Cognitive and Academic Late Effects Among Children Previously Treated for Acute Lymphocytic Leukemia Receiving Chemotherapy as CNS Prophylaxis

Ronald T. Brown, PhD, Avi Madan-Swain, PhD, Gary A. Walco, PhD, Irene Cherrick, MD, Carolyn E. Ievers, PhD, Paola M. Conte, MA, Roger Vega, MD, Beverly Bell, MD, and Stephen J. Lauer, MD

1Medical University of South Carolina, 2Emory University School of Medicine, 3Hackensack University Medical Center, 4University of Medicine and Dentistry, New Jersey Medical School, and 5State University of New York at Syracuse

Objective: Examine cognitive and academic late effects among children and adolescents who had received central nervous system (CNS) prophylactic chemotherapy alone for acute lymphocytic leukemia (ALL); none had received whole brain radiation therapy (RT).

Method: Subjects included 47 children and adolescents from 5 to 22 years of age who were treated on the same protocol and had been off treatment from 2 to 7 years at the time of assessment.

Results: As a group the survivors displayed generally average performance on measures of cognitive and academic abilities, although they differed from normative means on tests of nonverbal skills. Girls performed more poorly than the normative sample on nonverbal tasks, while no differences were found for boys. Age at diagnosis and time off treatment were not significantly associated with cognitive and academic functioning for survivors of this particular chemotherapy-only protocol.

Conclusions: Data were interpreted to support generally modest potential late effects in specific areas for children and adolescents surviving ALL. These findings suggest a need for monitoring nonverbal cognitive skills for childhood survivors of ALL, particularly for girls.

Key words: cognitive; late effects; leukemia; chemotherapy.

There has been significant interest in the cognitive and academic late effects of children who have survived acute lymphocytic leukemia (ALL) (for review, see, Mulhern, 1994). The standard treatment for these children includes prophylaxis of the central nervous system (CNS) to eradicate the leukemia cells in the CNS and thereby increase the probability of disease-free survival. Initially, protocols for CNS prophylaxis consisted of whole brain radiation therapy (RT) with or without intrathecal chemo-
therapy (i.e., administered into the spinal fluid). Due to the documented neuropsychological impairments associated with whole brain RT (for review, see Butler & Copeland, 1993; Mulhern, 1994), the Children's Cancer Group (CCG) and the Pediatric Oncology Group (POG) have limited this modality of treatment to children having CNS disease at diagnosis (Mulhern, 1994) and now employ only intrathecal chemotherapy for CNS prophylaxis across the majority of all risk groups.

Mixed results have been obtained concerning potential late effects of intrathecal chemotherapy for CNS prophylaxis on children's cognitive and academic functioning. In a cross-sectional investigation, Brown et al. (1992) found that patients with ALL who had completed a three-year course of systemic and intrathecal chemotherapy evidenced greater nonverbal neurocognitive impairments and learning disabilities in mathematics than did children who were newly diagnosed and still receiving treatment. Similarly, Brown et al. (1996) noted that patients with leukemia experienced more neurocognitive toxicities one year after completion of therapy than children with other types of cancer who received no CNS prophylaxis. Further, three years postdiagnosis the leukemia group scored more poorly on academic measures of reading, spelling, and mathematics. Likewise, Mulhern et al. (1992) found that children treated for ALL scored significantly lower than controls (a solid tumor group) on measures of visual and auditory memory and mathematical skills and required more special education services than controls.

In contrast to these findings, in prospective longitudinal studies, Copeland and associates (Copeland et al., 1988; Copeland, Moore, Francis, Jaffe, & Culbert, 1996; Dowell, Copeland, & Judd, 1989) found few differences in cognitive and academic functioning between children with solid tumors who received only systemic chemotherapy and children with leukemia who received systemic and intrathecal chemotherapy. Similarly, in a retrospective investigation, Butler, Hill, Steinherz, Meyers, and Finlay (1994) found that CNS prophylactic chemotherapy alone was not significantly associated with neurocognitive toxicities.

Research has identified variables that may contribute to the cognitive and academic late effects of children who have received CNS prophylaxis for leukemia. Specifically, diagnosis at an earlier age may place children at greater risk for later deficits since early childhood is a time of significant CNS maturation (for review, see Butler & Copeland, 1993; Mulhern, 1994). The potential vulnerability of gender for neurocognitive toxic late effects also has been an area of interest. Waber et al. (1990) found that girls who were treated with chemotherapy and whole brain RT (24 cGy) scored lower (over one half of a standard deviation) on the Arithmetic and Digit Span subtests of the Wechsler Intelligence Scale for Children-III (WISC-III) and in arithmetic achievement than their male counterparts. In fact, Waber et al. found an interaction of age at diagnosis and gender, with girls diagnosed with ALL and treated at a younger age evidencing the greatest neuropsychological impairment. It is unclear whether gender differences may have resulted from the combination of prophylactic therapies (i.e., the combination of whole brain RT and intrathecal chemotherapy) or from whole brain RT alone. In a previous investigation of whole brain RT or chemotherapy alone, no gender differences were found for chemotherapy alone (Schlieper, Esseltine, & Tarshis, 1989).

The present investigation examined the cognitive and academic functioning of children with ALL who received CNS prophylaxis without whole brain RT approximately 2 to 7 years posttermination of treatment ($M = 3$ years, $10$ months, $SD = 1$ year, $5$ months). This study offers a unique contribution as it increases our knowledge of the long-term functioning of children who were treated in the same clinical trial and received similar chemotherapy. We hypothesized that children and adolescents who had been successfully treated for ALL would show some impairments on tests of cognitive abilities and academic achievement relative to normed instruments. Based on previous research (e.g., Brown et al., 1992; Mulhern et al., 1992), survivors were expected to evidence greater deficits on nonverbal tasks and to display poorer academic performance in mathematics than the general population. Girls were predicted to perform more poorly relative to test norms on measures of nonverbal cognitive tasks and of academic functioning (particularly in arithmetic). Finally, younger age at diagnosis was expected to be related to greater neurocognitive toxicities, and an age at diagnosis by gender interaction was hypothesized, with younger girls at diagnosis predicted to evidence greater deficits on nonverbal tasks.
Method

Participants

Participants were 47 children and adolescents, 20 (43%) girls and 27 (57%) boys, diagnosed with ALL treated on POG protocol #8602 from 1986 to 1990. The majority of the participants were Caucasian ($n = 42, 89.4$%); two of the participants were African American (4.3%), two were Hispanic (4.3%), while one subject was Native American (2%). The participants ranged in age at testing from 5 years, 0 months to 22 years, 1 month ($M = 11$ years, 0 months, $SD = 3$ years, 1 month). Age at diagnosis ranged from 1 year, 6 months to 15 years, 6 months ($M = 4$ years, 4 months, $SD = 2$ years, 7 months). They received triple intrathecal therapy (TIT) with methotrexate, hydrocortisone, and cytosine arabinoside alone as the principle CNS prophylaxis. No children in this study received whole brain RT. Details of the treatment protocol have been presented elsewhere (Land, 1994). Patients were randomized to treatment arms that included intermediate dose methotrexate (IDM). All patients received continuation therapy with oral 6-mercaptopurine and intramuscular methotrexate with vincristine and prednisone pulses. The CNS preventive therapy was continued for 2 years and consisted of intermittent doses of TIT.

The present study was conducted at three major medical centers affiliated with POG. Only children who were off therapy $\geq 24$ months were included as participants in this investigation. A total of 184 children and adolescents were enrolled on POG #8602 at the three sites. Of all potential patients, 47 (25.5%) participated in the testing protocol. Reasons for nonparticipation in the study were as follows: 24 (13%) relapsed and were removed from the POG protocol, 7 (4%) experienced other medical complications that necessitated their removal, 17 (9%) relocated and their care was transferred to other sites, and 18 (10%) died. In addition, 3 (2%) of the children were ineligible due to chromosomal abnormalities (e.g., Trisomy 21) and preexisting intellectual deficits, 2 (1%) did not complete testing by the study’s deadline, 11 (6%) declined to participate in the testing, and the remaining 35 (19%) potential participants were lost to follow-up for unknown reasons. Finally, a total of 20 (10.9%) children were off treatment $< 24$ months and thus were not included in this particular investigation.

Our sample was drawn from a larger pool of 1,767 children and adolescents on this POG protocol. At diagnosis the mean age of the entire population was 6 years ($SD = 4$ years, 1 month). Fifty-five percent were boys ($n = 968$), and 45% were girls ($n = 799$). The ethnicities of the children in the population were Caucasian ($n = 1,378; 78$%), African American ($n = 164; 9$%), Hispanic ($n = 130; 7$%), and other ($n = 95; 6$%). At diagnosis 48 children were reported to have CNS involvement (3%). The children in the present investigation were slightly younger at diagnosis than the population treated on this POG protocol, but our sample was generally equivalent to the population with respect to gender and race.

All of the patients were participating in some type of school program, ranging from preschool to the second year of college ($M = 4.9, SD = 2.8$). More than one third of the entire sample (36.2%; $n = 17$) was receiving part-time special education assistance (resource placement), particularly in speech/language. An additional 6.9% ($n = 3$) of the sample was primarily in self-contained special education classrooms, and 23.4% ($n = 11$) had repeated a grade. For girls, 15% ($n = 3$) were in self-contained special education classrooms (i.e., full-time placement); no boys were in full-time special education placements. For girls, 40% ($n = 8$) were receiving special education resource assistance; 33.3% ($n = 9$) of boys were receiving resource assistance. While there was a higher frequency of girls in full-time special education placements than boys (Mann-Whitney $U = -2.06, p < .04$), no gender differences were found for number of boys versus girls receiving resource special education assistance (Mann-Whitney $U = .47, p = .64$).

Current parental marital status and occupation were obtained from medical chart reviews and were verified by an interview at a clinic visit. The majority of subjects were from middle-class backgrounds based on socioeconomic status (SES) occupation codes from Hollingshead (1975). For fathers, the mean occupation was characteristic of a small business owner, technician, or semiprofessional ($M = 5.5, SD = 2.0, range, 1–9$); for mothers, the mean occupation was characteristic of skilled manual workers, craftsmen, or smaller business owners ($M = 4.0, SD = 2.8, range 1–9$). The majority of par-
ents were married \((n = 31, 66\%)\), while 12.8\% \((n = 6)\) were divorced; the remaining parents were either single \((n = 5, 10.6\%)\) or their marital status was unknown \((n = 5, 10.6\%)\).

Approval by the respective institutional review boards and written informed consent from parents were obtained before patient entry. Data were gathered either through school records if there was a recent psychological evaluation or by psychologists in the three clinical settings.

**Assessment Battery**

The assessment battery\(^2\) for this investigation included measures related to general cognitive abilities, visual-motor functioning, and academic achievement. The measure of cognitive abilities was the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991), which allows for the assessment of verbal (VIQ) and performance (PIQ) intelligence as well as full-scale intellectual functioning (FSIQ). In addition, three factor scores were examined: Verbal Comprehension (VC), Perceptual Organization (PO), and Freedom from Distractibility (FFD). The Developmental Test of Visual-Motor Integration (VMI; Beery, 1989) was administered to assess visual-perceptual and fine-motor functioning. Finally, the Woodcock-Johnson Psychoeducational Test Battery: Tests of Achievement-Revised (WJ-R; Woodcock & Johnson, 1990) was employed to assess academic achievement in reading and mathematics.

**Results**

The means and standard deviations of the measures on standardized scores for the entire sample and for each gender relative to test norms are presented in Table I. In all areas, survivors attained scores generally within the average range. To test our hypothesis that survivors evidenced deficits on nonverbal and mathematics achievement relative to norms, comparisons were made between the entire sample and test norms using two-tailed one-sample \(t\) tests. While directional hypotheses were made, a non-directional test of significance was chosen in order to be more conservative. In addition, Bonferroni correction was employed for groups of related constructs (i.e., tests of cognitive ability and achievement) to mitigate the likelihood of Type I error due to the multiple comparisons (Keppel, 1982). Standard scores were employed in all of the analyses. Results revealed significant differences on the WISC-III PIQ, \(t(42) = -2.81, p < .008\), the PO factor, \(t(35) = -2.92, p < .006\), and the FFD factor, \(t(34) = -2.84, p < .007\). Significant effects also were revealed on the VMI, \(t(19) = -3.14, p < .005\). We found no differences for any of the academic measures. These findings support our hypothesis that survivors would evidence deficits on nonverbal tasks relative to test norms. Finally, to determine whether performance on nonverbal tasks was poorer than performance on verbal tasks, we performed a series of two-tailed paired-samples \(t\) tests across the various factors of the WISC-III. Results yielded significant effects between the PIQ and VIQ, \(t(40) = -3.32, p < .002\), the VC and PO factors, \(t(32) = -3.97, p < .001\), and the FFD and VC factors, \(t(31) = -2.86, p < .01\). In all cases, scores on the nonverbal tasks were significantly lower than on the verbal tasks. For the WISC-III nonverbal measures, where differences were hypothesized, post-hoc calculations ranged from .93 to .99 (see Table I).

To determine whether girls performed more poorly than their male counterparts on nonverbal tasks and arithmetic achievement, mean scores for boys and girls were compared separately to the test norms using two-tailed one-sample \(t\) tests with Bonferroni correction for each domain (cognitive and academic measures). Covariates were not employed because no differences on any of the demographic characteristics were found for gender. For girls, differences were revealed between our sample and the normative means on the WISC-III PIQ, \(t(18) = -2.70, p < .02\), and the PO factor, \(t(13) = -3.09, p < .009\). A trend was revealed for the WISC-III FFD factor, \(t(13) = -2.23, p < .04\). Differences also were revealed on girls' VMI scores, \(t(7) = -2.87, p < .02\). We found no significant differences for girls on the WISC-III FSIQ, VIQ, or VC factor. Further, no significant differences were obtained on any of the measures when boys were compared with the normative means. These results support our original hypothesis that girls would evidence defi-

\(^2\)Because some the data were gathered from school psychological evaluations and some children were administered either the WPPSI-R, the WAIS-R, or the Stanford-Binet-Revised, WISC-III scores were not available for all subjects. Thus, the following WISC-III scores were available for analyses: 43 FSIQs, VIQs, and PIQ scores, and 35 VC, 36 PO, and 35 FFD factor scores. The following WJ-R scores were available: 35 Broad Reading and 33 Broad Math scores. The VMI data were collected at only one site (Emory) and therefore only 20 test scores were available for the analyses. Finally, one individual was older than 18 at the time of the evaluation and was administered the WAIS-R. For this subject, WAIS-R scores were not included in the analyses, and only achievement scores were examined.
Cognitive and Academic Late Effects

Table I. Means and Standard Deviations of Measures

<table>
<thead>
<tr>
<th>Measures</th>
<th>Entire Sample</th>
<th>Gender</th>
<th>Test Norms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>WISC-III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-scale IQ</td>
<td>43</td>
<td>95.1</td>
<td>16.9</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>43</td>
<td>98.2</td>
<td>16.5</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>43</td>
<td>92.2</td>
<td>18.2</td>
</tr>
<tr>
<td>VC</td>
<td>35</td>
<td>98.8</td>
<td>17.0</td>
</tr>
<tr>
<td>PO</td>
<td>36</td>
<td>90.7</td>
<td>19.2</td>
</tr>
<tr>
<td>FFD</td>
<td>35</td>
<td>92.3</td>
<td>16.1</td>
</tr>
<tr>
<td>VMI</td>
<td>20</td>
<td>89.1</td>
<td>15.6</td>
</tr>
<tr>
<td>WJ-R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad Reading</td>
<td>35</td>
<td>102.2</td>
<td>17.8</td>
</tr>
<tr>
<td>Word Recall</td>
<td>34</td>
<td>102.8</td>
<td>19.8</td>
</tr>
<tr>
<td>Comprehension</td>
<td>35</td>
<td>100.6</td>
<td>15.1</td>
</tr>
<tr>
<td>Broad Math</td>
<td>33</td>
<td>96.8</td>
<td>15.0</td>
</tr>
<tr>
<td>Math Calc.</td>
<td>34</td>
<td>96.2</td>
<td>17.4</td>
</tr>
<tr>
<td>Applied Prob.</td>
<td>34</td>
<td>97.8</td>
<td>17.5</td>
</tr>
</tbody>
</table>

ES = effect size.
*p < .05.
**p < .01.

 deficits on nonverbal tasks relative to test norms. Finally, when the mean scores of girls versus the mean scores of boys were examined, we found a significant difference only for the PO factor of the WISC-III, t(32) = -2.42, p < .02. Again, we computed a series of post-hoc analyses for each of the variables to determine the probability of detecting an effect size of at least the magnitude obtained in these data (see Table I).

To determine whether age at diagnosis, time off therapy, and SES were associated with greater neurocognitive toxicities, we computed Pearson product-moment correlations between these variables and the dependent measures. None was significant with the exception of a trend between time off treatment and the WISC-III PIQ, r(41) = -.29, p < .07. The means of the correlations between time off therapy and age at diagnosis and each of the dependent measures are .15 and .08, respectively. Finally, to examine the interaction of gender and age at diagnosis, a series of 2 (gender) x 2 (age at diagnosis: children younger than 4 years at diagnosis compared with their older peers) MANOVAs were performed on each of the assessment domains (i.e., cognitive abilities and academic achievement). The partitioning of the sample for age approximated a median split of the frequency distribution at diagnosis and is a clinically significant cut-off. A separate 2 x 2 ANOVA was performed for VMI scores. No significant main effects were obtained for gender or age at diagnosis or their interactions. Thus, no support was provided for our hypothesis of an association between younger age at diagnosis and greater deficits on any of the dependent measures or the interaction between age at diagnosis and gender.

Discussion

Our data corroborate findings from other studies that have identified late effects in nonverbal cognitive functioning for childhood survivors of ALL (Brown et al., 1992; Mulhern, Wasserman, Fairclough, & Ochs, 1988). In this investigation, survivors' performance on nonverbal tasks (the PIQ, PO factor, and FFD factor of the WISC-III) was significantly lower than test norms. Additionally, their scores on these nonverbal tasks were significantly lower as a group than their scores on verbal tasks on the WISC-III (i.e., the VIQ, VC factor). Additionally, survivors exhibited mild impairments in fine-motor abilities on the VMI. These fine-motor deficits previously had been attributed to the use of vincristine, which typically affects the peripheral nervous system, not the CNS (Copeland et al., 1988; Dowell et al., 1989). However, alternative explanations for the observed long-term effects of impairments in fine-motor abilities should be an area for future investigation.
When the total sample was examined, the data did not support predictions regarding late effects in the area of academic achievement. All of the academic achievement scores were well within the average range compared to norms and generally were consistent with the children's mental abilities. These findings fail to provide support for prior studies of academic late effects for children receiving CNS prophylactic chemotherapy for ALL. Brown and associates (Brown et al., 1992; Brown et al., 1996) found that children who were three years off therapy displayed academic late effects; these deficits were not evident for children who were only one to two years off therapy. Our investigation included children who were off therapy 2–7 years. Although the variance in the time since completion of treatment may have diminished the probability of detecting significant late effects in the present investigation, this is unlikely as time off treatment was not associated with any of the measures of academic achievement.

Our data offer little support for an association between age at diagnosis and neurotoxicity of prophylactic chemotherapy during the late-effects period. These findings are in accord with other investigations of children with ALL that failed to obtain age-related effects (e.g., Copeland et al., 1985; Longeway et al., 1990; Rubinstein, Varni, & Katz, 1990) yet are inconsistent with studies that found such effects (e.g., Said, Waters, Cousens, & Stevens, 1989; Waber et al., 1990). Mulhern (1994) has suggested that differences in the amount and types of chemotherapy administered may in part explain the discrepancies across studies. For example, some of the investigations have examined the effects of intrathecal methotrexate chemotherapy only (Copeland et al., 1996), while others, including the present study, have examined triple intrathecal medications (Brown et al., 1992), possibly explaining some of the differences across trials. Additional research is needed to compare neurocognitive functioning across various types of chemotherapy regimens and to evaluate for any potential effects of age at diagnosis.

Consistent with our prediction, girls performed more poorly than boys and the normative sample on nonverbal cognitive tasks (i.e., WISC-III PIQ, PO and FFD factors; the VMI). These findings are in accord with some investigations (e.g., Mulhern, Fairclough, & Ochs, 1991; Waber et al., 1990) that included children who received RT as well as chemotherapy. Taken together, it appears that girls evidence greater vulnerability than do their male counterparts to prophylactic whole brain RT and chemotherapy alone as a CNS prophylactic.

That over one third of the sample was receiving some type of special education services might be interpreted to contradict the finding of normative performance on the majority of the dependent measures. Because so many of the children were receiving special education services, this may have provided protection against any learning late effects that may have occurred if these services were not available. Further, these children may have received services based on their disease status alone rather than meeting actual diagnostic criteria (i.e., delayed achievement). In support of this notion, the Office of Education designates Other-Health Impaired as a special education category (Sexson & Madan-Swain, 1995), and many of the children in the present investigation were provided with special education services based on this designation.

The contribution of the present findings must be considered in light of the following caveats. First, the use of normative data instead of a matched control group limits both the interpretation and generalization of findings as differences between our sample and population norms may be due to differing demographic characteristics rather than the effects of chemotherapy. Future investigations should compare children with ALL who have received chemotherapy only as a CNS prophylaxis to siblings or those with other types of cancer to control for the cancer experience (e.g. solid tumors, children who have received CNS prophylactic RT) (Brown et al., 1996). Second, the lack of a baseline assessment for the children makes it difficult to determine the presence or magnitude of any changes in their scores following treatment. Longitudinal studies with appropriate controls would address this issue. Further, a longitudinal study would allow a comparison of children who survived the cancer experience relative to peers who died since initial baseline data would also include information on children who did not survive. It is possible that children who did not survive were more severely cognitively affected by both their disease and its treatment. Whether neurocognitive functioning is a marker for children with a more guarded prognosis cannot be addressed in this investigation. Third, limited power existed for the comparisons on some of the measures due to differences in assessment batteries administered at the various sites. This may have mitigated significant findings that otherwise
might have been detected. Fourth, whether or not the obtained effects are in part due to sampling error cannot be ascertained since only a small percentage of children treated on this protocol participated in the present investigation. Finally, our sample consisted of children who were off therapy for varying periods of time, ranging from approximately two to seven years. Our sample was not large enough to permit an examination of children’s functioning at each year off therapy.

In summary, children and adolescents who have received CNS prophylactic chemotherapy for ALL appear to be functioning generally within the normal range after treatment. Our data, albeit limited to specific areas of cognition, lend some support to the notion that cognitive deficits associated with CNS prophylactic therapy are not limited to radiation. However, these deficits appear to be mild and limited to specific areas of cognitive functioning (i.e., perceptual-organizational skills, attention, short-term memory, and visual-motor skills). Girls tend to be more vulnerable to these effects than are their male counterparts. For this reason, it is important that survivors of ALL receive ongoing screening to assess for any possible late effects that may be the result of CNS prophylactic therapy.

Acknowledgments

This research is a collaborative multicenter investigation from the Pediatric Oncology Group. We thank Kelly Byars, Deborah Karl, Wendy Kuppenheimer, and Susan Shaw for their assistance in data collection.

Received September 16, 1997; accepted April 7, 1998

References


Holingshead, A. B. (1975). Four factor index of social status. Unpublished manuscript, Yale University, New Haven, CT.


