Motor and Perceptual Timing Deficits Among Survivors of Childhood Leukemia

E. Mark Mahone,1,2 PhD, M. Cristine Prahme,1 MS, Kathy Ruble,2 MSN, Stewart H. Mostofsky,1,2 MD and Cindy L. Schwartz,3 MD

1Kennedy Krieger Institute, 2Johns Hopkins University School of Medicine, and 3Brown University School of Medicine

Objective There is growing evidence of cerebellar-frontal system change in children treated for leukemia with chemotherapy alone (Lesnik et al., 1998). Methods We compared 22 long-term survivors of acute lymphoblastic leukemia (ALL), aged 8–18, to 22 age- and gender-matched controls on tasks emphasizing cerebellar-frontal functioning including judgment of time duration and motor timing. Groups were also compared on a judgment of pitch task, used as a control measure. Children with ALL were at least 5 years from diagnosis, treated with intrathecal chemotherapy (methotrexate in all, hydrocortisone and cytarabine in 20/22), but not radiation therapy, and free from recurrence of disease. Results After controlling for IQ, the ALL group had poorer performance than controls on judgment of long duration and motor timing, but not judgment of pitch. Conclusions Treatment with intrathecal and infusional chemotherapy for childhood ALL may be associated with skill deficits comparable to those seen in individuals with cerebellar-frontal abnormalities.

Key words cerebellum; executive function; late effects; leukemia; neuropsychological timing.

Prophylactic treatment of the central nervous system (CNS) to prevent meningeal relapse of Acute Lymphoblastic Leukemia (ALL) was introduced in the late 1960s. Neuropsychological effects of cranial radiation have been well-described (Stehbens et al., 1991) and led to an attempt to replace radiation with chemotherapy. Most children now receive intrathecal chemotherapy (methotrexate ± hydrocortisone/cytosine arabinoside), often with infusional methotrexate. Chemotherapy supplanted radiotherapy as CNS prophylaxis at Johns Hopkins in 1979. From 1983 to 1990, pilot studies evaluated infusional methotrexate (±6-MP) with either intrathecal methotrexate or triple intrathecal medications. Infusional 6-mercaptopurine (6-MP) was studied in the 1980s and 1990s. Current regimens continue to rely on intrathecal chemotherapy, and intravenous or infusional chemotherapy to avoid need for cranial radiation. Characterization of the long-term neuropsychological outcome of these long-term patients is important to predict risks for current patients.

There is growing evidence of cerebellar-frontal system changes associated with treatment for childhood leukemia (Ciesielski, Harris, Hart, & Pabst, 1997). Lesnik et al. (1998) found that children treated with intrathecal chemotherapy, but not radiation therapy, showed morphometric changes in cerebellar lobules I–V and VI–VII and bilateral prefrontal cortices, relative to controls. In turn, these changes were related to deficits in behavioral measures of executive function, including working memory, visual organization and planning. The authors concluded that the timing of the insult to the developing nervous system selectively impacted the later maturing neocerebellum and prefrontal cortices, while sparing other areas of the brain. This evidence is consistent with the findings of Dawson et al. (1999), who found different patterns of executive dysfunction in children treated for ALL before age 6 and after age 6. Intrathecal chemotherapy and high-dose systemic chemotherapy are associated with later cognitive difficulties among children treated for ALL (Moore et al., 2000), including
visual motor skills (Buizer, De Sonneville, Van Den Heuvel-Eibrink, Njokiktjien, & Veerman, 2005), reading (Brown, Sawyer, Antoniou, Toogood, & Rice, 1999), auditory attention (Precourt et al., 2002), and arithmetic (Kaemingk, Carey, Moore, Herzer, & Hutter, 2004). Often the impairments associated with chemotherapy are subtle (Kingma et al., 2001), and girls appear to be differentially affected (Brown et al., 1998).

Executive functions are supported by a distributed neural network with cortical and subcortical components including the frontal lobes in circuit with the basal ganglia and with the cerebellum. Current models of frontal lobe structure and function describe a series of parallel frontal-subcortical circuits (Lichter & Cummings, 2001), with a minimum of five circuits, two of which are related to motor function, originating in skeletonmotor and oculomotor regions of the cortex; the other three, originating in dorsolateral prefrontal, anterior cingulate, and orbitofrontal cortices, are thought to be crucial in cognitive (“executive”) and socioemotional control. The cerebellum also plays a central role in higher cortical and executive functions (Ravizza & Ivry, 2001), which is not surprising, given that it has reciprocal connections with the frontal lobes, including prefrontal regions, and may be part of a larger distributed network in the intentional (executive control) system (Middleton & Strick, 2001).

Timing appears to be a salient aspect of cognitive function for which cerebellar-frontal circuits may be critical. A number of investigators have demonstrated evidence for the role of the cerebellum and prefrontal regions in motor timing. Compared to controls, individuals with known cerebellar lesions (Dennis et al., 2004), as well as those with learning disorders (Waber et al., 2000), Attention Deficit Hyperactivity Disorder (ADHD) (Rubia, Noorloos, Smith, Gunning, & Sergeant, 2003), and general “clumsiness” (Williams, Woollacott, & Ivry, 1992) showed increased variability on repetitive tapping tasks. Further, using fMRI, Rubia et al. (1999) identified reduced activation in right mesial prefrontal cortex among children with ADHD during motor timing tasks. Findings of motor timing deficits in ADHD are not surprising, since imaging studies have shown ADHD to be a disorder associated with prefrontal (Mostofsky, Cooper, Kates, Denckla, & Kaufmann, 2002) and cerebellar (Mostofsky, Reiss, Lockhart, & Denckla, 1998) abnormalities.

Problems with time estimation (i.e., judgment of time duration) are observed among clinical groups presenting with executive dysfunction. Barkley (1997) described time estimation as a frontally mediated executive function, suggesting that difficulties in this area indicate “inter-temporal competence” or the ability to associate events separated by time, which in turn, contribute to difficulties in time management, estimation of time durations, and inability to hold rules in mind while completing tasks (Barkley, Koplowitz, Anderson, & McMurray, 1997). For certain paradigms, however, cognitive models of time estimation implicate neural mechanisms in addition to frontally mediated executive functions, including cerebellar timing mechanisms (Ivry & Fiez, 2000). For example, Ivry and Keele (1989) found that adults with cerebellar lesions performed more poorly on tasks of time (interval) estimation than those with cortical lesions or Parkinson’s disease. Deficits on tasks of interval judgment have been observed among children with posterior fossa tumors (Hetherington, Dennis, & Spiegler, 2000) and adolescents with ataxia-telangiectasia (Mostofsky, Kunze, Cutting, Lederman, & Denckla, 2000). Mangels (Ivry, & Shimizu 1998) reported that while individuals with cerebellar lesions had impairments in judgment of both short (milliseconds) and long (seconds) interval duration, those with frontal lesions had impairment only on judgment of long duration, but had normal performance when judging short intervals. Given the working memory demands associated with the long duration task, these findings were thought to implicate separate cerebellar and frontal contributions to processing of temporal information, with neocerebellar regions contributing to a central timing mechanism, and frontal cortex involved in acquisition and monitoring of information in working memory. Radonovich and Mostofsky (2004) reported findings in children with ADHD similar to those observed in the Mangels’ (1998) frontal lesion group. In the Radonovich and Mostofsky study, children with ADHD were impaired on judgments of long durations (4s), but not on short durations (550 ms), suggesting that the impairments in ADHD are at the level of utilization of temporal information, rather than a more central timing deficit. Thus, if prophylactic treatment for ALL produces a pattern of neurobiological differences similar to that seen developmentally in ADHD, ALL survivors may also present with patterns of impairments in time estimation and motor timing.

In summary, the literature suggests that the effect of prophylactic chemotherapy is related to abnormalities in development of cerebellar-frontal brain systems. In turn, neurobehavioral functions related to these affected brain systems (i.e., motor timing, judgment of duration)
may be disrupted in children treated with chemotherapy. Among tasks of duration judgment, the detrimental effects of chemotherapy are more likely to be observed under demands for judgment of long (vs. short) durations—similar to the findings in ADHD and frontal lesion groups. The purpose of the present study was to examine motor and perceptual timing in children with a history of ALL treated with chemotherapy, but not radiation therapy. We hypothesized that, compared to matched controls, the ALL group would show deficits on both motor timing and judgment of duration. Among judgment of duration tasks, we hypothesized that the ALL group would show greater deficits on judgment of long (vs. short) duration.

Method
Participants

Patients for the current study were identified through the Long-term Follow-up Program in the Johns Hopkins Hospital Department of Pediatric Oncology between 2001 and 2003. Children were included in the ALL group if they were more than 5 years from diagnosis, and had been diagnosed with ALL after age 12 months, and were free from recurrence of disease (POG protocols 9006, 9005, 8699, 8698). A total of 22 patients (12 boys, 10 girls) agreed to participate (from a pool of approximately 30 seen in that clinic during that time period who otherwise met general entry/exclusion criteria). Treatment included intrathecal chemotherapy (methotrexate in all, hydrocortisone and cytarabine in 20 of the 22), but not radiation therapy. Infusional methotrexate (1–2.5 g/m²/dose) was standard, with infusional 6-MP given to 16 of the 22. Children with ALL were excluded if they had previous diagnosis (through review of records) of mental retardation, or known visual or hearing impairment. Intellectual level was screened in the ALL group using the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). All, but two, of the children in the ALL group were diagnosed at or before age 6; one was diagnosed at age 9, and one at age 10 (mean age at diagnosis = 3.5 years; range = 12 months to 10 years). The mean number of years since completion of chemotherapy treatment was 6.2 (range = 3–11 years). A total of 22 healthy children were selected from ongoing research projects at the Kennedy Krieger Institute, and matched individually for age, gender, and SES to children in the ALL group. Children were confirmed as controls (i.e., no DSM-IV diagnoses) based on administration of the Diagnostic Interview for Children and Adolescents-IV (DICA-IV; Reich et al., 1997), and two sets of behavior rating scales, each of which was completed by parents and teachers: Conners’ Rating Scales (CPRS-R, CTRS-R; Conners, 1997) and ADHD Rating Scale-IV (ARS; DuPaul, Power, Anastopoulos, & Reid, 1998). For the control group, intellectual level was assessed using the Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 1991); only children with Full Scale IQ of ≥80 were included. None of the children in the control group were taking psychotropic medication.

Results

Participants completed the three experimental tests and IQ over two days as part of a larger battery of neuropsychological tests. Motor timing and judgment of explicit time intervals were studied using computer-based timing tasks based on previous work by Ivry and Keele (1989). An additional judgment of pitch (frequency perception) task was included to determine whether findings on the judgment-of-interval tasks might be due to a general auditory processing deficit.

Motor Timing Task

The test of motor timing is a repetitive finger-tapping test. Each trial begins with a series of computer tones presented at regular intervals (inter-tone interval of 550 ms). The subject is instructed to tap a button with the index finger (dominant hand), synchronizing responses with the tones produced by the computer. After 13 paced responses, the tones cease, and the subject’s task is to continue tapping at the same rate for 30 unpaced inter-tap intervals. The task takes approximately 10 min, and performance is measured by determining the standard deviation of the inter-tap trials.

Judgment of Interval Tasks

The judgment of short interval task consists of a series of trials, during which the two intervals are presented, a standard and a comparison interval. Two pairs of tones were presented at 73 dB, 50 ms in duration, and at a frequency of 1,000 Hz. On this task, the first pair of tones was separated by a 550 ms delay. After 1 s, a second pair of tones was presented at variable durations. Subjects had to say whether the second delay was shorter or longer than the first. A parameter estimation by sequential testing (PEST) procedure was used (Taylor & Creelman, 1967) to determine an upper and lower threshold. Half of the trials estimated the lower threshold (the point at which the subject correctly responded “shorter” on ~90% of trials) and half estimated the higher threshold (the point at which the subject correctly responded “longer” on ~90% of the trials). During these trials,
the closer the thresholds were to the standard interval of 550 ms, the better the performance. Overall scores were calculated based on the difference between the higher and lower thresholds, with lower scores indicating better performance.

The judgment of long interval task was the same as the judgment of short interval task, except that the delay in tone presentations was longer. Two pairs of tones were presented at 73 dB, 50 ms in duration, and at a frequency of 1000 Hz. The first pair of tones was separated by a 4000 ms delay. After 1 s, a second pair of tones was presented at variable durations. Subjects had to say whether the second delay was shorter or longer than the first. Half of the trials estimated the lower threshold (the point at which the subject correctly responded “shorter” on ~90% of trials) and half estimated the higher threshold (the point at which the subject correctly responded “longer” on ~90% of the trials). During these trials, the closer the thresholds were to the standard interval of 4000 ms, the better the performance. Overall scores were calculated based on the difference between the higher and lower thresholds, with lower scores indicating better performance.

Judgment of Pitch Task
Two pairs of tones were 73 dB loud, 50 ms in duration, and separated by 550 ms. The first pair of tones was presented at a standard pitch of 1000 Hz. A second pair of tones was presented at variable frequency. Subjects had to say whether the second pair was lower or higher (in pitch) than the first. PEST procedure was again used. Half of the trials estimated the lower threshold (the point at which the subject correctly responded “lower” on ~90% of trials) and half estimated the higher threshold (the point at which the subject correctly responded “higher” on ~90% of the trials). Overall scores were calculated based on the difference between the higher and lower thresholds, with lower scores indicating better performance. This task was analogous to the interval judgment task presented earlier.

Data Analyses
Group means for control and ALL groups were compared using a series of four ANCOVAs, controlling for FSIQ. Motor timing and the two judgment of interval tasks were normally distributed. Judgment of pitch task had a positively skewed distribution; therefore, a log transformation was used prior to the parametric analyses for that measure. Given the small sample size, Spearman rank-order correlations were used to guard against outliers influencing the magnitude of association, when examining the relationship between demographic variables (i.e., diagnosis age, time since treatment) and performance in the ALL group. The relative difficulty of judgment of long versus short duration was examined using a repeated measures ANCOVA with duration as the within-subjects factor. Effect size values were computed using Cohen’s $d$ statistic. Group differences were considered statistically significant if $p < .05$; however, given the relatively small sample size, findings of $d > .50$ were considered worth fully describing as they may represent meaningful effects.

Results
Demographic Information
The overall sample was drawn from a largely middle class SES and was predominantly Caucasian (89% Caucasian, 7% African-American, 2% Hispanic, and 2% mixed ethnic/racial groups). Children ranged in age from 8 to 18 years (mean age: ALL = 11.8, SD = 3.2; controls 11.6, SD = 2.9). None of the children in the study were taking stimulant medication at the time of the assessment, although one child in the ALL group was taking Neurontin. There were no significant differences between children with ALL and controls in age ($t \[42\] = 0.20, $p = .85$), SES/Hollingshead Index ($t \[32\] = 0.44, $p = .66$), handedness ($\chi^2 = 3.02, p = .22$), or racial distribution ($\chi^2 = 2.56, p = .46$). The ALL and control groups were also matched in sex distribution (12 boys, 10 girls in each group). The control group had higher mean FSIQ (111.6) than the ALL group (103.1; $t \[42\] = 2.05, $p = .05$). As such, IQ was used as a covariate in subsequent analyses. Within the ALL group, age at diagnosis did not significantly correlate with judgment of long duration ($\rho = -0.09, p = .71$), judgment of short duration ($\rho = -0.08, p = .70$), motor timing ($\rho = -0.24, p = .29$), or log transformed judgment of pitch ($\rho = -0.06, p = .79$). Number of years since treatment was not significantly correlated with judgment of long duration ($\rho = -0.06, p = .79$), judgment of short duration ($\rho = -0.07, p = .80$), or motor timing ($\rho = -0.20, p = .39$); however, log transformed judgment of pitch ($\rho = -0.45, p = .03$) was significantly negatively correlated with time since treatment, suggesting that performance on this task improves with time out of treatment.

Group Comparisons (ALL vs. Control) on Experimental Measures
Group comparisons on timing and pitch perception tasks are presented in Table 1. After controlling for FSIQ, children in the ALL group had significantly greater
variability than matched controls on the motor timing task ($F[1,41]=5.46$, $p=.02$, $d=.65$). Similarly, children in the ALL group had a larger difference between upper and lower threshold on the judgment of long interval task than controls ($F[1,41]=4.05$, $p=.05$, $d=.66$). There was a trend for greater difference between upper and lower threshold on the judgment short interval task ($F[1,41]=3.3$, $p=.07$, $d=.57$); however, there were no group differences observed on the log transformed judgment of pitch task ($F[1,41]=0.74$, $p=.39$, $d=.27$). Using repeated measures ANCOVA for the two duration tasks, the main effect for group ($F[1,41]=3.98$, $p=.05$, $d=.62$) the interaction effect between group and duration ($F[1,41]=4.09$, $p=.05$, $d=.63$) both showed strong trends. For the interaction, the relative magnitude of group differences (control > ALL) was greater on the long (vs. short) duration task.

### Discussion

The present study demonstrates that treatment with intrathecal and infusional chemotherapy for childhood ALL may be associated with impairment on measures of motor and perceptual timing, even when statistically controlling for IQ. Conversely, chemotherapy treatment does not appear to affect judgment of pitch, suggesting that the observed deficits in the ALL group are not due to more pervasive problems with attention (e.g., careful attention to task), or with auditory perception in general. Indeed (although not reported in the present analyses), the ALL group had a mean Digit Span scaled score of $9.95\pm2.66$—solidly in the average range (range = 6–17), suggesting that auditory attention is grossly intact. These findings support the hypothesis that late effects of chemotherapy treatment may preferentially affect brain systems supporting motor timing (i.e., cerebellum and prefrontal cortex) and judgment of longer time intervals (also cerebellum and prefrontal cortex). Further, the fact that children with ALL showed greater relative difficulty on the judgment of long duration, compared to judgment of short duration task, is generally consistent with previous research on individuals with frontal lobe abnormalities, including adults with frontal lobe lesions (Mangels et al., 1998) and children with ADHD (Radonovich & Mostofsky, 2004). Thus, while some clinical groups (e.g., children with ADHD) may have impairment at the level of utilization of temporal information (implicating greater frontal system involvement), the late effects of intrathecal and infusional chemotherapy may be associated with both utilization and more general timing deficits, similar to those with cerebellar lesions (Dennis et al., 2004; Mangels et al., 1998). Given the deficits in both motor and perceptual timing, dysfunction in both prefrontal and cerebellar systems appears most likely. Alternatively, the findings might also be explained by anomalies in the development of white matter, which have been observed in ALL survivors and associated with reduced attention and processing speed (Reddick et al., 2006).

The present study provides additional behavioral support to the notion that early chemotherapy treatment preferentially affects developing brain systems that have later and more prolonged course of postnatal development (e.g., the posterior neocerebellum—lobules VI–VIII, and prefrontal cortex). Neurogenesis and migration of granule cells in the cerebellum continue through the first years of life in humans. As such, early disruption by toxic insult can disrupt the normal structural and functional interaction between granule cells with Purkinje cells and radial glial fibers (Rakic, 1971), resulting in cerebellar hypoplasia and atrophy (Lesnik et al., 1998). The acquired cerebellar dysfunction following treatment for ALL may mimic the attention and motor control deficits observed in children with congenital disorders affecting development of the cerebellar vermis (e.g., ADHD, autism), both in terms of structure and function (Cielieski et al., 1997).

It has been argued that the adverse effects of intrathecal and systemic chemotherapy are most pronounced when administered at younger ages (Mulhern, 1994). Giralt et al. (1992) found that those children with ALL diagnosed and treated before age 5 years had higher

**Table I. Judgment of Duration, Pitch and Motor Timing for ALL Survivors and Matched Controls**

<table>
<thead>
<tr>
<th></th>
<th>ALL ($n=22$)</th>
<th>Control ($n=22$)</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor timing* (ms)</td>
<td>41.0 (10.7)</td>
<td>35.1 (7.8)</td>
<td>0.66</td>
</tr>
<tr>
<td>Judgment of long duration* (ms)</td>
<td>198.2 (60.1)</td>
<td>149.7 (92.5)</td>
<td>0.65</td>
</tr>
<tr>
<td>Judgment of short duration (ms)</td>
<td>49.6 (15.0)</td>
<td>38.9 (22.7)</td>
<td>0.57</td>
</tr>
<tr>
<td>Judgment of pitch (Hz)</td>
<td>17.1 (12.2)</td>
<td>14.2 (14.6)</td>
<td>0.22</td>
</tr>
<tr>
<td>Judgment of pitch [Log transformed] (Hz)</td>
<td>2.6 (0.7)</td>
<td>2.3 (0.8)</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Note: ALL, Acute Lymphoblastic Leukemia; Standard deviations in brackets. *$p<.05$; **$p<.01$; $d$ = effect size Cohen’s $d$. 


risk of presenting with long-term cognitive sequelae. Thus, most of the children in our ALL sample were at greater risk for cognitive dysfunction because of their early diagnosis and treatment. A total of 68% of our sample (15/22) were diagnosed by age 3 years, and 82% by 5 years (18/22). Indeed, we did observe a moderate (albeit nonsignificant) negative correlation between age at diagnosis and performance on motor timing, generally supporting the notion that earlier onset of disease and earlier treatment may be related to greater negative effects associated with cerebellar development. Early introduction of chemotherapy may disrupt the rapid development of cerebellar Purkinje cells and white matter, which occurs during the first five years of life. Autopsy studies have identified white matter reduction (Fletcher & Copeland, 1988) and cerebellar sclerosis (Wizniter, Packer, Rourke, & Meadows, 1987) among those treated with chemotherapy alone. Specifically, the use of intrathecal methotrexate, hydrocortisone and cytarabine in combination (i.e., triple intrathecal therapy) has been linked to frontal-cerebellar atrophy (Lesnik et al., 1998) as well as leukoencephalopathy in children (Rubenstein, Herman, Long, & Wilbur, 1975), while systemic treatment has been associated with cerebellar degeneration (Stentoft, 1990). More recently, white matter fractional anisotropy (FA) changes following chemotherapy treatment (with and without radiation therapy) have been identified via diffusion tensor magnetic resonance imaging (DTI), with FA changes associated with reduction in IQ (Khong et al., 2006). Whether or not these neurological and neuropsychological changes result in measurable academic decline, however, remains unclear (Kingma et al., 2001).

Strengths of the present study included matching on age, gender, and SES and statistical control for effects related to IQ. There were several limitations with the current study. First, it is possible that our sample ALL survivors may not be representative, either because those who actually attend long-term effects clinics are less healthy overall, or due to other biological (e.g., pharmacokinetics and drug metabolism), socioeconomic (i.e., healthcare access), or other unknown factors associated with racial/ethnic disparities in outcomes (Kadan-Lottick, Ness, Bhatia, & Gurney, 2003). It is also possible that group differences may have been enhanced by the strict exclusionary criteria for the control group. Second, the failure to find group differences on the control (pitch) task may have been because it was relatively less difficult (i.e., positively skewed) than the other tasks. Although the results remained the same even when analyzing the “normalized” log transformed data, future research should attempt to ensure that control tasks are of equivalent difficulty level. Third, although we inferred from our data that our findings coincided with changes in cerebral and cerebellar structures, we did not directly link our findings to imaging studies in these children. Future research should continue, using larger samples to examine whether these timing deficits in ALL are associated with frontal-cerebellar changes on MRI. More specifically, research should examine whether treatment induced changes in prefrontal cortex, frontal-striatal white matter, or cerebellar vermis are related to motor and perceptual timing deficits in long-term ALL survivors. The judgment of long (vs. short) duration task may place greater demands on working memory, and as such, may have a stronger association with MRI volumes in dorsolateral prefrontal cortex. It will also be important to clarify the functional life skills associated with deficits in motor timing, including those related to motor speed, efficiency, and inhibitory control. Judgment of duration tasks may have particular relevance in determining skill in higher level executive functions, including future planning, organization of time, prospective memory, and adherence to rules (Barkley et al., 1997). Motor timing deficits may interfere with tasks requiring motor coordination, such as handwriting (Waber et al., 2000). Perceptual and motor timing deficits may also be associated with increased difficulties in social communication—particularly with regard to timing of responses (Williams et al., 1992).


