Population-Based Exploration of Academic Achievement Outcomes in Pediatric Acute Lymphoblastic Leukemia Survivors

Lyndsay A. Harshman,1 MD, Sheila Barron,2 PhD, Anna M. Button,3 MS, Brian J. Smith,3 PhD, Brian K. Link,4 MD, Charles F. Lynch,5 MD, and Natalie L. Denburg,6 PhD

1Department of Pediatrics, University of Iowa Carver College of Medicine, 2Department of Psychological and Quantitative Foundations, University of Iowa College of Education, 3Holden Comprehensive Cancer Center, 4Department of Internal Medicine, University of Iowa Carver College of Medicine, 5Department of Epidemiology, University of Iowa College of Public Health, and 6Department of Neurology, University of Iowa Carver College of Medicine

All correspondence concerning this article should be addressed to Natalie L. Denburg, PhD, Department of Neurology, #2007 RCP, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, IA, 52242-1053, USA. E-mail: natalie-denburg@uiowa.edu

Received May 24, 2011; revisions received December 13, 2011; accepted December 15, 2011

Objective Examine academic achievement among pediatric acute lymphoblastic leukemia survivors diagnosed during the years 1993–2008. Method A deterministic linkage of the Iowa Cancer Registry and Iowa Testing Programs databases was performed and yielded 147 survivors. Achievement data, in the form of Iowa Percentile Rank scores, were obtained and analyzed by grade and content domain. Results Children diagnosed before age 5 evidenced more underachievement than those diagnosed later (p = .05). Underachievement was noted in mathematics in grades 8 and 11 (p’s < .05), in addition to a longitudinal decrease in scores from grades 4 through 11 (p = .01). No differences were found in academic achievement between males and females. Conclusions Utilization of a population-based approach with a nationally recognized, standardized instrument indicates that academic underachievement is subtle yet exists, most notably in mathematics.

Key words academic achievement; cancer survivorship; cognition; leukemia; population-based.

Introduction

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy and has a peak incidence between 2 and 5 years of age (Linabery & Ross, 2008; Reis et al., 2008). As treatment for ALL has improved over the past 30 years, healthcare teams can now provide a nearly 80% cure rate to those affected (Pui, Relling, & Downing, 2004). Despite treatment success, survivors are at risk for the development of long-term complications following therapy, such as psychosocial dysfunction (Mody et al., 2008; Schultz et al., 2007), neurocognitive deficits (Mahone, Prahme, Ruble, Mostofsky, & Schwartz, 2007; Moleski, 2000), and neuroanatomic abnormalities (Carey et al., 2008; Chan, Roebuck, & Yuen, 2001; Iuvone, Mariotti, & Colosimo, 2002). It has also been noted that young age at diagnosis and female gender increase risk for the aforementioned outcomes, specifically neurocognitive deficit (Buizer, de Sonneville, van den Heuvel-Eibrink, & Veerman, 2005; Mulhern & Butler, 2004).

Since the late 1970s, a large literature has focused on the neurocognitive sequelae of cancer treatment in pediatric ALL survivors. Although the specific mechanisms behind central nervous system (CNS) impairment following therapy for ALL have not been definitively elucidated, several possibilities have been raised including white matter injury and altered levels of neurotransmitters...
followed CNS chemotherapy (see Saykin, Ahles, & McDonald, 2003, for a review). To illustrate, methotrexate is a commonly used CNS chemotherapy in ALL for prevention of leukemic metastases to the CNS. Use of methotrexate has been associated with white matter injury and subsequent leukoencephalopathy in children, along with decreased hippocampal concentrations of neurotransmitters important for cognition in rodent models of malignancy (Iyer, Chaturvedi, Pruthi, Khanna, & Ishak, 2011; Yang et al., 2011).

Resultant neurocognitive deficits have a potential to create a profound impact on survivors’ future academic and occupational successes. Concern for academic underachievement in pediatric ALL survivors has been bolstered by two separate independent analyses of data from the Childhood Cancer Survivor Study revealing that survivors of ALL are less likely to graduate from high school compared to sibling controls (Mitby et al., 2003; Mody et al., 2008). Additionally, a meta-analysis by Campbell et al. (2007) demonstrated that ALL survivors remain susceptible to significant declines in multiple areas of neurocognitive functioning, including global academic achievement, following contemporary trends in ALL treatment emphasizing chemotherapy-based CNS protocols over cranial radiation therapy to prevent leukemic metastases to the brain and spinal cord. Academic achievement in long-term ALL survivors has been within expectations for the domains of reading and spelling; however, mathematics achievement tends to remain inferior to reading and spelling scores (Janzen & Spiegler, 2008). Difficulties with mathematics have been reported as one of the most frequently observed deficits for pediatric ALL survivors and families, and evidence exists to link mathematics underachievement in ALL survivors to deficits in neurocognition, such as memory function (Kaemingk, Carey, Moore, Herzer, & Hutter, 2004).

The current study sought to examine the effects of ALL on academic achievement throughout childhood and adolescence (i.e., school grades K-12) in a population-based cohort via a linkage involving the Iowa Cancer Registry (ICR) and the Iowa Testing Programs (ITP), two longstanding, statewide databases unique to the state of Iowa. Here, academic achievement outcomes were assessed with a comprehensive standardized instrument administered at multiple time points. We hypothesized that ALL survivors would demonstrate academic underachievement compared to Iowa normative data in all domains analyzed. Furthermore, considering the known pediatric oncology literature, we hypothesized that underachievement would be most prominent in the domain of mathematics across all years surveyed.

**Methods**

**Database Sources**

**Iowa Cancer Registry**

Since 1973, the ICR has participated in the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) program (About SEER, 2011; Data Collection, 2010). The ICR collects information on all newly diagnosed primary cancers among Iowans and follows these patients to death while capturing information on date of diagnosis, demographic characteristics, clinicopathologic data (i.e., cancer histology, stage, and grade), and cause/date of death. The ICR collects treatment data for first-course therapy information, outlining therapy that has either been provided or to be provided; however, no specific agent or dose data are collected, rather, the information is general. Prior to study initiation, the authors chose to exclude analysis of treatment data since, given its generality, it would be non-informative in relation to measures of academic achievement due to lack of specificity regarding chemotherapeutic agent, dose, and length of treatment used.

The ICR does not collect socioeconomic data (i.e., education, income) on registrants at the individual level. For adults, some general socioeconomic information is abstracted from medical records regarding occupational information. Nonetheless, Iowa is a relatively homogenous state with a stable population with 91.3% of state citizens being white, 93.6% of the state speaking English at home, and 83.4% of the population living in the same home for one year or more during the years 2005-2009 (U.S. Census Bureau, 2011).

**Iowa Testing Programs**

The ITP, a division of the College of Education at the University of Iowa, develops the Iowa Tests of Basic Skills (ITBS) and Iowa Tests of Educational Development (ITED). The ITP also manages the Iowa statewide testing program and maintains a historical database of ITBS and ITED records for Iowa students dating back to the 1930s. The tests began as an “academic contest” in Iowa and developed into a statewide voluntary testing program for kindergarten through 12th grade with 90–95% of schools in Iowa participating on a strictly voluntary basis (Peterson, 1983) prior to the institution of mandatory statewide testing in 2001 (No Child Left Behind, 2001).

These tests, which are well known, well reviewed, and also widely used outside the state of Iowa (Linn, 1989), serve three main purposes for educators: (a) identification of students’ areas of relative strength and weakness in specific subject areas, (b) monitoring yearly growth in basic skills, and (c) describing each student’s developmental
level within a test area ("Interpreting Test Scores," 2008). The historical ITP database at the University of Iowa provides a valuable tool for addressing a wide range of questions about educational achievement (Linn, 1989), as evidenced by two Congressional Budget Office (1986, 1987) studies that relied heavily on ITP data to draw conclusions about trends in education.

Although the ITBS and ITED are separate test batteries, they are developed and standardized simultaneously and designed to measure academic achievement from kindergarten through high school using a common developmental score scale to provide validity across grades. Extensive psychometric data exist from the past 60 years of ITBS and ITED test construction and evaluation have demonstrated outstanding psychometric properties. For example, the consistency of composite scores over a time interval comparable to the one utilized in the current study have demonstrated reliability coefficients of .97–.98 (Hoover et al., 2003). Additionally, elementary and secondary school scores on these tests carry predictive weight for both ACT composite score and collegiate grade point average (Loyd, Forsyth, & Hoover, 1980; Qualls & Ansley, 1995). Both the ITBS and ITED assess content domains such as Reading, Language, and Mathematics. Additionally, a Core Total and Complete Composite score may be obtained for each child. The Core Total score is derived from the Reading, Language and Mathematics test scores. The Complete Composite score is derived from the Core Total plus additional achievement scores in Social Studies, Science, and Sources of Information domains (Hoover et al., 2003). Students receive a raw score, percent correct, Iowa percentile rank, national percentile rank, grade equivalent, and developmental standard score for each test domain administered.

Procedure
A deterministic linkage between the ICR and ITP databases was undertaken. See Figure 1 for a summary of the linkage procedure. The ICR was queried for Iowa children aged birth through 18 years diagnosed with ALL as their first primary cancer between the years 1993 and 2008 and listed as residents of the state at the time of diagnosis. Cases were excluded if the ALL diagnosis was nonprimary malignancy, was diagnosed outside of the years 1993–2008, and/or if death occurred during the period for which data were available. The epoch of 1993–2008 was chosen for analysis to grossly reflect current treatment trends in ALL central nervous system prophylaxis (i.e., use of chemotherapy-based protocols as opposed to cranial radiation therapy) (Brown et al., 1992; Mulhern, 1994).

Selected demographic variables from the ICR database were provided to the ITPI for those participants having met the aforementioned inclusion and exclusion criteria. These variables included full name, date of birth, age, gender, domestic address(es), and date of cancer diagnosis. The demographic variables were used to perform a query of the ITPB database to search for matching students. To be considered a match and included in the sample, demographic data had to correspond to achievement data in one grade or more grades from the three surveyed (i.e., grades 4, 8, or 11). Inclusion in longitudinal data analyses required that a participant have data present in two or more sequential years surveyed (e.g., grades 4 and 8, grades 8 and 11, or grades 4 and 11). Data were found within the ITPB database for 147 of the 162 students provided from the initial ICR query, yielding a 90.7% match rate for the sample. Within the sample, 85 (57.8%) participants had data present for at least 2 years of academic data.

For each ICR–ITP match, the ITP provided a dummy-coded identification number, which eliminated certain demographic data (i.e., name, domestic address, and date of birth). The ITPI then provided the research team with the following dummy-coded variables for each participant: age at testing (in years and months), age at ALL diagnosis (year and month), grade in school, date(s) of test administration, exam scores, and test level. Before data were transferred to the research team, the linked files were secondarily assessed for accuracy by ITP staff to ensure data were complete and without error. All electronic files used in the transfer of information were destroyed upon satisfactory completion of the study.

The University of Iowa’s Institutional Review Board approved this study. Participant informed consent was waived in obtaining these data secondary to the archival nature of the linkage, wherein the investigators did not receive identifying patient information to perform the data procurement and analysis.

Data Analysis
This study referenced Iowa pediatric ALL survivor IPR data to state normative IPR data on the ITBS and ITED to maintain a relatively uniform curricular standard for academic achievement. IPR data hold a consistent meaning across grades, tests, and time. In any random sample of students, the expected value for the median IPR is the 50th percentile, with scores less than the 50th percentile indicating underachievement. To examine if ALL survivors displayed...
underachievement compared to peers, the IPR for the sample was compared to the benchmark of the 50th percentile. In the early years of mandated testing, schools were required to test at grades 4, 8, and 11 (No Child Left Behind, 2001). For this reason, and because there was some preference for testing these grades even when testing was voluntary, data from these grades were analyzed in this study. In the current study, domains tested by both the ITBS and ITED were emphasized, specifically Reading, Language, Mathematics (see Table I for a description of these domains).

Table I. Test Domains

<table>
<thead>
<tr>
<th>Test domain</th>
<th>ITBS</th>
<th>ITED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading (RT)</td>
<td>Concerned with a student’s ability to derive meaning from a text</td>
<td>Utilizes upper level comprehension skills, including analytic and evaluative understanding.</td>
</tr>
<tr>
<td>Language (LT)</td>
<td>Emphasis on expressing ideas in written form</td>
<td>Require students to revise written materials and make choices concerning organization, usage, and mechanics.</td>
</tr>
<tr>
<td>Mathematics (MT)</td>
<td>Focus on mathematics computation, estimation, and data interpretation</td>
<td>Primary purpose is to measure student’s ability to solve quantitative problems; however, problem-solving skills are also a focus.</td>
</tr>
</tbody>
</table>

A summary of the major domains tested as part of Iowa Tests of Basic Skills (grades kindergarten through 8) and Iowa Tests of Educational Development (grades 9 through 12). Each domain below was analyzed as part of the current study.


Achievement data were required from at least one school year for inclusion in the sample and at least two sequential academic years (e.g., grades 4 and 8 or grades 8 and 11) of data were necessary for inclusion in pair-wise (longitudinal) analyses. Linear mixed effects regression was used to model test scores as a function of domain, gender, grade, and age at diagnosis. Iowa percentile rank scores for participants were utilized as the dependent variable in the analysis. Main effects for content domain, gender, grade, and age at diagnosis were included as independent variables. The Akaike information criterion (AIC) was used to compare the fits of models with all possible combinations of two-way interaction added to the main effects. AIC-based comparisons yielded a final model with interactions between domain and gender, as well as domain and grade. Inclusion of the interactions allowed the effects of gender and grade to differ across test domains. An autoregressive correlation structure was specified in the mixed effects models to account for within-subject
correlation among repeated test scores collected over time, and unstructured correlation to account for correlation among the different domains (i.e., Reading, Mathematics, Language) within a complete testing battery. Gender and grade-specific mean estimates and 95% confidence intervals for the scores were obtained from the mixed model for each test group. Pair-wise comparisons were performed to test for specific differences in mean scores. All statistical tests were two-sided and carried out at the 5% level of significance using the SAS 9.2 software package (Cary, NC, USA).

Results
Out of the 162 participants provided by the ICR to the ITP, a total of 147 participants met the aforementioned inclusion and exclusion criteria and were matched to academic achievement data for at least one of the three grades surveyed across the 1993–2008 epoch. Eighty-five of the 147 participants included in the final sample (57.8%) had data for at least two of the three grades surveyed. Of the 147 participants, 41% were female. On average, participants were aged 7.2 years (SD = 3.8 years) at the time of diagnosis. Descriptive statistics are provided in Table II to summarize participant characteristics and test scores. Mean estimates for achievement scores were adjusted for age at diagnosis and correspond to a diagnosis age of 5 years.

Virtually all mean estimate scores for achievement were below the 50th percentile, although not all reached statistical significance at the 95% confidence interval (Table II). Achievement was significantly less than 50th percentile for the following grades and domains: 4th grade Language; 4th grade Reading; 8th and 11th grade Mathematics; 8th and 11th grade Core Composite (see Figure 2 for a distribution of Mathematics and Core Composite scores). As a whole, males evidenced significant underachievement on tests of Language; however, Language achievement was not significantly different between genders.

Because data were available from grades 4, 8, and 11, it was possible to analyze performance over time (Table III). There was a significant decline in Mathematics scores between grades 4 and 8 (p = .02) as well as between grades 4 and 11 (p = .01), with estimated mean achievement scores declining from the 50th percentile in grade 4 to the 39th percentile in grade 11. No significant difference in achievement was noted across time in either Reading or Language.

Overall mean achievement scores were estimated to decline by 1.8 percentile ranks for every one-year decrease in age at diagnosis (p = .05) in children diagnosed younger than age 5 years. There was no interaction between age at diagnosis and academic achievement performance between genders (p = .82), nor was there a relationship between academic achievement and increasing time since diagnosis (p = .54).

Discussion
Much research investigating academic underachievement in pediatric ALL survivors has been hindered by small sample sizes, variability in testing instruments, and lack of long-term follow-up with patients (Anderson & Kunin-Batson, 2009; Peterson et al., 2008). Further, no population-based study has been conducted in the United States to examine academic achievement outcomes in pediatric leukemia survivors. Here, we utilized
a well-validated academic achievement instrument and procured data on a large population-based cohort of ALL survivors. Overall, our ALL survivor sample fits well with known epidemiological data demonstrating a trend for a peak in ALL diagnoses at approximately 5 years of age (7.2 ± 3.8 years of age in our sample), with a slight male predilection for disease development (59% male in our sample) (Linabery, 2008; Reis, 2008).

As predicted, we found evidence for longitudinal decline in academic achievement in Mathematics between grades 4 and 11, with scores worsening progressively across grades. This finding supports past research demonstrating survivors’ scores in mathematics have the potential to worsen with increasing time from initiation of treatment on chemotherapy-only protocols (Brown et al., 1992), and additionally bolsters more recent research demonstrating persistence in mathematics deficits for ALL survivors having completed chemotherapy-only protocols (Buizer et al., 2005; Kaemingk et al., 2004). Contrary to our hypothesis, we did not find widespread academic dysfunction in the sample assessed. Rather, survivors performed generally around 50th percentile in Language and Reading, as well as for overall achievement (i.e., Core Composite); however, there was wide variability in performance across the sample for overall achievement as is displayed in Figure 2. For example, approximately 20% of the Core Composite scores in grade 11 were below the 10th percentile, whereas another 20% of the grade 11 Core Composite scores were below the 90th percentile.
scores were above the 80th percentile, and remaining scores varied markedly in between. It is believed that young age at diagnosis is a risk factor for general neurocognitive dysfunction in pediatric ALL survivors (Buizer et al., 2005; Von der Weid et al., 2003); however, less research exists examining the impact of early age at diagnosis and a more specific impact on academic achievement. In the current study, there was an interaction noted between age at diagnosis \((p = .05)\) and poorer academic achievement outcomes. Additionally, Mulhern and Butler (2004) indicate that young age at diagnosis, in combination with female gender, confers additional risk for poor neurocognitive outcome; however, no interaction between young age at diagnosis and gender was observed in our data.

Past studies have noted significant differences in overall neuropsychological performance between male and female pediatric ALL survivors (Brown et al., 1998; Waber, Tarbell, Kahn, Gelber, & Sallan, 1992), with females generally performing worse than males on neuropsychological tasks; conversely, this trend has not been well replicated in studies with academic achievement. More recent research specific to academic achievement has failed to note an association between gender and achievement using parental and child self-report measures to assess for potential academic dysfunction in ALL survivors (Kingma, Rammeloo, van der Does-van den Berg, Rekers-Mombarg, & Postma, 2000). In the current study, there was neither a main effect of gender on achievement nor any statistically significant differences in academic achievement for individual test domains between genders. Although males did perform consistently lower than the 50th percentile on tests of Language skills, scores were not significantly different than females. The ITP has investigated potential gender differences in achievement utilizing normative data, and specific to the ITBS and ITED, the performance of boys tends to be more variable than girls (Hoover et al., 2003). The Iowa Tests “Guide to Research and Development” describes that data from a 2000 normative sample showed gender differences could be found in the domain of Language whereby girls appear to perform more favorably and that score differences between gender increase with grade at the 50th and 10th percentiles. Equalization in achievement by gender tends to occur just above the median in most grades for domains including Vocabulary, Math Problem Solving, Social Studies, and Science (Hoover et al., 2003). Given the variability of achievement differences in a healthy population, it could be suggested that consistent gender differences noted in the pediatric oncology literature (e.g., females performing more poorly than males) among ALL survivors may not be as striking in academic achievement measures as in neuropsychological tasks.

There are limitations to the current study. The authors were not able to contact children/families for whom academic achievement data were procured due to the archival nature of this study and similar use of waived consent, thus limiting our ability to investigate additional individual participant variables that could impact academic performance, such as time out of school. On a related note, there were no socioeconomic data available for participants within either the ICR or ITP. Additionally, we were unable to address the impact of actual treatment on academic achievement, and although the ICR collects diagnostic data for all registrants, there is no specific treatment data in the ICR. It could be recommended that future research examine risk type of leukemia in addition to variables such as age at diagnosis, gender, and/or death as specific predictors for academic achievement outcomes to more fully understand whether academic underachievement is a consequence of disease and/or treatment or rather, a marker of a more guarded prognosis and more severe disease.

The results of this population-based analysis of academic achievement in pediatric survivors of ALL suggest that subtle differences in achievement can be found between survivors and state-referenced, normative data with more notable weaknesses found in the domain of Mathematics. We believe that a strength of the current study is its inherent ecological validity. Participation in the Iowa tests is a routine part of childhood in Iowa and furthermore, the content matter tested is considered standard curriculum for the state. By surveying a large population over fifteen years using a standardized testing instrument, we believe the results obtained provide a representative and generalizable picture of academic achievement for pediatric ALL survivors that parallels much of the known pediatric oncology literature, despite a lack of treatment data. As such, potential academic late effects encountered by survivors should continue to be a consideration when counseling family members regarding prognosis after treatment, as well as when working with educators throughout growth and development.

**Funding**

Preparation of this article was supported by a Doris Duke Clinical Research Fellowship (to L.A.H.) and a National Institute on Aging Career Development Award (K01 AG022033 to N.L.D.).

**Conflicts of interest:** None declared.
References


