Are Incidence Rates of Adult Leukemia in the United States Significantly Associated with Birth Cohort?

Philip S. Rosenberg, Katherine L. Wilson, and William F. Anderson

Abstract

Background: Leukemia is a common cancer among U.S. adults but there are few established risk factors. If leukemia risks are substantially influenced by exposures that vary in prevalence across generations, then population incidence rates should vary significantly by birth cohort. However, prior studies have not examined leukemia birth cohort effects using contemporary data and methods.

Methods: We used incidence data from the National Cancer Institute's Surveillance, Epidemiology and End Results Program from 1992 through 2009 for adults 25–84 years old and age period cohort models to estimate incidence rate ratios according to birth cohort for acute myeloid leukemia (AML), acute lymphoid leukemia (ALL), chronic myeloid leukemia (CML), and chronic lymphoid leukemia (CLL).

Results: Leukemia incidence varied significantly between birth cohorts for each major leukemia type in men and women except female AMLs; changes on the order of 1% per birth year or 20% per generation were observed. The most significant birth cohort signatures were observed for CLLs and AMLs in men, which were decreasing and increasing, respectively, in cohorts born since 1946.

Conclusions: Our results support the hypothesis that adult leukemia risks are significantly modulated by environmental and lifestyle exposures.

Impact: A number of well-established (smoking, certain chemicals, radiation) and newly recognized (obesity) leukemia risk factors are modifiable; ultimately, efforts to promote healthy lifestyles might also help reduce incidence rates of adult leukemia. Cancer Epidemiol Biomarkers Prev; 21(12); 2159–66. ©2012 AACR.

Introduction

Leukemia (all types) is a common cancer among adults in the United States with more than 40,000 new cases expected in 2012 (seer.cancer.gov). Adult leukemia incidence rates have significant international variation (1), secular trends (1–3), and characteristic patterns according to age, gender, and ethnic group (1, 2, 4–6). Nonetheless, whereas the risks of secondary leukemias attributable to cancer chemotherapy and radiotherapy have been well characterized (7, 8), there are currently only a few established primary risk factors for any major leukemia type (7–10). Therefore, it remains unclear whether adult leukemias are largely a stochastic consequence of biologic senescence (11–13) or alternatively if a substantial fraction might potentially be preventable through modification of environmental and lifestyle risk factors.

Assessment of birth cohort effects provides a potentially informative approach to shed light on this question (14). If cumulative exposures or exposures during sensitive ages to known or unknown leukemia risk factors vary in prevalence from one generation to the next, then incidence rates in the population should also vary significantly by birth cohort (15). Cohort patterns for childhood leukemias have been described (16, 17); however, associations of birth cohort and adult leukemia rates have not been examined using contemporary data and methods. Therefore, we used nationally representative leukemia incidence data from the National Cancer Institute’s Surveillance, Epidemiology and End Results (SEER) Program (18) and age period cohort models (19) to assess incidence trends and patterns according to birth cohort (20).

Materials and Methods

Data

We obtained leukemia incidence data from the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) program for the 18 year period 1992–2009 (seer.cancer.gov). As previously described (21, 22), case and population data were combined from SEER’s 13 (23) and 18 (24) Registries Databases; collectively, SEER’s 18 Tumor Registries cover up to 28% of the U.S. population.

We conducted separate analyses for each of the 4 major leukemia types acute myeloid leukemia (AML), acute lymphoid leukemia (ALL), chronic myeloid leukemia (CML), and chronic lymphoid leukemia (CLL), as
recorded in the SEER\textsuperscript{3} Stat Database version  7.0.9 (25). We analyzed incidence rates for males and females separately to account for gender-related differences in the natural history. We limited our analysis to adult cases ages 25–84 years given the distinct etiologies of infant and childhood leukemias (26–30) and the small numbers of cases ages 15–24. We extracted single-year rates for all races combined (primary analyses) and for persons of non-Hispanic white race/ethnicity (sensitivity analyses). To obtain stable estimates using age period cohort analyses (see below), we grouped the single-year data into twenty 3-year age groups (25–27, . . . , 82–84) and six 3-year calendar periods (1992–1994, . . . , 2007–2009) spanning 25 partially overlapping 6-year birth cohorts (1910, 1913, . . . , 1982; referred to by mid-year of birth).

### Statistical analysis

The age period cohort model provides a flexible framework for characterizing cancer trends and patterns according to age at diagnosis, year of diagnosis (period), and year of birth (cohort; refs. 31–33). Importantly, period effects in the age period cohort model adjust for the influences of new treatments or diagnostic routines that can lead to changes in a specific diagnosis among all age groups at a certain point in time. Using age period cohort models, we examined the net drift, which estimates the average annual percentage change in incidence per calendar year or per year of birth (32, 34), the longitudinal age-at-onset curve, which summarizes the age-associated natural history for birth cohorts (14, 35) and the cohort rate ratio curve (CRR; ref. 20), a recently introduced measure that describes the age-specific incidence rates in each cohort relative to an arbitrary reference cohort, adjusted for any calendar period effects influencing all age groups simultaneously. We used the central 1946 cohort as the reference. The log linear component of the CRR equals the net drift, and the nonlinear component equals the Holford cohort deviations (15). We considered 2 likelihood ratio tests of the CRR values, a log linear trend test and a test of nonlinear birth cohort effects, equivalent to the Holford tests (15) of drift and cohort deviations, respectively. In exploratory analyses, we used the Tarone–Chu method (36) to examine local changes in incidence trends around the 1946 reference cohort. For all of our analyses, $P < 0.05$ was considered statistically significant and CI denoted confidence interval.

### Results

Our study included a total of 91,448 leukemias (4 major types diagnosed in males and females ages 25 through 84 years) and $7.01 \times 10^6$ person-years of follow-up (Table 1). A majority of cases, 58%, were diagnosed in men. The overall age-standardized rate was 17.8 and 10.5 per 100,000 person-years in men and women, respectively. Age period cohort models were successfully fitted. There was no significant over- or underdispersion, residual plots did not reveal any systematic lack of fit, and fitted rates closely tracked observed rates.

Net drifts for the 4 major leukemia types were significantly heterogeneous. CML decreased significantly over time among men and women, by around 1% per year (Table 1). CLLs also declined significantly by around 1% per year among men and 0.6% per year among women. In contrast, ALL rates increased significantly by around 2.5% per year among women and nonsignificantly by 0.7% per

### Table 1. Male and female leukemia cases in the SEER 13 and 18 Registries Databases, 1992–2009

<table>
<thead>
<tr>
<th>Sex</th>
<th>Leukemia type\textsuperscript{a}</th>
<th>No. of cases\textsuperscript{b}</th>
<th>Age-standardized rate\textsuperscript{c} per 100,000 person-years (SD)</th>
<th>Net drift, % per y\textsuperscript{d} (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>AML</td>
<td>17,394</td>
<td>5.85 (0.04)</td>
<td>+0.04 (−0.36 to +0.43)</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>3,185</td>
<td>0.96 (0.02)</td>
<td>+0.65 (−0.16 to +1.48)</td>
</tr>
<tr>
<td></td>
<td>CML</td>
<td>8,922</td>
<td>2.93 (0.03)</td>
<td>−1.03 (−1.52 to −0.52)</td>
</tr>
<tr>
<td></td>
<td>CLL</td>
<td>23,819</td>
<td>10.05 (0.07)</td>
<td>−0.98 (−1.35 to −0.61)</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>53,320</td>
<td>17.8 (0.09)</td>
<td>−0.36 (−0.59 to −0.14)</td>
</tr>
<tr>
<td>Female</td>
<td>AML</td>
<td>14,296</td>
<td>3.95 (0.03)</td>
<td>+0.14 (−0.29 to +0.58)</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>2,627</td>
<td>0.73 (0.01)</td>
<td>+2.54 (−1.53 to +3.56)</td>
</tr>
<tr>
<td></td>
<td>CML</td>
<td>6,115</td>
<td>1.69 (0.02)</td>
<td>−0.95 (−1.59 to −0.31)</td>
</tr>
<tr>
<td></td>
<td>CLL</td>
<td>15,090</td>
<td>5.13 (0.04)</td>
<td>−0.55 (−1.05 to −0.05)</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>38,128</td>
<td>10.5 (0.06)</td>
<td>+0.05 (−0.22 to +0.32)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Leukemia types based on ICD-O-3 codes. AML: 9840, 9861, 9866, 9867, 9871–9874, 9895–9897, 9910, 9920; ALL: 9826, 9835–9837; CML: 9863, 9875, 9876, 9945, 9946; and CLL: 9823.

\textsuperscript{b}Leukemia as first primary cancer: Patients diagnosed at ages 25–84 years during 1992–2009 in SEER catchment areas with $3.42 \times 10^6$ and $3.59 \times 10^6$ person-years of follow-up among males and females, respectively.

\textsuperscript{c}2000 U.S. standard population (Census P25-1130). Age-standardized rate and its SD estimated using weighted least-squares.

\textsuperscript{d}Calculated from age period cohort model.
year among men. AML trends were not significant in either sex.

The longitudinal age incidence curves for the 4 major types (Fig. 1) were broadly consistent with prior studies examining cross-sectional age incidence curves (1, 4); however, the longitudinal curves shown here are adjusted for period and cohort effects. On the logarithmic scale shown, age incidence curves for the chronic leukemias CML (Fig. 1A) and CLL (Fig. 1B) appear essentially parallel in males and females, with males consistently higher except perhaps for the male excess of CML, which appears slightly less prominent around the age of 50 years. In contrast, a pronounced male excess of AML emerges after the age of 50 years (Fig. 1C) and a pronounced male excess of ALL diminishes after the age of 65 years (Fig. 1D). We used the Wald test for parallelism (35) to compare the shape of the age incidence curve in males versus females. On the logarithmic scale shown, age incidence curves for the chronic leukemias CML (Fig. 1A) and CLL (Fig. 1B) appear essentially parallel in males and females, with males consistently higher except perhaps for the male excess of CML, which appears slightly less prominent around the age of 50 years. In contrast, a pronounced male excess of AML emerges after the age of 50 years (Fig. 1C) and a pronounced male excess of ALL diminishes after the age of 65 years (Fig. 1D). We used the Wald test for parallelism (35) to compare the shape of the age incidence curve in males versus females. $P$ values were 0.19 and 0.40, respectively, for CML and CLL, indicating little difference in shape for each type in males versus females, versus $6 \times 10^{-4}$ and 0.03, respectively, for AML and ALL, indicating significant age–sex interaction.

The longitudinal age incidence curves in Fig. 1 are set to the level that best matches absolute rates in the reference cohort of persons born circa 1946. Under the age period cohort model, the adjusted age incidence curve for any other cohort is proportionally higher or lower according to the corresponding value of the CRR curve (Fig. 2). CRR values varied significantly for each major leukemia type except AMLs in females (Table 2), on account of significant log linear trend, significant nonlinear birth cohort effects, or both. Specifically, among women, log linear trend was significant for ALLs, CMLs, and CLLs. Among men, the log linear trend was also significant for ALLs, CMLs, and CLLs, and in addition, the nonlinear birth cohort effects were significant for AMLs and CLLs and trending to significance for ALLs.

CRR values for AMLs in males (Fig. 2A) increased then decreased among successive cohorts born before 1946, peaking around 1.25-fold higher for cohorts born from 1916 through 1931. Thereafter, the log linear trend was stable or consistently increasing. In exploratory analysis, a nominally significant local acceleration was detected for men ($P = 0.01$) when the 7 adjacent birth cohorts on either
side of the reference cohort (1949–1967 vs. 1925–1943) were contrasted using the Tarone–Chu method. AML birth cohort patterns appeared qualitatively similar in women compared with men (Fig. 2B vs. (A; $P = 0.29$ for differences between males and females). However, among women, neither the log linear nor the nonlinear birth cohort effects were statistically significant (Table 2).

For AMLs, the distribution of the underlying 13 ICD-O-3 codes changed markedly over time. The frequency of the most common code (9861) decreased abruptly around the middle of our study period, from 78% of cases in 1992 to 54% of cases in 2009, and a number of less common codes (including 9871–9874) increased substantially. Therefore, we did not attempt to fit age period cohort models for specific codes.

CRR values for CLL in males declined significantly (Fig. 2C); furthermore, the rate of decline was significantly faster in cohorts born after 1946 than before (Table 2 and Fig. 2C). Compared with the 1946 cohort, the CLL rate in the 1964 cohort was 0.63-fold lower (95% CI, 0.50- to 0.79-fold lower). CRR values for CLL in females also declined significantly (Fig. 2D); however, there was no acceleration of the decline in younger cohorts.

For ALL in males (Fig. 2E), nonlinear birth cohort effects were trending to significance ($P = 0.08$; Table 2), and there was a nominally significant local acceleration around the differences between males and females). However, among women, neither the log linear nor the nonlinear birth cohort effects were statistically significant (Table 2).

For AMLs, the distribution of the underlying 13 ICD-O-3 codes changed markedly over time. The frequency of the most common code (9861) decreased abruptly around the middle of our study period, from 78% of cases in 1992 to 54% of cases in 2009, and a number of less common codes (including 9871–9874) increased substantially. Therefore, we did not attempt to fit age period cohort models for specific codes.

CRR values for CLL in males declined significantly (Fig. 2C); furthermore, the rate of decline was significantly faster in cohorts born after 1946 than before (Table 2 and Fig. 2C). Compared with the 1946 cohort, the CLL rate in the 1964 cohort was 0.63-fold lower (95% CI, 0.50- to 0.79-fold lower). CRR values for CLL in females also declined significantly (Fig. 2D); however, there was no acceleration of the decline in younger cohorts.

For ALL in males (Fig. 2E), nonlinear birth cohort effects were trending to significance ($P = 0.08$; Table 2), and there was a nominally significant local acceleration around the

---

**Table 2. Statistical significance of CRRs**

<table>
<thead>
<tr>
<th>Sex</th>
<th>Leukemia type</th>
<th>Log linear trend per birth year, $P$</th>
<th>Nonlinear birth cohort effects, $P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All males</td>
<td>AML</td>
<td>0.66</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>CML</td>
<td>$7 \times 10^{-8}$</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>CLL</td>
<td>0.01</td>
<td>$8 \times 10^{-6}$</td>
</tr>
<tr>
<td>All females</td>
<td>AML</td>
<td>0.24</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>$2 \times 10^{-7}$</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>CML</td>
<td>$4 \times 10^{-4}$</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>CLL</td>
<td>0.05</td>
<td>0.30</td>
</tr>
</tbody>
</table>

*See Methods for details. $P \leq 0.05$ are noted in bold.*

---
1946 reference cohort ($P = 0.03$), such that CRR values were consistently increasing in younger cohorts born after 1946. For example, compared with the 1946 reference cohort, the ALL rate in the 1964 cohort was 1.37-fold higher (95% CI, 1.01- to 1.87-fold higher). Among females, the CRR values for ALL (Fig. 2F) also increased consistently and significantly by cohort, but there was no evidence for any changes around the 1946 reference cohort. As for AMLs, the frequency of specific ALL codes also changed abruptly over time, therefore separate models were not considered.

Birth cohort patterns for CMLs appear very similar in men and women (Figs. 2G and H, respectively). In both sexes, rates were highest in the oldest cohorts born circa 1910. The rates clearly declined by around 1% per birth year through the cohorts born circa 1946. The CRR values appear to have remained stable from that cohort on. However, the apparent leveling off of the CRR curves does not result in a statistically significant test of nonlinear birth cohort effects in males or females (Table 2). Interestingly, in exploratory analysis, we detected a nominally significant local deceleration for men ($P = 0.004$ for 1949–1967 vs. 1925–1943). The corresponding difference in slopes was similar for women, but the slope contrast was not statistically significant ($P = 0.11$). As a sensitivity analysis, we excluded the 22% of CML cases with ICD-O-3 code 9945 corresponding to the clinically and cytogenetically distinct entity of chronic myelomonocytic leukemia (CMML) and obtained very similar results.

As a sensitivity analysis, we estimated CRR curves specifically for non-Hispanic whites. The values were very similar to those shown for all races combined. However, for a number of the curves, the levels of statistical significance were reduced, as expected given the diminished numbers of cases.

**Discussion**

The 4 major leukemia types are established endpoints for descriptive (37) and analytic (38–41) epidemiologic studies; however, it is increasingly clear from molecular studies that these categories do not represent homogeneous groups of similar diseases but instead constitute heterogeneous groups of related diseases. Within the major types, the etiologies of the component diseases remain largely unclear and are very difficult to study because of the small numbers. For this reason, we and others have focused on the broad-based major types, recognizing that the mixing of subgroups might weaken or obscure any underlying association signals.

Therefore, it is quite striking to report that leukemia incidence varied significantly between birth cohorts for each major leukemia type in men and women except female AMLs; changes on the order of 1% per birth year or 20% per generation were observed. Our results are consistent with the hypothesis that many leukemia risks among adults are substantially affected by known, suspected, and perhaps unrecognized environmental and lifestyle exposures that change in prevalence from one birth cohort to the next (32, 34). Compared with the patterns found for some other malignancies, notably testicular (42–44) and breast (45, 46) cancers, associations of leukemia risks and birth cohort are comparatively subtle. Of course, the comparatively weak signal strengths might reflect the broad mixture of diseases under study.

A key strength of our study is that we analyzed the sexes separately. Indeed, our longitudinal age incidence curves revealed significant age–sex interactions for AMLs and ALLs after accounting for period and cohort effects. This finding has repeatedly been noted in prior cross-sectional studies in which cohort artifacts cannot entirely be ruled out. Therefore, our analysis strongly supports the hypothesis that acute leukemia risks are modified by hormonal changes over the lifespan in males or females. Important-ly, our separate estimates of birth cohort effects for males and females are adjusted for these differences in the natural histories between the sexes.

The strongest specific (i.e., nonlinear) birth cohort signatures were observed for CLls and AMLs in men. The significant birth cohort signature for AMLs in males is intriguing, given there are a number of established AML risk factors (including benzene and formaldehyde, and tobacco (7, 47, 48). Therefore, the significant increase and decrease of AML incidence in cohorts of men born before 1946 might reflect, at least in part, increases and decreases in these well-established carcinogens.

There is increasing evidence that obesity should be recognized as an important established risk factor for leukemia, particularly AMLs, based on diverse single-population studies (38, 40, 49–52) and in meta-analyses (53, 54). In the United States, obesity has become prevalent in all age groups (55). In exploratory analysis, we also detected a significant acceleration in AML risk (consistent with either a genuine increase in risk or a slowing of a decreasing risk) beginning with the baby boomer cohorts. Importantly, these cohorts of men are not yet elderly. Therefore, if the significant acceleration in AML risk in successive cohorts born since 1946 is, in part, a consequence of obesity, our findings suggest that obesity, and any other putative risk factor increasing in these cohorts, must modulate the risks beginning in early adulthood or middle age.

Nonlinear birth cohort effects for CLls were highly significant in men but not in women, even though the shape of the age incidence curve for CLL was remarkably similar in women compared with men. This discrepancy might reflect smaller numbers of CLL cases in women compared with men. Alternatively, if CLL is increasingly diagnosed and managed outside of the hospital setting especially in younger men, the cohort patterns reported here might reflect decreasing completeness of hospital-based CLls reporting to SEER rather than genuine declines in incidence (56). Existing (small) adjustments for reporting delay apply to all ages and therefore modify period effects but not cohort effects in age period cohort models. Development of reporting delay adjustments that
vary by age, sex, and calendar year might help determine whether cohort patterns for CLLs reflect genuine changes in risk or reporting artifacts.

CML is characterized by a signature lesion, the BCR-ABL translocation (57). Because a necessary cause of CML is so specific and the disease is so rare in all populations, one might hypothesize that the incidence of CML might be more driven by stochastic events than specific environmental risk factors. We and others (1, 3) have reported significant overall declines in CML incidence; in this study, we also observed apparently stable CML rates in cohorts born since 1946. This new observation is consistent with the hypothesis that CML is less environmentally sensitive than AML which is increasing in the same cohorts but needs to be replicated in other populations.

ALL is a highly heterogeneous disease and little is known about its causes (58). We observed significant gender differences in natural history (Fig. 1), time trends (Table 1), and birth cohort effects (Fig. 2). ALL does appear to be increasing in both younger men and women.

A major limitation of our study is the primary analysis included persons of all race/ethnic groups, because even using SEER 18, there were relatively limited numbers of cases to estimate birth cohort effects with sufficient precision within distinct racial/ethnic groups. Therefore, our primary results could potentially be biased by changes in population structure. However, we did have sufficient numbers of cases to restrict our analyses to the largest subtypes defined by ICD-O-3 codes over time that might reflect changes in risk or reporting artifacts. Therefore, we did not pursue separate analyses for the various subtypes defined by ICD-O-3 codes.

In summary, our results are consistent with the hypothesis that risks of adult leukemia are significantly modulated by environmental and lifestyle exposures and therefore may be modifiable. However, we did not identify notably sharp or specific changes in risk associated with birth cohort, except for increases in AMLs and decreases in CLLs in younger men born since 1946. In light of the limitations noted above, our results also suggest that identification of stronger signals may require that future studies split each major type into a few consistently defined and biologically related families of subtypes (59–60). This strategy has proven very informative in studies using estrogen receptor status of breast tumors (61–62). For population-based studies of the leukemias, such subtypes might reflect advances in cytogenetics (13, 63).

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: P.S. Rosenberg, W.F. Anderson
Development of methodology: P.S. Rosenberg
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): W.F. Anderson
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): P.S. Rosenberg, K.L. Wilson, W.F. Anderson
Writing, review, and/or revision of the manuscript: P.S. Rosenberg, K.L. Wilson, W.F. Anderson
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P.S. Rosenberg, K.L. Wilson, W.F. Anderson
Study supervision: P.S. Rosenberg

Acknowledgments
The authors thank Ju-Hyun Park for his thorough statistical review and the reviewers for their helpful comments that greatly improved the content of the manuscript.

Grant Support
This research was supported entirely by the Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 1, 2012; revised September 11, 2012; accepted September 28, 2012; published OnlineFirst October 12, 2012.

References