and the address at diagnosis was available only for cases, the main focus of the study was SES at birth. For subjects whose birth address could not be resolved to a census tract, census-based SES could not be obtained and was unknown. For cases only, we performed a comparative analysis of census-based SES at birth and at diagnosis.

We adjusted all models for child's race. Although racial origin of the child was collected in birth records, more than 50% of values were missing. We used more complete variables encoding racial origin of mother and father from birth records to derive the race of the child and combined those into four main racial groups: White, Black, Asian and Other. More detailed information for race/ethnicity is described elsewhere.²²

Statistical analysis: As a part of descriptive analysis, we examined relationships among father's education, mother's education, source of payment for delivery and census-based SES using kappa statistics and Spearman correlation and the distribution of individual-level proxies for SES by census-based SES.

We tested the association of childhood leukemia and variables of interest using unadjusted and adjusted conditional logistic

regression.²³ Since results were not substantially different, we present results of adjusted analyses only. Various models with different subsets of covariates were fit and checked for potentially influential observations. For adjusted analysis, the models for SES were chosen based on information on known or potential confounders and model fit statistics; models with the lowest Akaike information criterion (AIC) value and lowest number of potential confounders are presented.

To test whether census-based SES or individual level SES measures were predictive of childhood leukemia risk, we also conducted joint significance tests of the sets of coefficients associated with the SES variables (null hypothesis that all coefficients equal zero vs. alternative that at least one is nonzero).

Since risk factors for the two main subtypes of childhood leukemia, acute lymphocytic leukemia (ALL) and acute myeloid leukemia (AML) could be different, analyses were also conducted for these subtypes separately.

Despite the large number of cases and controls, sample sizes were smaller for some analyses due to missing data. Most missingness was due to variation in the variables collected on birth records from year-to-year, and no systematic differences in patterns of missingness were detected between cases and controls. As a sensitivity analysis, multivariate imputation techniques were used to impute missing or unknown values for father's and mother's education, payment source for delivery and census-based SES under a missing at random assumption using the MI procedure in SAS.²⁴⁻²⁶ The imputation models included all variables used in the analytic models. Analyses were repeated using the multiply imputed data using the MIANALYZE procedure.²⁴ Analyses were conducted using SAS 9.3.²⁴

Results

A total of 6645 childhood leukemia cases were identified from the CCR. Linkage to birth records was successful for 87.1% (5788/6645) of cases. Of the 5788 cases (55.8% males and 44.2% females) included in this analysis, 4721 were ALL cases (56.2% males and 43.8% females), 852 were AML cases (53.3% males and 46.7% females), and 215 were other childhood leukemia subtypes. The median age at diagnosis was 3.8 years with a range of 0 to 15.6 years with the peak for ALL between 2 and 5 years of age and between 0 and 2 years of age for AML. Table 1 shows other characteristics of study subjects.

To compare individual- and community-level measures, we present the distribution of father's education and mother's education, the source payment for delivery and census-based SES by each other presented in Figure 1. As shown in the figure, the proportion of high census-based SES increased and the proportion of low census-based SES decreased with increasing years of paternal and maternal education. A similar pattern was exhibited for the relationship between source of payment for delivery and census-based SES: the proportion of governmental programs decreased and the proportion of private insurance and other sources increased with increasing census-based SES.

Kappa statistics indicated slight agreement between father's education and census-based SES and between father's education and source of payment for delivery (weighted kappa=0.17 for both). The highest weighted kappa was observed between mother's and father's education (kappa=0.62). (Table 2) Similar results were obtained using Spearman correlation coefficients.

In adjusted analysis, we detected a tendency toward slightly decreased risk of total childhood leukemia and ALL in children of the

middle and middle-high census-based SES compared to the lowest SES class, but the associations were imprecise. Compared to children whose parents (mother and father) were least educated (<12years of school), children with parents having higher education levels (13-16 and ≥17years) had slightly decreased risk of total childhood leukemia and ALL although confidence intervals were wide. Sizable decreased risk of AML was noted for children with higher educated fathers compared to those whose fathers were least educated (<12years of school) (Table 3).

When we used a principal component score for father's and

mother's years of education, results did not differ from separate models containing either mother's education or father's education (OR=0.96, 95% CI: 0.89-1.04).

After performing complete case analyses, analyses were repeated using multiply imputed data. Results (not presented) were very similar to complete case analysis, except the association of father's education and AML. The negative association found in the complete case analysis was not replicated in analysis with multiply imputed data. For 13-16 years of paternal education, OR=0.78 with 95% CI: 0.51-1.17, and for 17years of education, OR=0.82 with 95% CI: 0.47-1.44.

Table 1 Characteristics of study subjects, California birth registry, 1986-2007

Variables	Cases (%)	Controls (%)	ALL # Cases/Controlsa	AML # Cases/Controlsa
All	5788	5788	4721	852
Census-Based SES Quintile				
Lowest	1264 (26.6)	1249 (26.3)	1023/1009	184/195
Lower-Middle	1126 (23.7)	1131 (23.8)	918/914	167/182
Middle	969 (20.4)	990 (20.9)	816/809	128/135
Higher-Middle	713 (15.0)	749 (15.8)	559/609	122/107
Highest	672 (14.2)	628 (13.2)	557/522	87/84
Unknown	1044	1041	848/858	164/149
Source for Payment for Delivery (Classification 1)a				
Governmental Programs (Low SES)	2173 (44.03)	2284 (46.43)	1741/1861	346/350
Military, Worker's Compensation, Self Pay, Other (Medium SES)	157 (3.18)	170 (3.46)	137/135	17/28
Private Insurance, Prepaid Plans (High SES)	2605 (52.79)	2465 (50.11)	2159/2031	361/339
Missing (Not Collected)*	853 (845)	869 (850)	684/694	128/135
Father's Education				
<12 Years	2552 (62.3)	2510 (61.7)	2056/2034	405/378
12 Years	641 (15.7)	649 (15.9)	544/548	79/83
13-16 Years	646 (15.8)	661 (16.2)	554/562	68/76
≥17 Years	258 (6.3)	251 (6.2)	213/208	35/36
Missing (Not Collected)*	1691 (1509)	1717 (1505)	1354/1369	265/279
Mother's Education				
<12 Years	801 (33.1)	784 (32.7)	644/624	127/131
12 Years	690 (28.5)	651 (27.1)	583/548	90/86
13-16 Years	696 (28.8)	726 (30.2)	586/617	84/84
≥17 Years	232 (9.6)	239 (10.0)	192/197	32/34
Missing (Not Collected)*	3369 (3335)	3388 (3340)	2716/2735	519/517

* Patterns of missingness varied by year due to differences in data collection year-to-year

Table 2 Kappa statistics for agreement between individual- and community-level measures of socio-economic status of childhood leukemia cases and their matched controls, California birth registry, 1986-2007

	Kappa statistics (Adjusted)	Level of agreement	Level of agreement
Father's Education and Source of Payment for Delivery	0.17	Slight	Poor
Father's Education and Census-Based SES	0.17	Slight	Poor
Father's Education and	0.62	Substantial	Good

Table continued...

	Kappa statistics (Adjusted)	Level of agreement	Level of agreement
Mother's Education			
Source of Payment for Delivery and Census-Based SES	0.29	Fair	Poor
Mother's Education and Source of Payment for Delivery	0.27	Fair	Poor
Mother's Education and Census-Based SES	0.27	Fair	Poor

Table 3 Conditional odds ratios (95% CIs) for individual level and community level socio-economic status and childhood leukemia matched on child's age and sex, California birth registry, 1986-2007^a

	All Types			ALL			AML		
	OR	95% CI		OR	95% CI		OR	95% CI	
Community-Level Measures									
Census-Based SES (5 categories)									
Reference - Lowest	1	-	-	1	-	-	1	-	-
Low-middle	1.06	0.92	1.21	1.07	0.92	1.24	1.01	0.72	1.44
Middle	0.92	0.79	1.05	0.99	0.84	1.15	0.72	0.5	1.05
Middle-High	0.87	0.74	1.01	0.83	0.7	0.98	1.13	0.76	1.68
Highest	0.95	0.81	1.11	0.95	0.8	1.14	1.07	0.69	1.66
p-value for trend	0.093								
Individual-Level Measure									
Source for Payment for Delivery (Classification 1)									
Governmental Programs (low SES)	1	-	-	1	-	-	1	-	-
Private Insurance (Middle-High SES)	1.04	0.95	1.14	1.08	0.97	1.19	1.03	0.82	1.3
(Classification 2)									
Governmental Programs (low SES)	1	-	-	1	-	-	1	-	-
Military, Self-Pay, and Other (Middle SES)	0.94	0.74	1.19	1.08	0.83	1.4	0.58	0.3	1.12
Private Insurance (High SES)	1.05	0.96	1.15	1.08	0.97	1.19	1.07	0.84	1.35
Father Education									
<12 Years	1	-	-	1	-	-	1	-	-
12 Years	0.96	0.82	1.14	1.02	0.85	1.22	0.6	0.38	0.95
13-17 Years	0.89	0.75	1.05	0.93	0.78	1.12	0.59	0.37	0.94
17 and + Years	0.9	0.71	1.13	0.92	0.71	1.18	0.75	0.4	1.39
p-Value for Trend	0.26								
Mother Education									
<12 Years	1	-	-	1	-	-	1	-	-
12 Years	1.05	0.9	1.24	1.09	0.92	1.3	0.79	0.5	1.27
13-17 Years	0.88	0.74	1.04	0.88	0.73	1.06	0.76	0.46	1.26
17 and + Years	0.87	0.68	1.1	0.89	0.68	1.17	0.62	0.31	1.22
p-Value for Trend	0.11								

^a Adjusted for child's race, birth order, birth weight and mother's age

Discussion

We assessed the association of childhood leukemia and socio-economic status (SES) using individual and community level measures. We used father's education and mother's education as individual-level SES proxies similar to previous research. As a departure from previous work, we defined a new way to assess individual-level SES, namely, payment source for medical services, such as delivery. We

were also able to assess the association between individual-level and community-level measures for this population-based sample.

Many studies that used individual-level proxies for SES have detected an inverse association with childhood leukemia, i.e. low SES was associated with higher risk of leukemia.^{2,4-6,16,27-29} The majority of these studies assessed individual-level SES using self-administered questionnaires or interviews. In such studies, selection bias may occur

because controls of high SES are usually more likely to participate [30]. A major strength of our study was that data were obtained from population registries with almost complete registration of births and cancers in California. Also, controls were randomly selected from the CBR rather than directly involving participants, as in many case-control studies. Since these registries are independent of each other and participation of subjects was not required for our data collection, selection bias was unlikely.

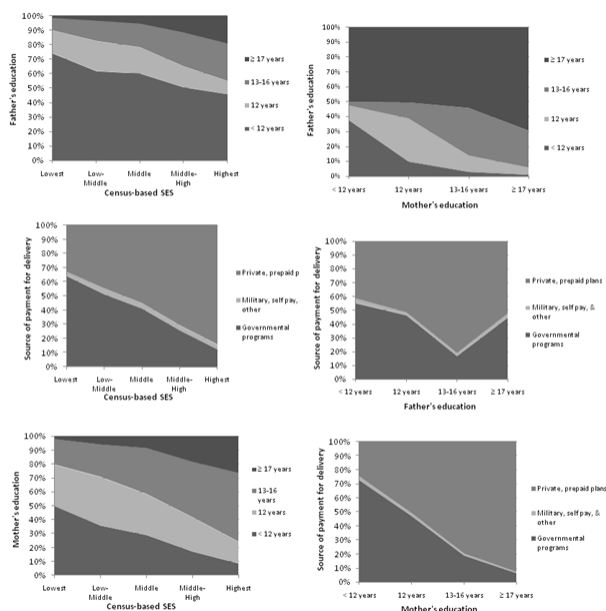


Figure 1 The distribution of individual- and community-level SES by each other; California birth registry, 1986-2007.

- Father's education by census-based SES and by Mother's education.
- Source of payment for delivery by census-based SES and by Father's education.
- Mother's education by census-based SES and by Source of payment for delivery.

Although the majority of studies that used community-level proxies for SES found an increased risk of childhood leukemia for high SES, some studies had reported an inverse association with childhood leukemia (higher risk of leukemia for low SES).^{7,8,14,16} Many of the studies that detected a positive association were purely ecological in design and thus were prone to ecological bias. We detected a tendency toward negative association between childhood leukemia and census-based SES.

High risk of childhood leukemia associated with higher SES would be more consistent with the viral hypothesis suggested by Greaves.^{31,32} In this hypothesis, late exposure to common infections, potentially associated with higher SES, could lead to abnormal immune response and play an important role in development of childhood leukemia.^{10,11} An inverse association of SES with childhood leukemia is more consistent with a "population mixing" theory.³³ According to this theory, some childhood leukemia cases could be a rare immune response to unidentified infections introduced by high level of personal contacts and/or large migration of new people into "closed" communities by increased commuting and travel ("population mixing") that is potentially associated with lower SES.¹⁰

Another advantage of this study was that the large size of the dataset allowed us to carry out analyses for two main subtypes of childhood leukemia, ALL and AML. We did not find a clear association of any SES measure with either subtype, which may further indicate that SES is not a determinant of childhood leukemia.

One of the limitations of this study is potential misclassification of variables of interest and covariates. Misclassification of the outcome was unlikely due to high accuracy of the CCR.³⁴ Misclassification of SES was possible due to potential inaccuracy of reporting and/or categorization of parental education and sources of payment for delivery. Parental education, particularly maternal education, was not consistently recorded on birth certificates, its completeness varied from year to year, and we do not know how accurate it is, but we believe that misclassification of parental education was not differential for cases and controls.³⁵ We also think that sources of payment for delivery were reported fairly accurately because medical entities relied on this information for payment for delivery care.³⁶ Misclassification could possibly arise from our categorizations of this variable. However, in our sensitivity analysis of various categorizations of these two classifications, our results were unaltered.

Another limitation of the study was missing data. However, since information was missing mainly due to year-to-year differences in the information collected on birth certificates rather than non-response, the potential for systematic biases was probably small and the impact was mainly on the precision of the estimates. For example, information on parental education was available only for more recent years. Since census-based SES did not come from CBR, we assessed if subjects that were missing census-based SES were different from subjects with this variable. We found that subjects with missing census-based SES had slightly younger maternal age at birth, slightly lower paternal education and higher proportion of governmental programs as source of payment for delivery. There was no difference in the pattern of missingness between cases and controls for census-based SES nor for any other variable. Therefore, missingness patterns were not differential and probably did not bias our results. We re-analyzed the data using multiple imputations and obtained very similar results, except for the association for father's education and AML, which became weaker and less precise.

Overall, we found little support for the suggestion that SES, as measured by individual- or community-level proxies, is associated with the risk of childhood leukemia or of its major subtypes. It is likely that previous studies finding an association between childhood leukemia and SES were prone to selection or ecological bias.

Acknowledgments

This project was supported by a research contract from the Electric Power Research Institute to UCLA and by UCLA Faculty Grants Program. Crespi was also partially supported by National Institutes of Health grant P30 CA16042. The authors want to thank California Department of Public Health for providing support and access to birth registry. The study was approved by University of California, Los Angeles Office for the Protection of Research Subjects and California Committee for the Protection of Human Subjects.

Conflicts of interest

The authors declare there is no conflict of interests.

Funding

None.

References

- Ma SK, Wan TS, Chan LC. Cytogenetics and molecular genetics of childhood leukemia. *Hematol Oncol*. 1999;17(3):91–105.
- Dockerty JD, Elwood JM, Skegg DC, et al. Electromagnetic field exposures and childhood cancers in New Zealand. *Cancer Causes Control*. 1998;9(3):299–309.
- Dockerty JD, Skegg DC, Elwood JM, et al. Infections, vaccinations, and the risk of childhood leukaemia. *Br J Cancer*. 1999;80(9):1483–1489.
- Ma X, Buffler PA, Selvin S, et al. Daycare attendance and risk of childhood acute lymphoblastic leukaemia. *Br J Cancer*. 2002;86(9):1419–1424.
- McBride ML, Gallagher RP, Theriault G, et al. Power–frequency electric and magnetic fields and risk of childhood leukemia in Canada. *Am J Epidemiol*. 1999;149(9):831–842.
- Reynolds P, Von Behren J, Elkin EP. Birth characteristics and leukemia in young children. *Am J Epidemiol*. 2002;155(7):603–613.
- Muirhead CR. Childhood leukemia in metropolitan regions in the United States: a possible relation to population density? *Cancer Causes Control*. 1995;6(5):383–388.
- Reynolds P, Elkin E, Scalf R, et al. A case–control pilot study of traffic exposures and early childhood leukemia using a geographic information system. *Bioelectromagnetics*. 2001;Suppl 5:S58–S68.
- Richardson S, Monfort C, Green M, et al. Spatial variation of natural radiation and childhood leukaemia incidence in Great Britain. *Stat Med*. 1995;14(21–22):2487–24501.
- Stillier CA, Kroll ME, Boyle PJ, et al. Population mixing, socioeconomic status and incidence of childhood acute lymphoblastic leukaemia in England and Wales: analysis by census ward. *Br J Cancer*. 2008;98(5):1006–1011.
- Borugian MJ, Spinelli JJ, Mezei G, et al. Childhood leukemia and socioeconomic status in Canada. *Epidemiology*. 2005;16(4):526–531.
- Ribeiro KB, Buffler PA, Metayer C. Socioeconomic status and childhood acute lymphocytic leukemia incidence in Sao Paulo, Brazil. *Int J Cancer*. 2008;123(8):1907–1912.
- Dockerty JD, Draper G, Vincent T, et al. Case–control study of parental age, parity and socioeconomic level in relation to childhood cancers. *Int J Epidemiol*. 2001;30(6):1428–1437.
- McKinney PA, Juszczak E, Findlay E, et al. Pre–and perinatal risk factors for childhood leukaemia and other malignancies: a Scottish case control study. *Br J Cancer*. 1999;80(11):1844–1851.
- Smith A, Roman E, Simpson J, et al. Childhood leukaemia and socioeconomic status: fact or artefact? A report from the United Kingdom childhood cancer study (UKCCS). *Int J Epidemiol*. 2006;35(6):1504–1513.
- Poole C, Greenland S, Luetters C, et al. Socioeconomic status and childhood leukaemia: a review. *Int J Epidemiol*. 2006;35(2):370–384.
- MacMahon B. Prenatal x–ray exposure and childhood cancer. *J Natl Cancer Inst*. 1962;28:1173–1191.
- Ghosh JK, Wilhelm MH, Dunkel–Schetter C, et al. Paternal support and preterm birth, and the moderation of effects of chronic stress: a study in Los Angeles county mothers. *Archives of women's mental health*. 2010;13(4):327–338.
- Lewis C, Suffet IH, Hoggatt K, et al. Estimated effects of disinfection by–products on preterm birth in a population served by a single water utility. *Environ Health Perspect*. 2007;115(2):290–295.
- Liu L, Deapen D, Bernstein L. Socioeconomic status and cancers of the female breast and reproductive organs: a comparison across racial/ethnic populations in Los Angeles County, California (United States). *Cancer Causes Control*. 1998;9(4):369–380.
- Yost K, Perkins C, Cohen R, et al. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. *Cancer Causes Control*. 2001;12(8):703–711.
- Oksuzyan S, Cockburn M, Crespi CM, et al. Race/ethnicity and childhood leukemia: a case–control study in California. *J Epidemiol Community Health*. 2013;69(8):795–802.
- Hosmer DW, Lemeshow S. Applied Logistic Regression. 3rd edn. New York, USA: Wiley; 2000.
- SAS Institute I. SAS Statistical Software (9.1 edn). Cary NC2006.
- Rubin DB, Schenker N. Multiple imputation in health–care databases: an overview and some applications. *Stat Med*. 1991;10(4):585–598.
- Little R, Rubin D. Statistical Analysis with Missing Data. 2nd edn. New York, USA: Wiley; 2000.
- Kaye SA, Robison LL, Smithson WA, et al. Maternal reproductive history and birth characteristics in childhood acute lymphoblastic leukemia. *Cancer*. 1991;68(6):1351–1355.
- Perrillat F, Clavel J, Auclerc MF, et al. Day–care, early common infections and childhood acute leukaemia: a multicentre French case–control study. *Br J Cancer*. 2002;86(7):1064–1069.
- Shu XO, Linet MS, Steinbuch M, et al. Breast–feeding and risk of childhood acute leukemia. *J Natl Cancer Inst*. 1999;91(20):1765–1772.
- Mezei G, Kheifets L. Selection bias and its implications for case–control studies: a case study of magnetic field exposure and childhood leukaemia. *Int J Epidemiol*. 2006;35(2):397–406.
- Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer*. 2006;6(3):193–203.
- Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia*. 1988;2(2):120–125.
- Kinlen LJ, Stillier C. Population mixing and excess of childhood leukemia. *BMJ*. 1993;306(6882):930.
- The California Cancer Registry. Mission Statement & Purpose Sacramento, CA; 2009.
- Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd edn. Philadelphia, USA: Lippincott Williams & Wilkins; 2008.
- Baumeister L, Marchi K, Pearl M, et al. The validity of information on “race” and “Hispanic ethnicity” in California birth certificate data. *Health services research*. 2000;35(4):869–883.