

# UCSF

## UC San Francisco Previously Published Works

### Title

Daycare attendance and risk of childhood acute lymphoblastic leukaemia

### Permalink

<https://escholarship.org/uc/item/2r4454bj>

### Journal

British Journal of Cancer, 86(9)

### ISSN

0007-0920

### Authors

Ma, X  
Buffler, PA  
Selvin, S  
[et al.](#)

### Publication Date

2002-05-01

### DOI

10.1038/sj.bjc.6600274

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Daycare attendance and risk of childhood acute lymphoblastic leukaemia

X Ma<sup>1</sup>, PA Buffler<sup>\*1</sup>, S Selvin<sup>1</sup>, KK Matthay<sup>2</sup>, JK Wiencke<sup>3</sup>, JL Wiemels<sup>3</sup> and P Reynolds<sup>4</sup>

<sup>1</sup>Division of Public Health Biology and Epidemiology, University of California, Berkeley, California, CA 94720-7360, USA; <sup>2</sup>Department of Pediatrics Oncology, University of California, San Francisco, California, CA 94143, USA; <sup>3</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, California, CA 94143, USA; <sup>4</sup>Environmental Health Investigations Branch, California Department of Health Services, Oakland, California, CA 94612, USA

The relationship between daycare/preschool ('daycare') attendance and the risk of acute lymphoblastic leukaemia was evaluated in the Northern California Childhood Leukaemia Study. Incident cases (age 1–14 years) were rapidly ascertained during 1995–1999. Population-based controls were randomly selected from the California birth registry, individually matched on date of birth, gender, race, Hispanicity, and residence, resulting in a total of 140 case–controls pairs. Fewer cases ( $n=92$ , 66%) attended daycare than controls ( $n=103$ , 74%). Children who had more total child–hours had a significantly reduced risk of ALL. The odds ratio associated with each thousand child–hours was 0.991 (95% confidence interval (CI): 0.984–0.999), which means that a child with 50 thousand child–hours (who may have, for example, attended a daycare with 15 other children, 25 h per week, for a total duration of 30.65 months) would have an odds ratio of  $(0.991)^{50}=0.64$  (95% CI: 0.45, 0.95), compared to children who never attended daycare. Besides, controls started daycare at a younger age, attended daycare for longer duration, remained in daycare for more hours, and were exposed to more children at each daycare. These findings support the hypothesis that delayed exposure to common infections plays an important role in the aetiology of childhood acute lymphoblastic leukaemia, and suggest that extensive contact with other children in a daycare setting is associated with a reduced risk of acute lymphoblastic leukaemia.

British Journal of Cancer (2002) 86, 1419–1424. DOI: 10.1038/sj/bjc/6600274 www.bjcancer.com  
© 2002 Cancer Research UK

**Keywords:** childhood leukaemia; daycare; delayed exposure

Despite decades of research, the aetiology of childhood leukaemia remains obscure. Greaves hypothesised that delayed exposure to common infections leads to an increased risk of childhood leukaemia, especially common pre-B acute lymphoblastic leukaemia (ALL) (cALL), which has a maximum incidence in 2–5 year olds (Greaves, 1988, 1997). The essence of Greaves' hypothesis is that two separate genetic events may be responsible for the development of cALL; the first event occurs spontaneously pre- or peri-natally during the expansion of B-cell precursors; the second occurs in the same mutant clone following antigenic challenge early in life. A delay in a child's exposure to common childhood infections may result in an improperly modulated immune system and subsequent risk of aberrantly high levels of lymphoblastic cell proliferation following the barrage of infections when the child enters daycare or preschool. Intrigued by a leukaemia cluster in Seascale, UK, Kinlen suggested that childhood leukaemia might result from a rare response to a common (but identified) infection. Increased risks would occur when populations from different geographic areas were mixed so that infected and susceptible individuals had an increased level of contact (Kinlen, 1988, 1998). The hypotheses by both Greaves and Kinlen are consistent with the proposition that leukaemia may be a rare response to a common infection or infections.

In developed countries, most of a child's exposure to common infections is from contact with other children. It has been well documented that daycare attendance increases the risk of infections, which seems to increase with the number of children in a group (Istre *et al*, 1985; Ferson, 1993; Nystad *et al*, 1999). Alexander suggested that daycare of the index child and of the siblings is probably the best available proxy measure of exposure to infections (Alexander *et al*, 1993). To date, daycare/preschool (referred to as 'daycare' from here on) attendance, along with other variables such as birth order and number of siblings, has been used as an indicator of exposure to common infections in several studies (Petridou *et al*, 1993, 1997; Dockerty *et al*, 1999; Infante-Rivard *et al*, 2000; Neglia *et al*, 2000; Rosenbaum *et al*, 2000). However, the exposure assessment was crude in these studies and the results are inconsistent. The Northern California Childhood Leukaemia Study (NCCLS), which collected detailed data on contact with other children in a daycare setting, offered an opportunity to critically explore the quantitative relationship between daycare attendance and the risk of ALL.

## METHODS

### Study population

The NCCLS is currently ongoing, and is composed of two phases. Phase I consists of cases diagnosed between January 1, 1995 and November 30, 1999, and Phase II is planned to include cases diagnosed between December 1, 1999 and December 31, 2003. The cases in this analysis are from Phase I. Incident cases of newly diag-

\*Correspondence: PA Buffler; E-mail: pab@uclink4.berkeley.edu  
Received 2 October 2001; revised 11 February 2002; accepted 4 March 2002

nosed childhood leukaemia (age 0–14 years) were prospectively ascertained (usually within 24 h after diagnosis) from major clinical centres. Although case ascertainment was hospital-based, a comparison with cases ascertained by the statewide California Cancer Registry (1997–1999) shows that the NCCLS protocol successfully identified 88% of all newly diagnosed childhood leukemia cases in the San Francisco–Oakland Metropolitan Statistical Area. Controls were randomly selected from the statewide birth certificate files maintained by the California Department of Health Services and 1:1 matched to cases on date of birth, gender, mother's race (white, black, or other), Hispanicity (either parent is Hispanic), and mother's county of residence at the time of child's birth. These controls were then traced by using commercially available searching tools. For cases not born in California (less than 10% of all cases), county of residence at diagnosis was used for the matching. For each case, the search continued until an eligible control subject consented to participate in the study.

To be eligible, each case or control had to (1) reside in the study area; (2) be under 15 years of age; (3) have at least one parent/guardian who spoke English or Spanish; and (4) have no previous history of any malignancy. Of a total of 273 eligible Phase I cases, 229 (83%) consented to participate. This analysis is based on 183 cases for whom a matched control subject was available at the time of data analyses. The characteristics (age, gender, race, ethnicity, subtype, maternal education, household income, and daycare attendance) of these 183 cases were comparable to the cases who were not included in this analysis. In order to enrol the 183 controls, a total of 475 searches were conducted, from which 355 (75%) potential controls were successfully located. Of the 355 potential controls, 279 (79%) were eligible, of which 183 (66%) consented to participate. The actual number of searches per case ranged from 1 to 15. Of the 183 controls, 121 (66%) were the first choices, i.e. the first eligible controls contacted. The study protocol was approved by the Internal Review Board of all collaborating institutions, and written informed consents were obtained for all participating subjects. Epidemiologic data were obtained by a self-administered questionnaire and a subsequent in-home personal interview. Of the 183 cases, 148 had a diagnosis of ALL. Since infant ALL (diagnosed at a age of less than 12 months) is often considered to have a distinctive aetiology, eight infant ALL cases were excluded, resulting in a total of 140 ALL case–control pairs for the daycare analysis.

## Data collection

Data on daycare attendance were first collected by a self-administered questionnaire and then verified during an in-home personal interview. The interviews were conducted with the primary care givers, usually the biological mother (97%) or biological father (2.5%). All data for daycare attendance were censored on the date of diagnosis (corresponding dates for controls) or 6 years of age, whichever was earlier.

For this analysis, the following variables were included or constructed: (1) attendance at daycare (yes/no); (2) age when first started daycare (in months); (3) duration of stay at all daycare facilities (in months), regardless of hours attended; (4) mean hours per week attending daycare facilities; (5) mean number of children exposed to at each daycare facility; (6) total number of children ever exposed to in a daycare setting; and (7) total child–hours, which is an overall summary of daycare attendance. If a child only attended one daycare facility, then the mean number of children would be the same as the total number of children. Child–hours at each daycare is calculated as follows: number of months attending a daycare  $\times$  mean hours per week at this daycare  $\times$  number of other children exposed to at this daycare multiplied by 4.35 (i.e. number of weeks per month). The child–hours at each daycare were added to obtain the total child–hours for each child. For children who never attended daycare, an age of 72 months was assigned as the age when

first started daycare, and a value of 0 was assigned as duration of stay, mean hours per week, mean number of children, total number of children, and total child–hours.

Data on the number of other children in household before the index child went to first grade were also collected. This was used as a potential indicator of exposure to infectious agents by contact with other children in the same household. In the United States, children usually start first grade when they are around 6 years of age.

**Table 1** Characteristics of 140 non-infant ALL cases and 140 matched controls

Characteristics	Cases n (%)	Controls n (%)	P-value $\chi^2$ test
Age <sup>a</sup>			–
12–23.9 months	10 (7)	10 (7)	
2–5 years	86 (61)	86 (61)	
6–10 years	34 (24)	34 (24)	
11–14 years	10 (7)	10 (7)	
Mean (s.e. <sup>b</sup> ) (years)	5.4 (0.3)	5.4 (0.3)	
Gender <sup>a</sup>			–
Male	81 (58)	81 (58)	
Female	59 (42)	59 (42)	
Race/Ethnicity <sup>a</sup>			–
Hispanic	43 (31)	43 (31)	
Non-Hispanic White	73 (52)	73 (52)	
Non-Hispanic Black	3 (2)	2 (1)	
Other	21 (15)	22 (16)	
Household income (K\$/year)			0.03
<15	15 (11)	12 (9)	
15–29.9	29 (21)	15 (11)	
30–44.9	20 (14)	13 (9)	
45–59.9	21 (15)	17 (12)	
60–74.9	19 (14)	24 (17)	
≥75	36 (26)	59 (42)	
Maternal education			0.80
≤High school	50 (36)	45 (32)	
>High school but <bachelor's	45 (32)	49 (35)	
≥Bachelor's	45 (32)	46 (33)	
Paternal education			0.68
≤High school	58 (42)	53 (38)	
>High school but <bachelor's	30 (22)	36 (26)	
≥Bachelor's	50 (36)	49 (36)	
Unknown	2	2	
Maternal age (years)			0.74
<20	9 (6)	9 (6)	
20–24	28 (20)	21 (15)	
25–29	34 (25)	42 (30)	
30–34	47 (34)	45 (32)	
≥35	21 (15)	23 (17)	
Unknown	1	0	
Mean (s.e. <sup>b</sup> ) (years)	28.9 (0.5)	29.4 (0.5)	
Breastfeeding (months)			0.81
Never	30 (22)	28 (20)	
1–2	18 (13)	24 (17)	
3–5	25 (18)	26 (19)	
6–11	38 (28)	32 (23)	
≥12	25 (18)	28 (20)	
Unknown	4	2	
Mean (s.e. <sup>b</sup> ) (months)	6.6 (0.7)	6.7 (0.7)	
Birth order			0.87
First	54 (40)	59 (43)	
Second	46 (34)	45 (33)	
Third or higher	36 (26)	34 (25)	
Unknown	4	2	

<sup>a</sup>These are matching variables. Age is age at diagnosis for cases and age at the corresponding dates for controls. <sup>b</sup>Standard error.

Statistical analysis

Pearson’s Chi-square test was used to compare the distributions of demographic, socio-economic, and birth characteristics between cases and controls. The matched design of the NCCLS dictates a matched analysis. Paired *t*-tests were used to compare daycare attendance measures between cases and controls, one at a time. A quantile–quantile plot, which is an important graphic technique for comparing shapes of distributions (Hoaglin *et al*, 1985), was used to illustrate the difference in total child–hours between cases and controls. Nonparametric Wilcoxon signed-rank test and Kolmogorov–Smirnov test, which are robust and free from assumption of the distribution of data, were also conducted. Hotelling’s multivariate *T*<sup>2</sup> test provided a single assessment of the overall difference between cases and controls, measured in terms of the six variables. Multivariate conditional logistic regression was employed to estimate odds ratios and 95% confidence intervals, adjusting particularly for annual household income.

RESULTS

Data on demographic, socio-economic, and birth characteristics of cases and controls were obtained from the in-home personal interviews, during which a respondent, usually the biological mother,

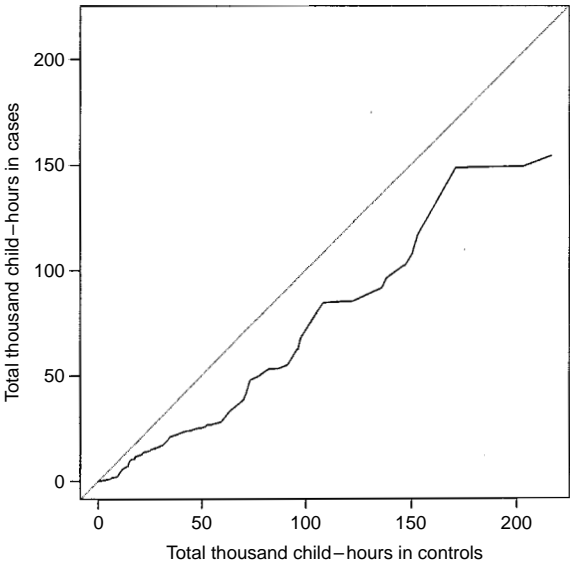
was asked to indicate the race and ethnicity of the case or control child (Table 1). As expected, there was only a minor difference on race and Hispanicity between cases and controls. The distribution of cases and controls were similar with respect to maternal education, maternal age, duration of breastfeeding, and birth order. Controls appeared to have higher annual household income (*P* value=0.03).  
Among the 140 pairs, 92 (65.7%) of the cases and 103 (73.6%) of the controls attended at least one daycare facility. Compared to the cases, controls started daycare at younger ages, stayed at daycare for longer, were exposed to more children in the daycare setting, and acquired more child–hours (Table 2). Generally, significance levels from Wilcoxon signed-rank tests, which do not assume normal distributions, were smaller than those from paired *t*-tests. The mean number of other children in the household before the index child went to first grade was slightly higher in cases than in controls. Although not all the tests were statistically significant, there was a consistent pattern.

The quantile–quantile plot (cases *vs* controls) clearly showed that cases had fewer total child–hours than the controls (Figure 1). If there were no difference in the distribution of total child–hours between cases and controls, the points would have randomly deviated from the dotted reference line, as opposed to clustering in one side of the reference line. A nonparametric Kolmogorov–

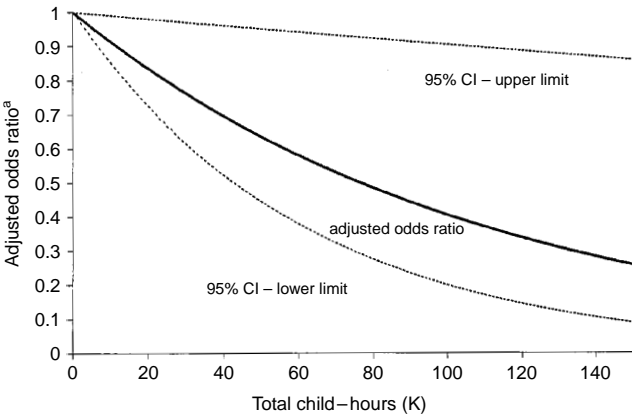
**Table 2** Age started first daycare, duration of daycare attendance, number of children exposed to, and total child-hours by case–control status

Variables	Cases mean (s.e. <sup>a</sup> )	Controls mean (s.e. <sup>a</sup> )	Control–Case Mean difference (95% CI)	P value (paired t-test)	P value (paired Wilcoxon signed rank test)
Age first started daycare (months)	41.7 (2.2)	35.7 (2.2)	–6.0 (–11.7, –0.3)	0.04	0.02
Months of stay (regardless of hours)	16.9 (1.6)	19.9 (1.8)	3.0 (–1.0, 6.9)	0.14	0.08
Mean hours per week	12.7 (1.2)	18.4 (1.4)	5.8 (2.3, 9.3)	0.001	0.001
Mean number of other children exposed to at each daycare	8.8 (0.7)	10.6 (0.8)	1.8 (–0.2, 3.8)	0.08	0.06
Total number of other children exposed to in a daycare setting	16.5 (1.7)	17.3 (1.5)	0.8 (–3.5, 5.1)	0.71	0.31
Total thousand child-hours	20.0 (2.8)	31.2 (3.9)	11.2 (2.9, 19.5)	0.01	0.002
Number of other children in household before the index child went to 1st grade	2.1 (0.2)	1.9 (0.2)	–0.2 (–0.7, 0.30)	0.48	0.59

<sup>a</sup>Standard error.



**Figure 1** Quantile–quantile plot of total thousand child–hours.



**Figure 2** Total child–hours and the risk of childhood ALL.

Smirnov test was conducted to compare the cumulative density functions of total child-hours between cases and controls, which produced a *P* value of 0.094. As expected, the other five daycare attendance variables in Table 2 followed the same pattern, since these daycare measures are all related.

Hotelling's multivariate  $T^2$  test was used to assess the overall difference between cases and controls, in terms of age started first daycare, duration of daycare attendance, mean hours per week, mean number of children at each daycare, total number of children at all daycares, and total child-hours. The  $T^2$  statistic was 3.74 with a corresponding *P* value of 0.003, which again indicates that one or more of the variables reflecting daycare exposure are associated with the risk of ALL.

The results presented above are not adjusted for household income. Since income is one of many measures of socio-economic status, and socio-economic status is possibly related to daycare attendance, household income was considered an important contributor in the conditional logistic regression analysis. Each unit of mean hours per week and total thousand child-hours is significantly associated with the risk of ALL, with and without controlling for household income (Table 3). The association between age first started daycare and ALL was statistically significant only when household income was not adjusted for. A decreased risk was also observed for children who stayed for longer, or exposed to more children, although the associations did not reach statistical significance.

The odds ratios displayed in Table 3 measure the change in risk associated with each unit of change of the independent variables. For example, each thousand child-hours is associated with an odds ratio of 0.991, which means that a child who has 50 thousand child-hours (who may have, for example, attended a daycare with 15 other children, 25 h per week, for a total duration of 30.65 months) would have an odds ratio of  $(0.991)^{50}=0.64$ , compared to children who never attended daycare (Figure 2).

A logistic regression model allows the simultaneous assessment of the impact of multiple daycare attendance variables on the difference in risk between cases and controls. In addition, the relative roles of each variable can be described and the confounding influence of income measured. The results from a conditional logistic analysis, based on a purely additive model employing the daycare attendance variables directly as measured, are presented in Table 4. Both mean hours per week and mean number of children exposed to at each daycare are excluded because they are highly correlated with other important variables included in the model. The exposure measured by total child-hours is the dominant predictor of case-control status ( $z=-2.612$ ; *P* value=0.009). The other three exposure variables add little additional independent information to the evaluation of association between daycare attendance and the risk of ALL. Income is an important predictor of difference in risk between cases and controls, but is only a slight confounder of the daycare attendance variables. For example, the coefficient associated with

total-child hours is  $-0.016$  when income is included in the model, and is slightly smaller ( $-0.013$ ) when income is not included, changing the *P* value from 0.009 to 0.022. The other daycare variables are similarly influenced by income, each showing a slightly less but still unimportant association with case-control status.

The protective effect of daycare attendance sustained in both Hispanic and White non-Hispanic subgroups. Separate analyses were conducted for 86 ALL cases who were diagnosed at 2–5 years old and are likely to be cases of cALL. The results derived from the analyses of these 86 case and birth certificate control pairs were not much different. The odds ratio adjusted for income is 0.991, which is the same as the odds ratio obtained from the analysis of the 140 case-control pairs. The odds ratio adjusted for age first started daycare, months of stay, total number of children, and income simultaneously is 0.985, which is very similar to the result based on the 140 case-control pairs (Table 4).

Daycare attendance did not appear to be associated with the risk of AML (OR=1.000, 95% CI: 0.984–1.016). However, due to the very limited number of AML case-control pairs available for analysis ( $n=35$ ), the result should be interpreted with caution.

Excluding cases who were born out of the state of California (less than 10% of all the cases) did not alter any of the results.

## DISCUSSION

Our observation that extensive contact with other children in a daycare setting is associated with a reduced risk of childhood ALL provides strong indirect support to the hypothesis that delayed exposure to common infections is associated with the risk of ALL. Although the magnitude of the association we observed between daycare attendance and the risk of ALL is not large, it is consistent, not likely to be confounded by income, and statistically significant. Daycare attendance only reflects exposure to common infections, and is certainly, at best, an indirect measure. There are also background exposures that a child may experience through sources other than daycare attendance, such as contact with other children who live in the same household or in the public settings

**Table 4** A multivariate conditional logistic regression model with 140 matched pairs

Variables	Odds ratio			Z statistic	P value
	Coefficient	ratio	Standard error		
Child-hours (K)	−0.016	0.984	0.006	−2.612	0.009
Age first started daycare	0.011	1.011	0.009	1.192	0.233
Months of stay	0.022	1.022	0.015	1.525	0.127
Total number of children	0.015	1.015	0.009	1.581	0.114
Income	−0.338	0.713	0.097	3.479	0.001

**Table 3** Daycare attendances and the risk of ALL

Variables	Crude OR <sup>a</sup> (95% CI)	P value	Adjusted OR <sup>a</sup> (95% CI)	P value
Age first started daycare (months)	1.011 (1.000, 1.021)	0.04	1.007 (0.996, 1.017)	0.22
Months of stay (regardless of hours)	0.989 (0.975, 1.004)	0.14	0.995 (0.980, 1.010)	0.53
Mean hours per week	0.974 (0.958, 0.991)	0.002	0.978 (0.961, 0.996)	0.01
Mean number of other children exposed to at each daycare	0.976 (0.946, 1.003)	0.08	0.980 (0.952, 1.010)	0.20
Total number of other children exposed to in a daycare setting	0.998 (0.985, 1.011)	0.71	1.001 (0.988, 1.015)	0.85
Total child-hours (thousands)	0.990 (0.983, 0.998)	0.01	0.991 (0.984, 0.999)	0.03

<sup>a</sup>The odds ratios were derived from conditional logistic regression models. Adjusted odds ratios were calculated controlling for annual household income.

(hospitals, parks, shopping malls, etc.), and infectious agents brought home by parents.

Household income may reflect to some extent possible differences in socio-economic status not controlled for by the matching on race and Hispanicity. No income data were available for the nonparticipating controls. However, we were able to obtain information on maternal education, a commonly used marker for socio-economic status, from the birth certificates of all potential controls, whether or not they participated. The eligible first-choice controls (i.e. ideal controls), regardless of participation, have very similar maternal education levels to the controls who actually participated in the study (i.e. actual controls).

Three previous studies also reported that a larger proportion of controls attended daycare than cases (Petridou *et al*, 1993, 1997; Infante-Rivard *et al*, 2000). In a Greek study, childhood leukaemia cases in metropolitan Athens were less likely than controls to have attended daycare (OR=0.7, 95% CI: 0.4–1.1), especially for attendance of at least 3 months prior to the age of 2 years (OR=0.3, 95% CI: 0.1–0.9) (Petridou *et al*, 1993). Another study in Greece reported a similar association comparing the proportion of children who ever attended any daycare among the cases and controls, but the association was not statistically significant (OR=0.83, 95% CI 0.51–1.37) (Petridou *et al*, 1997). A significantly protective effect of daycare attendance was observed in a population-based Canadian study with ALL cases aged 0–9 years. Compared with children who never attended daycare, the odds ratios for children who started daycare at  $\leq 2$  years old and  $> 2$  years old were 0.49 (95% CI: 0.31–0.77) and 0.67 (0.45–1.01), respectively (Infante-Rivard *et al*, 2000). No detailed information on daycare attendance, however, was reported in these three studies, which precluded directly comparing our study with theirs.

Dockerty *et al* (1999) found no significant association between attendance at particular types of preschool groups and ALL risk in New Zealand (OR=0.67, 95% CI 0.40–1.12). The relationship between daycare attendance and ALL was also evaluated in a recent study, which recruited cases from many Children's Cancer Group (CCG) member institutions across the United States (Neglia *et al*, 2000). Neither attendance at daycare, age first started daycare, or time enrolled at daycare, was associated with risk of overall ALL or cALL in the CCG study. Several explanations exist for the discrepancy between the CCG study and the NCCLS. Compared to the large, predominantly White CCG study population that included cases diagnosed at hospitals in many different states, our study population is more geographically defined, ethnically diverse, better educated, and has a much higher household income. In the CCG study, less than 50% of control subjects attended any daycare, while 74% of controls subjects in our study did so. Another difference between the two studies is the selection of control subjects. The CCG study used the method of random-digit dialling, while the NCCLS selected controls from population-based statewide birth registry. The most noteworthy difference, however, is that daycare attendance data obtained in the CCG study were not as detailed or complete as the data obtained in the NCCLS. For example, neither number of children the index child was exposed to nor the hours the child spent at each daycare facility was collected in the CCG study. None of the five variables evaluated in the CCG study, including any daycare, age at start of daycare, time at daycare, daycare before age two, time at daycare before age 2 years, were found to be important predictors of the risk of ALL in our study. Rather, it was the more specific indicators of exposure such as total child-hours that appeared to be associated with a significantly modified risk in our study. Despite the large sample size, the CCG study did not provide adequate data to evaluate the relationship between daycare attendance and the risk of childhood ALL. In a more recent study that included ALL cases diagnosed at one of the four referral centres in western and central New York

State, the odds ratio for children who stayed at home compared with those who attended daycare for greater than 36 months was 1.32 (95% CI: 0.70–2.52); the odds ratios for 1–18 and 19–36 months of daycare were 1.74 (95% CI: 0.89, 3.42) and 1.32 (95% CI: 0.64, 2.71), respectively (Rosenbaum *et al*, 2000). When the analysis was restricted to B-lineage ALL, the magnitude of association increased. The authors concluded, however, that their results did not support an association between daycare attendance and ALL, probably because the 95% confidence intervals of several odds ratios included 1.0. In fact, combining the data presented for B-lineage ALL in Table 5 in the publication by Rosenbaum *et al* (2000) produces a summary Mantel–Haenszel odds ratio of 1.63 (95% CI: 1.06–2.52), which indicates an overall association with the length of daycare attendance. Breaking continuous variables into categories, although conventional in epidemiological studies, does not use all the information available and certainly decreases statistical power under most circumstances (Selvin, 1996). Also, the choice of the number of categories and corresponding cut points can change the results, sometimes substantially. In addition to the difference in analytical strategy, our study and the study by Rosenbaum *et al* (2000) differed by participation (83% and 66% in cases and controls, respectively, vs 71% and 55% in cases and controls, respectively), as well as the method of data collection (in-home personal interview vs mailed self-administered questionnaire). In our study, the conditional logistic regression analysis indicates that total child-hours, which combine information from several measures of daycare attendance into a single value, effectively summarise the protective effect of attending daycare (Table 4).

A particular strength of our study is a detailed and timely exposure assessment. The NCCLS commenced in 1995. The investigators had the advantage of learning from and improving upon data collection instruments that were used and kindly made available by researchers from previous studies. Detailed data on when a subject started and stopped attending each daycare facility, number of hours on average he/she spent at the facility, as well as how many other children that he/she was exposed to at the facility, were obtained. As opposed to a telephone interview or a mailed questionnaire, all data on daycare attendance were collected first by a self-administered questionnaire and then verified at an in-home personal interview with the primary care giver of the subject, usually the biological mother. In addition, special efforts were made to rapidly ascertain incident cases and shorten the time between diagnosis (corresponding dates for controls) and exposure assessment. Bias pertaining to the difficulty of recalling should have been reduced. Another important feature of the NCCLS is the selection of population-based controls from statewide birth registry. Close collaboration with the California Department of Health Services resulted in an expeditious process of control selection. A series of methodological evaluations (manuscripts in preparation) indicate that the birth certificate controls are representative of the population base from which the cases arose.

A limitation of the study is the relatively small sample size, which reduces the power. On the other hand, the matched design and detailed daycare exposure measures should improve the precision of statistical analyses. The NCCLS is currently ongoing, and it will provide more information with regard to daycare attendance and the risk of ALL with the expansion of study population in the near future.

In summary, our study offered strong but indirect support to the hypothesis that delayed exposure to common infections plays an important role in the aetiology of childhood ALL. Daycare attendance appears to be associated with a reduced risk of childhood ALL. Our findings, if confirmed by future studies with detailed exposure assessment, will likely shed light on the immunological aetiology of childhood ALL and might eventually lead to preventative strategies.

## ACKNOWLEDGEMENTS

The study was supported by two research grants from the National Institute of Environmental Health Sciences, United States (PS42 ES04705 and RO1 ES09137). Dr Xiaomei Ma was supported by a grant from the Centers for Disease Control and Prevention, United States (T01-CCT917644-01). We thank Dr Jim Feusner at Children's Hospital Oakland, Dr Gary Dahl

at Packard Children's Hospital, Dr Stacy Month at Kaiser Oakland Hospital, Dr Vonda Crouse at Valley Children's Hospital, Dr Kenneth Leung at Kaiser San Francisco Hospital, and Dr Vincent Kiley at Kaiser Sacramento Hospital for assistance with recruiting patients, Monique Does for supervising field work and interviews, Dr Gladys Block, Dr David Ragland, and Kyndaron Reinier for helpful discussions, and Jennifer Ahern for technical support.

## REFERENCES

- Alexander FE, McKinney PM, Cartwright RA (1993) Migration patterns of children with leukaemia and non-Hodgkin's lymphoma in three areas of northern England. *J Public Health Med* **15**: 9–15
- Dockerty JD, Skegg DC, Elwood JM, Herbison GP, Becroft DM, Lewis ME (1999) Infections, vaccinations, and the risk of childhood leukaemia. *Br J Cancer* **80**: 1483–1489
- Ferson MJ (1993) Infections in day care. *Curr Opin Pediatr* **5**: 35–40
- Greaves MF (1988) Speculations on the Cause of Childhood Acute Lymphoblastic Leukemia. *Leukemia* **2**: 120–125
- Greaves M (1997) Aetiology of acute leukaemia. *Lancet* **349**: 344–349
- Hoaglin DC, Mosteller F, Tukey JW (1985) *Exploring Data, Table, Trends and Shapes*. pp 432–437. New York: John Wiley and Sons
- Infante-Rivard C, Fortier I, Olson E (2000) Markers of infection, breast-feeding and childhood acute lymphoblastic leukaemia. *Br J Cancer* **83**: 1559–1564
- Istre GR, Conner JS, Broome CV, Hightower A, Hopkins RS (1985) Risk factors for primary invasive Haemophilus influenzae disease: increased risk from day care attendance and school-aged household members. *J Pediatr* **106**: 190–195
- Kinlen L (1988) Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain [see comments]. *Lancet* **2**: 1323–1327
- Kinlen LJ (1998) Infection and childhood leukemia. [editorial; comment] *Cancer Causes Control* **9**: 237–239
- Neglia JP, Linet MS, Shu XO, Severson RK, Potter JD, Mertens AC, Wen W, Kersey JH, Robison LL (2000) Patterns of infection and day care utilization and risk of childhood acute lymphoblastic leukaemia. *Br J Cancer* **82**: 234–240
- Nystad W, Skrandal A, Magnus P (1999) Day care attendance, recurrent respiratory tract infections and asthma. *Int J Epidemiol* **28**: 882–887
- Petridou E, Kassimos D, Kalmanti M, Kosmidis H, Haidas S, Flytzani V, Tong D, Trichopoulos D (1993) Age of exposure to infections and risk of childhood leukaemia. *BMJ (Clinical Research Edn)* **307**: 774
- Petridou E, Trichopoulos D, Kalapothaki V, Pourtsidis A, Kogevinas M, Kalmanti M, Kolioukas D, Kosmidis H, Panagiotou JP, Piperopoulou F, Tzortzatou F (1997) The risk profile of childhood leukaemia in Greece: a nationwide case-control study. *Br J Cancer* **76**: 1241–1247
- Rosenbaum PF, Buck GM, Brecher ML (2000) Early child-care and preschool experiences and the risk of childhood acute lymphoblastic leukemia. *Am J Epidemiol* **152**: 1136–1144
- Selvin S (1996) Statistical power and sample-size calculations. In: *Statistical analysis of epidemiologic data*, Vol 25 Monographs in epidemiology and biostatistics, pp 95–100. New York: Oxford University Press