Title
Race/ethnicity and the risk of childhood leukaemia: a case-control study in California.

Journal
Journal of epidemiology and community health, 69(8)

ISSN
0143-005X

Authors
Oksuzyan, Sona
Crespi, Catherine M
Cockburn, Myles
et al.

Publication Date
2015-08-01

DOI
10.1136/jech-2014-204975

Peer reviewed

Race/ethnicity and the risk of childhood leukaemia: a case–control study in California

Sona Oksuzyan,[1] Catherine M Crespi,[2] Myles Cockburn,[3] Gabor Mezei,[4] Ximena Vergara,[5] Leeka Kheifets[1]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jech-2014-204975).

[1]Department of Epidemiology, UCLA Fielding School of Public Health, Los Angeles, California, USA
[2]Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, California, USA
[3]Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA
[4]Department of Epidemiology and Computational Biology, Exponent Engineering and Scientific Consulting, Menlo Park, California, USA
[5]Environment Department, Electric Power Research Institute, Palo Alto, California, USA

Correspondence to
Dr Sona Oksuzyan, University of California, Los Angeles, Semel Institute, Centre for Community Health, 10920 Wilshire Blvd., Suite 350, Los Angeles, CA 90024, USA; soksuzyan@ucla.edu

Received 19 September 2014
Revised 17 February 2015
Accepted 23 February 2015
Published Online First 19 March 2015

ABSTRACT

Background We conducted a large registry-based study in California to investigate the association between race/ethnicity and childhood leukaemia focusing on two subtypes: acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML).

Methods We obtained information on 5788 cases and 5788 controls by linking California cancer and birth registries. We evaluated relative risk of childhood leukaemia by race and ethnicity of the child and their parents using conditional logistic regression, with adjustment for potential confounders.

Results Compared with Whites, Black children had lower risk of ALL (OR=0.54, 95% CI 0.45 to 0.66) as well as children of Black/Asian parents (OR=0.31, 95% CI 0.10 to 0.94). Asian race was associated with increased risk of AML with OR=1.643, 95% CI 1.10 to 2.46 for Asian vs Whites; and OR=1.67, 95% CI 1.04 to 2.70 for Asian/Asian vs White/White. Hispanic ethnicity was associated with increased risk of ALL (OR=1.37, 95% CI 1.22 to 1.52). A gradient in risk of ALL was observed while comparing Hispanic children with both parents Hispanic, one parent Hispanic and non-Hispanic children (p Value for trend <0.0001). The highest risk of ALL was observed for children with a combination of Hispanic ethnicity and White race compared with non-Hispanic whites (OR=1.27, 95% CI 1.12 to 1.44). The lowest risk was observed for non-Hispanic blacks (OR=0.46, 95% CI 0.36 to 0.60). Associations for total childhood leukaemia were similar to ALL.

Conclusions Our results confirm that there are ethnic and racial differences in the incidence of childhood leukaemia. These differences indicate that some genetic and/or environmental/cultural factors are involved in aetiology of childhood leukaemia.

INTRODUCTION

A limited number of studies have specifically examined race and/or ethnicity in relation to childhood leukaemia risk.[1 2] Most studies on childhood leukaemia have considered race and/or Hispanic ethnicity only as a covariate in their analyses. One interview-based study found that the proportion of Whites among controls was higher than among childhood leukaemia cases[3] while another study found no association between race and childhood leukaemia;[4] both studies were prone to biases. The majority of studies have shown that Black race was associated with decreased risk of childhood leukaemia[1 2 5–10] compared with Whites. The definition of race/ethnicity differed in all these studies.

Fewer studies have looked at relationships between race/ethnicity and risk of major subtypes of childhood leukaemia, acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML). Most of these studies observed similar results for total childhood leukaemia and ALL.[1 2 6 11–13] Two studies found no association with any subtype.[4 13] For AML the findings varied: two studies found increased risk associated with White race, one study detected higher risk for Asian children,[6] others did not find any association.[12]

Most studies on Hispanic origin of the child have report an increased risk of total childhood leukaemia and ALL for Hispanic ethnicity;[1 11 14 15] however, some studies found no association between Hispanic origin and the risk of childhood leukaemia.[4 16]

The aims of this large-scale study were to examine the relationships between race/ethnicity of child, mother and father and childhood leukaemia and its subtypes. California is particularly suitable for studying racial and ethnic differences in the incidence of childhood leukaemia due to its diverse racial/ethnic distribution. In addition to being one of the most diverse states in the USA,[17] California has cancer and birth registries that have almost complete (99%) registration.[18 19]

Most previous record-based studies used a single definition of child's race and/or ethnicity. In our study, we explored several definitions of child's race and ethnicity and their combination. Registry-based child's race/ethnicity may have sizeable amount of missing data. We reconstructed child's race and ethnicity from the mother's and father's race and ethnicity and hence, reducing missing data.

MATERIALS AND METHODS

Eligible childhood leukaemia cases included in this analysis were diagnosed between 1988 and 2008 in California-born children younger than 16 years and who had resided in California at the time of diagnosis. Information about cases, cancer types and characteristics was extracted from the population-based California Cancer Registry (CCR). Cancer registry was linked to the California Birth Registry (CBR) to select controls and to obtain information on sociodemographic and other factors. Paper birth certificates were obtained from CBR for years prior to 1997 when no electronic birth records existed. Owing to high cost of each certificate and low number of cases prior to 1986, we restricted selection of cases and controls to1986–2007 birth years. Controls were selected randomly from CBR and matched to cases (1 to 1) on date of birth (±6 months) and sex.

To cite: Oksuzyan S, Crespi CM, Cockburn M, *et al*. *J Epidemiol Community Health* 2015;**69**:795–802.

Measures

In CCR, child's race and ethnicity was available for cases only; maternal and paternal race was not available. Therefore, we used CBR to extract information on child's and parental race/ethnicity. Racial groups for parents available from CBR were as follows: White, Black, American Indian, Asian-unspecified and Asian-specified, Chinese, Japanese, Korean, Vietnamese, Cambodian, Thai, Laotian, Indian, Filipino, Guamanian, Samoan, Eskimo, Aleut, Pacific Islander (PI), Hawaiian, and Other. Following the Surveillance, Epidemiology and End Results (SEER) classification, these groups were combined into five main racial categories as follows: White, Black, Asian (Asian-unspecified and Asian-specified, Chinese, Japanese, Korean, Vietnamese, Cambodian, Thai, Laotian, Indian and Filipino), American Indian and Alaskan (Eskimo, Aleut) and PI and Other (Guamanian, Samoan Hawaiian and all others).[20] These five categories were used in the analysis for the mother's and father's race.

Although child's race was available in CBR, more than 30% of values were missing. Significantly fewer individualss were missing information on parental race (see table 1).

We, therefore, created classifications of child's race based on the race of birth parents as recorded in child's birth certificate. We used several alternative approaches to classify child's race based on aforementioned five categories. By the reconstructed classification, a child was considered White if both parents were White; Black if either parent was Black; Asian if both parents were Asian or if one of the parents was White and another was Asian; American Indian and Alaskan if both parents were American Indian or Alaskan or if one parent was American Indian/Alaskan and the other was either White, or Asian, or PI and Other; and PI and Other if both parents were PI or Other, if one parent was PI or Other and the second was either White or Asian.

To check how sensitive our results were to differences in classification of race in some analyses we used the original child's race from birth records categorised into the same five racial groups. Findings did not vary; therefore, we present results for the reconstructed child race.

Similar to a system used by Chow *et al*,[16] we constructed a 14-category variable for child's race consisting of the combinations of parental races: White/White, Black/Black, Asian/Asian, American Indian/American Indians, PI and Other/PI and Other, White/Black, White/Asian, White/American Indian, White/PI and Other, Black/Asian, Black/American Indian, Black/PI and Other, Asian/American Indian, Asian/PI and Other (see table 2).

Parental Hispanic ethnicity was considered as a dichotomous variable separately from racial categorisation: Hispanic (Mexican, Puerto Rican, Cuban, Central/South American, Other Hispanic) and non-Hispanic. Hispanic origin of child was not recorded in registries and was derived from parental Hispanic ethnicity from CBR. A child was classified as Hispanic if either parent was Hispanic. Analyses were also conducted that classified children of Hispanic ethnicity as having one or both parents Hispanic.

For some analyses, we combined child's race and Hispanic ethnicity to create a variable with 10 categories: non-Hispanic white, Hispanic white, non-Hispanic black, Hispanic black, non-Hispanic Asian, Hispanic Asian, non-Hispanic American Indian, Hispanic American Indian, non-Hispanic PI and Other and Hispanic PI and Other.

We used parental education as a proxy for socioeconomic status (SES) at the individual level. Although maternal education was available from CBR, it was missing for about 60% of individuals. Therefore, we used paternal education to adjust for SES. It was categorised into four levels: <12, 12, 13–16 and 17 years and more. We used a measure of community-based SES derived from US Census data using principal components analysis based on seven indicator variables at a census block level.[21] Available to us were component scores from principal component analysis for SES index categorised into quintiles. We adjusted models for father's education and census-based SES separately.

Statistical analysis

The primary analysis method was conditional logistic regression utilising the matched case–control pairs. Several models using different subsets of covariates were fitted and checked for potentially influential observations. The models chosen were based on information on known or potential confounders and model fit statistics; the most parsimonious models with the lowest Akaike information criterion values are presented. Models for child, maternal and paternal race/ethnicity were fit separately to avoid close correlations. We have also considered models with interaction between child's race and other variables in the model.

Large sample size of the study allowed us to conduct analyses for two main subtypes of childhood leukaemia (ALL and AML) using the same models.

Despite the large number of cases and controls, sample sizes for some analyses were reduced due to missing data. Owing to differences in data collection by year, the pattern of missingness varied by year but no differences in patterns of missingness were detected between cases and controls. Under a missing at random assumption, multivariate imputation techniques were used to impute missing values for all variables in models by the MI procedure in SAS.[22–24] Analyses were repeated using the multiply imputed data in the MIANALYZE procedure.[24]

Analyses were conducted using SAS V.9.3.[24]

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.

RESULTS

A total of 6645 childhood leukaemia cases were identified from the CCR. Linkage to birth records was successful for 87.1% of cases. Of the 5788 cases (55.8% males and 44.2% females) included in this analysis, 4721 were ALL cases, 852 were AML cases and 215 were other types. The median age at diagnosis was 3.8 years (range 0–15.4) with the peak for ALL at 2–5 years and AML at 0–2 years of age. Table 1 shows other characteristics of study subjects.

Results of conditional logistic regression analyses assessing the association of childhood leukaemia with child's and parental races, presented in the table 3, indicate that Black race of child and of mother and father were each associated with a decreased risk of childhood leukaemia compared with White race. Similar findings were observed for ALL. For AML, increased but imprecise risk was observed for Black children, Asian children and children of Asian fathers. Adjusted analysis of a child's race, as defined by the father's and mother's races combined, showed similar results to the main analysis with reconstructed 5-category child race.

We repeated the analyses adjusting for census-based SES instead of father's education. Results were similar, except that the association of Asian race with AML was more precise. Children of Asian fathers had higher risk of AML compared with children of White fathers (OR=1.75, 95% CI 1.13 to

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

Variables	Cases (%)	Controls (%)	ALL cases/controls*, n	AML cases/controls*, n
All	**5788**	**5788**	**4721**	**852**
Birth weight (g)				
<2000	73 (1.3)	119 (2.1)	56/96	13/21
2000–3000	1009 (17.4)	1071 (18.5)	796/872	177/159
3000–4000	3941 (68.1)	3953 (68.3)	3232/3217	565/590
≥4000	764 (13.2)	644 (11.1)	636/535	97/82
Missing	1	1	1/1	0/0
Birth order				
First	2223 (38.5)	2333 (40.4)	1848/1896	309/356
Other	3559 (61.5)	3449 (59.7)	2878/2819	542/496
Missing	6	6	3/6	1/0
Mother's age (years)				
<25	1885 (32.6)	2067 (35.7)	1530/1689	296/301
25–35	3025 (52.3)	2973 (51.4)	2495/2430	415/435
35–45	863 (14.9)	740 (12.8)	687/594	136/116
≥45	14 (0.2)	7 (0.1)	9/7	0/0
Missing	1	1	0/1	0/0
Father's education (years)				
<12	2552 (62.3)	2510 (61.7)	2056/2034	405/378
12	641 (15.7)	649 (15.9)	544/548	79/83
13–16	646 (15.8)	661 (16.2)	554/562	68/76
≥17	258 (6.3)	251 (6.2)	213/208	35/36
Missing (not collected)†	1691 (1509)	1717 (1505)	1354/1369	265/279
Socioeconomic status				
Low	2390 (50.4)	2380 (50.1)	1941/1923	351/377
Middle	969 (20.4)	990 (20.9)	816/809	128/135
High	1385 (29.2)	1377 (29.0)	1116/1131	209/191
Missing	1044	1041	848/858	164/149
Child's race				
White	4550 (81.9)	4339 (78.8)	3777/3559	610/622
Black	290 (5.2)	490 (8.9)	198/399	76/68
Asian	614 (11.1)	569 (10.3)	480/462	105/87
American Indians and Alaskan	81 (1.5)	82 (1.5)	67/67	12/9
PI and Other	24 (0.4)	27 (0.5)	19/22	5/4
Missing	229	281	180/212	44/62
Mother's race				
White	4859 (84.7)	4675 (81.6)	4020/3815	665/686
Black	230 (4.0)	416 (7.3)	154/342	62/56
Asian	578 (10.1)	550 (9.6)	457/452	94/79
American Indians and Alaskan	33 (0.6)	37 (0.7)	28/31	5/5
PI and Other	37 (0.6)	52 (0.9)	30/41	7/8
Missing	51	58	32/40	19/18
Father's race				
White	4706 (84.6)	4514 (81.9)	3910/3697	630/649
Black	274 (4.9)	458 (8.3)	186/375	73/62
Asian	507 (9.1)	467 (8.5)	389/381	91/71
American Indians and Alaskan	31 (0.6)	30 (0.5)	27/26	4/2
PI and Other	43 (0.8)	40 (0.7)	31/32	10/6
Missing	227	279	178/210	44/62
Hispanic origin of child				
Both parents Hispanics	2481 (43.7)	2204 (39.1)	2074/1809	331/314
One parent Hispanic	633 (11.1)	603 (10.7)	513/485	93/94
Both parents non-Hispanic	2569 (45.2)	2827 (50.2)	2047/2304	411/417
Missing	105	154	87/123	17/27
Hispanic origin of mother				
Hispanic	2858 (49.6)	2559 (44.5)	2380/2082	387/381
Non-Hispanic	2899 (50.4)	3190 (55.5)	2317/2607	458/464
Missing	31	39	24/32	7/7

Continued

Table 1 Continued

Variables	Cases (%)	Controls (%)	ALL cases/controls*, n	AML cases/controls*, n
All	5788	5788	4721	852
Hispanic origin of father				
Hispanic	2737 (49.1)	2452 (44.3)	2281/2021	368/341
Non-Hispanic	2842 (50.9)	3087 (55.7)	2269/2501	451/459
Missing	209	258	171/199	33/52

*Number of cases and controls for ALL and AML do not add up for the total number of cases and controls for childhood leukaemia because there were few other subtypes in the data set.
†Patterns of missingness varied by year due to differences in data collection for different years.
ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; PI, Pacific Islander.

2.72). Asian children had higher risk of AML in all analyses (OR=1.64, 95% CI 1.10 to 2.46 for Asian vs Whites and OR=1.67, 95% CI 1.04 to 2.70 for Asian/Asian vs White/White children).

Analysis with Hispanic ethnicity considered independent of race revealed that Hispanic origin of child and of parents was associated with increased risk of total childhood leukaemia and ALL. Estimates and CIs for parental Hispanic ethnicity were almost identical to those for child; therefore, in table 4 we presented results for child's Hispanic ethnicity only.

A trend in risk of total childhood leukaemia and ALL was observed when comparing Hispanic children with both parents Hispanic, one parent Hispanic and non-Hispanic children, with the highest risk observed for children with both parents Hispanic (p Value for the trend <0.0001).

Models with interactions between child's race and birth order, mother's age, SES proxies and child's Hispanic ethnicity were also considered. Interactions were detected between child's race and ethnicity (p Value=0.02). For further investigation, we combined child's race and Hispanic ethnicity. The highest OR for total childhood leukaemia and for ALL was observed in Hispanic white and the lowest in non-Hispanic black children compared with non-Hispanic whites. Non-Hispanic Asians were at slightly increased risk of total leukaemia, ALL and AML, but with imprecise estimates (table 5).

After performing complete case analyses, these analyses were repeated using multiply imputed data. Results were very similar to complete case analysis. Some of these results are presented in online supplementary table S1.

DISCUSSION

Our results on the relationships between childhood leukaemia and parental or child's race indicate that, compared with White race, being Black was highly protective for the development of childhood leukaemia, particularly for ALL. This association was observed for paternal, maternal and child's races, regardless of the classification used. Although there were several studies that observed similar results,[1 2 5–8 11 16] our study had a larger sample size, used several classifications for child's race, did not have issues with subject selection and, additionally, included maternal and paternal races.

One possible explanation of this association, as some researchers have suggested, could be the underlying SES;[25 26] others suggest the associations may be explained by low birth weight among Blacks.[27–30] Many studies have shown that high birth weight was associated with increase in risk of childhood leukaemia;[31 32] consequently, low birth weight could have a protective effect on incidence of childhood leukaemia. However, since we controlled for SES and for birth weight in our study, these factors are unlikely to account for the finding.

We observed an association between Asian race and AML after adjusting for census-based SES. Elevated risk of AML for Asian children was less precise after adjustment for father's education. The association of AML with Asian race was also observed by Reynolds *et al*[6] in the unadjusted analysis and the estimate became imprecise when adjusted for father's education. We cannot offer any specific explanation for these findings, but as noted by many researchers, AML may have different risk factors than ALL.

We observed that Hispanic ethnicity of parents and child was associated with approximately 1/3 increase in risk of total childhood leukaemia and ALL. We observed a trend in the risk of total childhood leukaemia and ALL when we compared Hispanic children with both parents Hispanic, one parent Hispanic and non-Hispanic children, with the highest risk observed for children with both parents Hispanic. Our findings are in line with results of other studies showing increased risk of childhood leukaemia for children of Hispanic origin.[15]

Table 2 Race of child based on combination of parental races, California birth registry, 1986–2007

	Father's race						
Mother's race	Missing	White	Black	Asian	American Indian	PI and Other	Total
Missing	105	3	1	0	0	0	109
White	312	8889	147	114	44	28	9534
Black	41	41	556	5	2	1	646
Asian	36	217	19	852	1	3	1128
American Indian	4	48	4	0	14	0	70
PI and Other	8	22	5	3	0	51	89
Total	506	9220	732	974	61	83	11 576

PI, Pacific Islander.

Table 3 Conditional ORs (95% CIs) for childhood leukaemia and race of child (N=7982), mother (N=8096) and father (N=7984) matched on child's age and sex and adjusted for birth order, birth weight, mother's age and father's education

	All types			ALL			AML		
	OR	95% CI		OR	95% CI		OR	95% CI	
Child's race (reconstructed)									
White	1.00		–	1.00	–	–	1.00	–	–
Black	0.54	0.45	0.66	0.45	0.36	0.57	1.19	0.75	1.90
Asian	0.99	0.85	1.16	0.94	0.80	1.12	1.26	0.85	1.89
American Indian and PI	0.96	0.65	1.40	0.99	0.66	1.50	1.22	0.40	3.73
Other	0.75	0.37	1.49	0.79	0.36	1.69	0.74	0.12	4.51
Child's race (combined parental race)									
White/white	1.00	–	–	1.00	–	–	1.00	–	–
Black/black	0.51	0.40	0.64	0.40	0.30	0.53	1.35	0.77	2.37
Asian/Asian	1.04	0.87	1.24	0.98	0.81	1.20	1.34	0.83	2.17
American Indian/American Indian	0.36	0.07	1.80	0.19	0.02	1.57	*	*	*
PI and Other/PI and Other	0.76	0.38	1.51	0.80	0.37	1.72	0.75	0.12	4.58
White/black	0.74	0.52	1.05	0.69	0.46	1.04	1.10	0.50	2.39
White/Asian	0.89	0.68	1.17	0.85	0.63	1.15	1.13	0.57	2.21
White/American Indian	1.29	0.79	2.11	1.31	0.78	2.20	1.46	0.23	9.14
White/PI and Other	0.59	0.28	1.25	0.58	0.24	1.42	0.93	0.20	4.21
Black/Asian	0.31	0.10	0.94	0.23	0.07	0.80	†	†	†
Black/American Indian	*	*	*	*	*	*	*	*	*
Black/PI and Other	0.31	0.03	2.82	0.30	0.03	2.73	*	*	*
Asian/American Indian	*	*	*	*	*	*	†	†	†
Asian/PI and Other	2.93	0.30	28.35	*	*	*	†	†	†
Mother's race									
White	1.00	–	–	1.00	–	–	1.00	–	–
Black	0.53	0.43	0.66	0.43	0.33	0.56	1.24	0.74	2.08
Asian	0.98	0.84	1.15	0.95	0.80	1.12	1.13	0.74	1.73
American Indian and PI	0.73	0.41	1.30	0.65	0.35	1.21	1.22	0.20	7.55
Other	0.67	0.38	1.18	0.72	0.39	1.35	0.69	0.16	2.92
Father's race									
White	1.00	–	–	1.00	–	–	1.00	–	–
Black	0.55	0.45	0.67	0.45	0.36	0.57	1.30	0.80	2.11
Asian	1.01	0.86	1.20	0.95	0.79	1.14	1.42	0.92	2.19
American Indian and PI	1.07	0.57	2.01	1.18	0.60	2.31	1.03	0.13	8.17
Other	0.83	0.49	1.42	0.89	0.49	1.62	0.93	0.24	3.52

California birth registry, 1986–2007.
*Not estimable due to small cell counts.
†No observations.
ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; PI, Pacific Islander.

Table 4 Conditional ORs (95% CIs) for childhood leukaemia and Hispanic origin of child matched on child's age and sex and adjusted for birth order, birth weight, mother's age and father's education (N=8078)

	All types			ALL			AML		
	OR	95% CI		OR	95% CI		OR	95% CI	
Hispanic ethnicity of child									
Non-Hispanic	1.00	–	–	1.00	–	–	1.00	–	–
Hispanic	1.29	1.17	1.43	1.37	1.22	1.52	0.98	0.74	1.30
Non-Hispanic	1.00	–	–	1.00	–	–	1.00	–	–
One parent Hispanic	1.17	1.00	1.38	1.20	1.01	1.44	1.08	0.70	1.64
Both parents Hispanics*	1.33	1.20	1.49	1.42	1.26	1.60	0.96	0.71	1.29

California birth registry, 1986–2007.
*Trend test p value <0.0001.
ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia.

Table 5 Conditional ORs (95% CIs) for childhood leukaemia and combined child's race and Hispanic ethnicity, adjusted for birth order, birth weight, mother's age, father's education and matched on child's age and sex (N=7968)

Combined child's race and Hispanic ethnicity	All types			ALL			AML		
	OR	95% CI		OR	95% CI		OR	95% CI	
Non-Hispanic white	1	–	–	1	–	–	1	–	–
Hispanic white	1.23	1.1	1.38	1.27	1.12	1.44	1.11	0.8	1.53
Non-Hispanic black	0.58	0.46	0.72	0.46	0.36	0.6	1.23	0.72	2.11
Hispanic black	1.05	0.65	1.69	0.95	0.55	1.65	1.66	0.55	4.98
Non-Hispanic Asian	1.16	0.97	1.37	1.1	0.92	1.33	1.39	0.88	2.19
Hispanic Asian	0.91	0.58	1.42	0.76	0.45	1.27	1.53	0.55	4.28
Non-Hispanic American Indian	1.13	0.65	1.96	1.2	0.67	2.15	2.17	0.39	12.08
Hispanic American Indian	1.12	0.65	1.91	1.48	0.74	2.98	0.81	0.17	3.79
Non-Hispanic PI and Other	0.72	0.32	1.64	0.85	0.31	2.33	0.33	0.03	3.26
Hispanic PI and Other	1.32	0.35	4.98	0.92	0.39	2.15	*	*	*

California birth registry, 1986–2007.
*Not estimable due to small cell counts.
ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; PI, Pacific Islander.

Incidence in several Latin American countries is the highest in the world: 5.65 cases/100 000 in Costa Rica, 5.54/100 000 in Ecuador and 4.43/100 000 in Uruguay.[33] Analysis with combined race and Hispanic ethnicity showed that non-Hispanic black children had the lowest risk of childhood leukaemia compared with White non-Hispanics. The highest risk was observed in White Hispanic children.

Some researchers have suggested that nutrition and diet could contribute to racial and ethnic differences in cancer incidence.[34] Emerging evidence suggests that genetic risk factors may also explain the markedly different risk of childhood leukaemia in Hispanics and Blacks. A recent study by Xu *et al*[35] detected new susceptibility variants at 10p12.31–12.2 of the BMI1-PIP4K2A gene. This polymorphism is common in Hispanic ethnic groups and rare in black populations; it could, at least partially, explain the findings.[35]

Another potential explanation of observed associations for race, ethnicity and childhood leukaemia could be population mixing, in which immunologically naive susceptible individuals experience an increase in leukaemia incidence as population increases.[20 36–41] Muirhead[42] reported increasing incidence rates of childhood leukaemia with increasing population density in three US metropolitan areas, including San Francisco and California. During the years covered by this study, California experienced population growth[43] potentially allowing the opportunity for susceptible individuals to be exposed to some new infectious agents for which their immune systems have not been modulated, resulting in a rare abnormal response of childhood leukaemia.[44]

A major strength of our study was that data were obtained from nearly complete population registries and controls were randomly selected from the birth registry. Interview-based case–control studies are prone to selection bias. Meta-analysis by Slusky *et al*[45] has shown that interview-based case–control studies of childhood cancer could suffer from over-representation of Whites and under-representation of other races in participating controls. This is not a case for registry-based non-contact studies. Since birth and cancer registries were independent of each other and participation of patients was not required, selection bias due to participation was unlikely.

We matched on age and gender, and adjusted for other potential confounders available in registries; however, we cannot completely exclude residual confounding due to other unknown/unmeasured factors.

Another advantage of this study was that the large sample size allowed us to carry out analyses for two main subtypes of childhood leukaemia, ALL and AML. The risk patterns observed for AML were different from the risk patterns of ALL, and included higher risk for AML for Asian children. This highlights the importance of conducting disaggregated analyses by subtype.

Although there were two other studies on childhood leukaemia conducted in California, our study does not have a large overlap with those. One of the two studies (Reynolds *et al*[4]) included only children born in San Diego county during limited number of years (1988–1994), and had very limited sample size (90 cases/349 controls). The Northern California Childhood Leukaemia Study enrolled only newly diagnosed patients from hospitals in the northern part of California since 1995.[46] Although these cases may have some overlap with our study, it is unlikely that we had a large overlap for controls because of their random selection in both studies. Our study had a larger sample size, included all leukaemia cases from the whole of California from 1988 to 2008 and did not have any subject selection issues. The results of the study are generalisable in any racially/ethnically heterogeneous population.

A potential bias could arise if controls of a particular race/ethnicity selectively moved out of the state before their 'pseudo-diagnosis' date and became cases there. For example, if Hispanic children moved out of the state and became leukaemia cases outside of California, we would underestimate the association. These scenarios are very unlikely because the probability of controls moving out of California and subsequently developing childhood leukaemia, a rare cancer, is quite low. Additionally, the literature indicates higher mobility of childhood leukaemia cases compared with controls and not vice versa; therefore, this could not affect our results.[3 47 48]

One of the study limitations is potential misclassification on variables of interest and covariates. Misclassification of the outcome was unlikely due to the completeness and high accuracy of the CCR. Misclassification of race/ethnicity may have happened because it could be reported by parents, abstracted from a medical chart or recorded by hospital staff based on their own observations.[16] Nonetheless, in validation studies in California where birth certificate data were compared with structured postpartum interviews, the sensitivity of birth records to correctly identify most racial and ethnic groups was greater than 94%, with the exception of Native Americans.[49] Some

misclassification of race and ethnicity is still possible due to categorisation of these variables. We attempted to address this issue by examining sensitivity to reclassification which, reassuringly, did not alter our results. Even if misclassification was present, we believe that it was not different for cases and controls, which would pull the estimates towards the null in the case of binary categorisation.[50]

Another weakness of the study was missing data. To address the issue of missing values, at least partially, instead of only using records containing child's race that was missing for about a third of participants, we constructed a classification of child's race using more complete information on maternal and paternal races. For a majority of factors, missingness did not vary considerably by race and ethnicity. Black and Other races of children had slightly higher missing data on father's education. However, since information was missing mainly due to differences in the collected information between years rather than non-response and did not differ between cases and controls, the potential for biases was probably small and the impact was mainly on the precision of estimates. We reanalysed data using multiple imputations and obtained similar results with slightly narrower CIs.

In summary, we found that children of black race were at lower risk of childhood leukaemia and ALL. Hispanic ethnicity was associated with high risk of childhood leukaemia and ALL. A new finding was the association of Asian race and AML. Such ethnic and racial differences in incidence of childhood leukaemia indicate that some genetic and environmental/cultural factors may be involved in the aetiology of childhood leukaemia.

What is already known on this subject

Several studies noted decreased risk of childhood leukaemia for Black children compared with White children and increased risk for children of Hispanic ethnicity compared with non-Hispanics. Many of these studies had problems with subject selection, misclassification of race/ethnicity and missing data.

What this study adds

- This registry-based study examined relationships between race/ethnicity and childhood leukaemia in California. The large scale of the study allowed us to look at the association of race/ethnicity and the major subtypes of childhood leukaemia (acute lymphoblastic leukaemia and acute myeloid leukaemia) which was not possible in previous research. The risk pattern observed for these two subtypes was quite different and revealed new associations.
- Our study, being registry based, did not suffer from selection bias like the majority of previous studies.
- The study addressed potential misclassification of race/ethnicity by using several classifications for these.
- Missing data was at least partially addressed by using not only the original child's race from birth records but also a reconstructed child's race/ethnicity based on mother's and father's race/ethnicity. To the best of our knowledge, this type of analysis was never applied in previous research. In addition, we repeated our analyses using multiply imputed data sets which also helped in addressing the missing data issue.

Acknowledgements The authors want to thank California Department of Public Health for providing support and access to birth registry.

Contributors SO made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript and critical revision of the manuscript for important intellectual content. She is the corresponding author. CMC made substantial contributions to conception and design, analysis and interpretation of data. She participated in the drafting of the manuscript and critical revision of the manuscript for important intellectual content. MC made substantial contributions to conception and design, acquisition of data, and analysis and interpretation of data. He participated in critical revision of the manuscript for important intellectual content. GM made substantial contributions to conception and design, acquisition of data and interpretation of data. He participated in drafting of the manuscript and critical revision of the manuscript for important intellectual content. XV made substantial contributions to conception and design, and interpretation of data. She participated in drafting of the manuscript and critical revision of the manuscript for important intellectual content. LK made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data. She participated in drafting of the manuscript and critical revision of the manuscript for important intellectual content.

Funding This project was supported by a research contract from the Electric Power Research Institute (EPRI) to UCLA and by UCLA Faculty Grants Programme. CMC was also partially supported by National Institutes of Health grant P30 CA16042.

Competing interests None.

Ethics approval UCLA IRB and California CPHS.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Matasar MJ, Ritchie EK, Consedine N, *et al*. Incidence rates of the major leukemia subtypes among US Hispanics, Blacks, and non-Hispanic Whites. *Leuk Lymphoma* 2006;47:2365–70.
2 Okcu MF, Goodman KJ, Carozza SE, *et al*. Birth weight, ethnicity, and occurrence of cancer in children: a population-based, incident case-control study in the State of Texas, USA. *Cancer Causes Control* 2002;13:595–602.
3 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;149:831–42.
4 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):S58–68.
5 Macdougall LG, Jankowitz P, Cohn R, *et al*. Acute childhood leukemia in Johannesburg. Ethnic differences in incidence, cell type, and survival. *Am J Pediatr Hematol Oncol* 1986;8:43–51.
6 Reynolds P, Von Behren J, Elkin EP. Birth characteristics and leukemia in young children. *Am J Epidemiol* 2002;155:603–13.
7 Ross JA, Davies SM, Potter JD, *et al*. Epidemiology of childhood leukemia, with a focus on infants. *Epidemiol Rev* 1994;16:243–72.
8 Sandler DP, Ross JA. Epidemiology of acute leukemia in children and adults. *Semin Oncol* 1997;24:3–16.
9 Gurney JG, Severson RK, Davis S, *et al*. Incidence of cancer in children in the United States. Sex-, race-, and 1-year age-specific rates by histologic type. *Cancer* 1995;75:2186–95.
10 Linabery AM, Ross JA. Trends in childhood cancer incidence in the U.S. (1992–2004). *Cancer* 2008;112:416–32.
11 Dores GM, Devesa SS, Curtis RE, *et al*. Acute leukemia incidence and patient survival among children and adults in the United States, 2001–2007. *Blood* 2012;119:34–43.
12 Milne E, Laurvick CL, Blair E, *et al*. Fetal growth and acute childhood leukemia: looking beyond birth weight. *Am J Epidemiol* 2007;166:151–9.
13 Podvin D, Kuehn CM, Mueller BA, *et al*. Maternal and birth characteristics in relation to childhood leukaemia. *Paediatr Perinat Epidemiol* 2006;20:312–22.
14 McNeil DE, Cote TR, Clegg L, *et al*. SEER update of incidence and trends in pediatric malignancies: acute lymphoblastic leukemia. *Med Pediatr Oncol* 2002;39:554–7; discussion 52–3.
15 Campleman SL, Wright WE. Childhood Cancer in California 1988 to 1999. Volume I: Birth to Age 14. In: California Department of Health Services, ed. *Cancer Surveillance Section*. Sacramento, CA: California Cancer Registry, 2004:16–20.
16 Chow EJ, Puumala SE, Mueller BA, *et al*. Childhood cancer in relation to parental race and ethnicity: a 5-state pooled analysis. *Cancer* 2010;116:3045–53.
17 Humes K, Jones NA, Ramirez RR. *Overview of race and Hispanic origin: 2010. 2010 census briefs*. U.S. Census Bureau, 2011.
18 The California Cancer Registry. Mission Statement & Purpose. Secondary Mission Statement & Purpose. http://www.ccrcal.org/abouttheccr.html
19 Schoendorf KC, Branum AM. The use of United States vital statistics in perinatal and obstetric research. *Am J Obstet Gynecol* 2006;194:911–15.
20 Kinlen LJ. Infection and childhood leukemia. *Cancer Causes Control* 1998;9:237–9.

21 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;12:703–11.
22 Auvinen A, Hietanen M, Luukkonen R, *et al*. Brain tumors and salivary gland cancers among cellular telephone users. *Epidemiology* 2002;13:356–9.
23 Little R, Rubin D. *Statistical analysis with missing data*. 2nd edn. New York: Wiley, 2002.
24 SAS Institute I. SAS Statistical Software. 9.1 edn. Cary, NC: 2006.
25 Stiller CA, Parkin DM. Geographic and ethnic variations in the incidence of childhood cancer. *Br Med Bull* 1996;52:682–703.
26 Draper G. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966–83*. London: H.M. Stationery Office, 1991.
27 David RJ, Collins JW Jr. Differing birth weight among infants of U.S.-born blacks, African-born blacks, and U.S.-born whites. *N Engl J Med* 1997;337:1209–14.
28 Fang J, Madhavan S, Alderman MH. Low birth weight: race and maternal nativity—impact of community income. *Pediatrics* 1999;103:E5.
29 McGrady GA, Sung JF, Rowley DL, *et al*. Preterm delivery and low birth weight among first-born infants of black and white college graduates. *Am J Epidemiol* 1992;136:266–76.
30 Shiono PH, Klebanoff MA, Graubard BI, *et al*. Birth weight among women of different ethnic groups. *JAMA* 1986;255:48–52.
31 Caughey RW, Michels KB. Birth weight and childhood leukemia: a meta-analysis and review of the current evidence. *Int J Cancer* 2009;124:2658–70.
32 Hjalgrim LL, Westergaard T, Rostgaard K, *et al*. Birth weight as a risk factor for childhood leukemia: a meta-analysis of 18 epidemiologic studies. *Am J Epidemiol* 2003;158:724–35.
33 Howard SC, Metzger ML, Wilimas JA, *et al*. Childhood cancer epidemiology in low-income countries. *Cancer* 2008;112:461–72.
34 Ross JA, Kasum CM, Davies SM, *et al*. Diet and risk of leukemia in the Iowa Women's Health Study. *Cancer Epidemiol Biomarkers Prev* 2002;11:777–81.
35 Xu H, Yang W, Perez-Andreu V, *et al*. Novel susceptibility variants at 10p12.31–12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J Natl Cancer Inst* 2013;105:733–42.
36 Dickinson HO. The causes of childhood leukaemia. *BMJ* 2005;330:1279–80.
37 Dickinson HO, Parker L. Quantifying the effect of population mixing on childhood leukaemia risk: the Seascale cluster. *Br J Cancer* 1999;81:144–51.
38 Kinlen L. Infections and immune factors in cancer: the role of epidemiology. *Oncogene* 2004;23:6341–8.
39 Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* 1995;71:1–5.
40 Kinlen LJ, Stiller C. Population mixing and excess of childhood leukemia. *BMJ* 1993;306:930.
41 Law GR, Parslow RC, Roman E, *et al*. Childhood cancer and population mixing. *Am J Epidemiol* 2003;158:328–36.
42 Muirhead CR. Childhood leukemia in metropolitan regions in the United States: a possible relation to population density? *Cancer Causes Control* 1995;6:383–8.
43 Demographic Research Unit California Department of Finance, ed. Current Population Survey: California Two-Year Average Series. In: *California Two-Year Average Series*. Sacramento, CA, 2011:1–2.
44 Clark BR, Ferketich AK, Fisher JL, *et al*. Evidence of population mixing based on the geographical distribution of childhood leukemia in Ohio. *Pediatr Blood Cancer* 2007;49:797–802.
45 Slusky DA, Mezei G, Metayer C, *et al*. Comparison of racial differences in childhood cancer risk in case-control studies and population-based cancer registries. *Cancer Epidemiol* 2012;36:36–44.
46 Guha N, Ward MH, Gunier R, *et al*. Characterization of residential pesticide use and chemical formulations through self-report and household inventory: the Northern California Childhood Leukemia study. *Environ Health Perspect* 2013;121:276–82.
47 Green LM, Miller AB, Villeneuve PJ, *et al*. A case-control study of childhood leukemia in southern Ontario, Canada, and exposure to magnetic fields in residences. *Int J Cancer* 1999;82:161–70.
48 Michaelis J, Schuz J, Meinert R, *et al*. Childhood leukemia and electromagnetic fields: results of a population-based case-control study in Germany. *Cancer Causes Control* 1997;8:167–74.
49 Baumeister L, Marchi K, Pearl M, *et al*. The validity of information on "race" and "Hispanic ethnicity" in California birth certificate data. *Health Serv Res* 2000;35:869–83.
50 Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.