Feasibility of neuropsychological assessment in leukaemia patients shortly after diagnosis: directions for future prospective research

N C Jansen, A Kingma, P Tellegen, R I van Dommelen, A Bouma, A Veerman, W A Kamps

Aims: To study neuropsychological functioning of newly diagnosed children with acute lymphoblastic leukaemia (ALL) within two weeks after diagnosis in order to determine the feasibility of a sibling controlled prospective study design.

Methods: Fifty consecutive patients (median age at testing 6.6 years, range 4–12) were included in a prospective, longitudinal, nationwide study. Treatment would include intrathecal and systemic chemotherapy according to the DCLSG ALL-9 protocol. Children were evaluated with an extensive neuropsychological battery including measures of intelligence, memory, attention, language, visual-constructive function, and fine-motor abilities within two weeks after start of the chemotherapy. The control group consisted of 29 healthy siblings (median age at testing 8.2 years, range 4–12), who were tested <4 weeks after the patients’ assessment.

Results: Mean scores on Wechsler Intelligence Scales did not differ significantly between patients and siblings; mean IQ scores for both the patients and the controls were high average. To examine specific neuropsychological functions, norm scores based on the exact age were acquired by fitting procedures, but no significant differences were found.

Conclusions: Neuropsychological assessment of patients during early hospitalisation is feasible. The results indicate no adverse effect of illness and psychological factors on IQ and neuropsychological functioning of patients with recently diagnosed ALL. The prospective design of this study of cognitive late effects of chemotherapy will allow discrimination between adverse sequelae of disease and treatment.

Abbreviations: ALL, acute lymphoblastic leukaemia; CI, cranial irradiation; DCLSG, Dutch Childhood Leukaemia Study Group; SES, socioeconomic status
Table 1 Characteristics of patients and siblings at the first neuropsychological evaluation shortly after diagnosis of ALL

<table>
<thead>
<tr>
<th>Group</th>
<th>Male n (%)</th>
<th>Female n (%)</th>
<th>Age at testing Median (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>30 (60)</td>
<td>20 (40)</td>
<td>6.6 (4–12)</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>11 (38)</td>
<td>18 (62)</td>
<td>8.2 (4–12)</td>
</tr>
</tbody>
</table>

Table 2 Neuropsychological battery

<table>
<thead>
<tr>
<th>Neuropsychological domain</th>
<th>Measures</th>
<th>Age</th>
<th>No. patients</th>
<th>No. siblings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental, motor, and social development</td>
<td>Denver Developmental Scales</td>
<td>4–6</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>Intelligence</td>
<td>Wechsler Pre-School and Primary Scales of Intelligence (WPPSI-R), 10 subtests</td>
<td>4–6</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>FS-IQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V-IQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-IQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligence</td>
<td>Wechsler Intelligence Scale for Children (WISC-R), 10 subtests</td>
<td>6–13</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>FS-IQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V-IQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P-IQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptual organisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal-auditory learning and memory</td>
<td>Dutch version of Rey’s Auditory-Verbal Learning Test (RAVLT)</td>
<td>6–13</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>Visual memory</td>
<td>Ray-Osterreith Complex Figure Test delayed (CFT) recall</td>
<td>6–13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal fluency</td>
<td>Animal naming fluency test</td>
<td>6–13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustained attention/speed</td>
<td>Bourdon-Vos; self-paced, continuous performance cancellation task</td>
<td>6–13</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Cognitive flexibility</td>
<td>Wisconsin Card Sorting Test (WCST)</td>
<td>6–13</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>Perceptual-motor skills</td>
<td>Beery Developmental Test of Visual-Motor Integration (VMI)</td>
<td>4–13</td>
<td>47</td>
<td>28</td>
</tr>
<tr>
<td>Visual/spot constructional ability/planning</td>
<td>Ray-Osterreith Complex Figure Test (CFT) copy</td>
<td>6–13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fine-motor function</td>
<td>Purdue Pegboard</td>
<td>5–13</td>
<td>29</td>
<td>27</td>
</tr>
</tbody>
</table>

Treatments

Patients had just started treatment according to the national chemotherapy only DCLSG ALL-9 protocol, including vincristine, dexamethasone, daunorubicin, and triple intrathecal (IT) therapy as CNS prophylaxis. This protocol is similar to the DCLSG ALL-6 protocol. The patients had received one cycle of vincristine, dexamethasone, daunorubicin, and triple IT therapy before their first assessment.

Study design

Patients were individually evaluated within two weeks after diagnosis and start of treatment. Siblings were individually assessed within four weeks after the patients’ evaluation. Patients and siblings did not significantly differ with respect to age at testing and gender. There were no indications that patients with or without siblings differed in socioeconomic status (SES). To maximise standardisation, all participants were nationwide tested by one qualified child neuropsychologist who travelled to the hospitals were the children were treated.

Patients and healthy sibling controls were evaluated with an age appropriate comprehensive standardised neuropsychological test battery (table 2). Children aged 4–6 years were administered a developmental screening test and measures of intelligence, visual-motor integration, and, if necessary, the assessment was split into two sessions.

Statistical analysis

Performances of patients were compared to those of sibling controls using non-directional two tailed Student’s t tests for paired groups.

For the Wechsler Pre-School and Primary Scales of Intelligence (WPPSI-R), Experimental Dutch-Flemish version, and for the Wechsler Intelligence Scale for Children-revised (WISC-R, Dutch version), mean norm scores are 100 (SD = 15). For the remaining tests, norms have been acquired by fitting procedures based on the raw scores and the exact ages resulting in norm scores (mean = 50; SD = 10). The fitting procedures were based on the published norm data (means and standard deviations for different age groups) in the respective test manuals or other publications. This procedure enables comparisons of standardised scores between subjects of any specific age.22

Significance levels were established at p < 0.05. Statistical analyses were performed using version 10 of the Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL, USA).

RESULTS

Included patients (n = 50) did not significantly differ from missing patients (n = 35) in terms of sex, age at diagnosis, and initial characteristics of disease and prognostic risk group. We had no indication of differences in socioeconomic status between included and missing patients. The latter mainly emanated from two hospitals; patients were missed due to illness of the psychologists who should have referred eligible patients. Patients and siblings aged 4–6 years at diagnosis were assessed as essentially normal on the Denver Developmental Scales. Patients aged 4–6 years scored significantly higher than siblings on WPPSI-R FS-IQ and WPPSI-R VIQ (table 3). Comparing patients aged 6–13 and siblings, no significant differences were found for any WISC-R factor. IQs were high average for patients on the WPPSI-R and both patients and siblings on the WISC-R.

Table 4 shows results for the remaining cognitive measures. No significant differences between the groups were found for any test. Overall, patients and siblings had average scores.
DISCUSSION

We have shown that recently diagnosed children with the life threatening disease ALL can be reliably assessed with an extensive standardised neuropsychological test battery shortly after diagnosis. An important observation in this study was that the majority, even the very young children, enjoyed the assessment, which was rather a distraction among numerous medical procedures than an emotional burden. Moreover, this study is strengthened by the inclusion of healthy siblings as controls, who were also pleased to be involved in the study and enjoyed the special attention. This control group enables appropriate comparison with the healthy population. Increments in test results within the patient group can be detected, even if the results are still above average. For accurately assessing changes, precise standardised age scores are essential. Therefore the validity of this study is enhanced by using a fitting procedure for the construction of test norms which provides standardised age scores. 

Our data correspond with the few other studies offering neuropsychological tests comparing the ALL group with siblings at the first evaluation shortly after diagnosis of ALL. 

The present study can be criticised for the high number of missing patients, which could possibly account for bias in these test results. However, this is unlikely because included patients did not significantly differ from missed patients concerning demographic and initial disease characteristics. Missed patients should mainly have been referred by two ill psychologists. Fortunately, patients in these hospitals represent a random patient population, hence we have no indication that characteristics of the missed children differed from those who could be included. With the current numbers we would detect IQ differences of 0.7 SD (10.5 IQ points) to obtain an adequate power of 80%. To illustrate the meaning of 10.5 SD, a difference between 105 and 95 would be significant, but both IQs are considered average and children with both IQs would be in a regular school class. We could not control for SES. Given the overall average results, bias does not seem likely. In general, patients showed greater standard deviation on both the intelligence tests and neuropsychological tests. However, differences between patients and siblings did not result from one or few individuals with extreme scores.

The norm scores of the Experimental Dutch-Flemish version of the WPPSI-R were recently evaluated as disputable, which could explain the above-average IQs in the young patients (table 3). However, if the patients’ IQs are overrated, we could expect above average IQs in the siblings as well. There were no demographic differences explaining the IQ differences between patients and siblings aged 4–6 years. The scores of the children tested with the WISC-R are high average as a result of the Flynn effect, accounting for an indication that characteristics of the missed children differed from those who could be included. With the current numbers we would detect IQ differences of 0.7 SD (10.5 IQ points) to obtain an adequate power of 80%. To illustrate the meaning of 10.5 SD, a difference between 105 and 95 would be significant, but both IQs are considered average and children with both IQs would be in a regular school class. We could not control for SES. Given the overall average results, bias does not seem likely. In general, patients showed greater standard deviation on both the intelligence tests and neuropsychological tests. However, differences between patients and siblings did not result from one or few individuals with extreme scores.

The norm scores of the Experimental Dutch-Flemish version of the WPPSI-R were recently evaluated as disputable, which could explain the above-average IQs in the young patients (table 3). However, if the patients’ IQs are overrated, we could expect above average IQs in the siblings as well. There were no demographic differences explaining the IQ differences between patients and siblings aged 4–6 years. The scores of the children tested with the WISC-R are high average as a result of the Flynn effect, accounting for an indication that characteristics of the missed children differed from those who could be included. With the current numbers we would detect IQ differences of 0.7 SD (10.5 IQ points) to obtain an adequate power of 80%. To illustrate the meaning of 10.5 SD, a difference between 105 and 95 would be significant, but both IQs are considered average and children with both IQs would be in a regular school class. We could not control for SES. Given the overall average results, bias does not seem likely. In general, patients showed greater standard deviation on both the intelligence tests and neuropsychological tests. However, differences between patients and siblings did not result from one or few individuals with extreme scores.

The norm scores of the Experimental Dutch-Flemish version of the WPPSI-R were recently evaluated as disputable, which could explain the above-average IQs in the young patients (table 3). However, if the patients’ IQs are overrated, we could expect above average IQs in the siblings as well. There were no demographic differences explaining the IQ differences between patients and siblings aged 4–6 years. The scores of the children tested with the WISC-R are high average as a result of the Flynn effect, accounting for an indication that characteristics of the missed children differed from those who could be included. With the current numbers we would detect IQ differences of 0.7 SD (10.5 IQ points) to obtain an adequate power of 80%. To illustrate the meaning of 10.5 SD, a difference between 105 and 95 would be significant, but both IQs are considered average and children with both IQs would be in a regular school class. We could not control for SES. Given the overall average results, bias does not seem likely. In general, patients showed greater standard deviation on both the intelligence tests and neuropsychological tests. However, differences between patients and siblings did not result from one or few individuals with extreme scores.

The norm scores of the Experimental Dutch-Flemish version of the WPPSI-R were recently evaluated as disputable, which could explain the above-average IQs in the young patients (table 3). However, if the patients’ IQs are overrated, we could expect above average IQs in the siblings as well. There were no demographic differences explaining the IQ differences between patients and siblings aged 4–6 years. The scores of the children tested with the WISC-R are high average as a result of the Flynn effect, accounting for an indication that characteristics of the missed children differed from those who could be included. With the current numbers we would detect IQ differences of 0.7 SD (10.5 IQ points) to obtain an adequate power of 80%. To illustrate the meaning of 10.5 SD, a difference between 105 and 95 would be significant, but both IQs are considered average and children with both IQs would be in a regular school class. We could not control for SES. Given the overall average results, bias does not seem likely. In general, patients showed greater standard deviation on both the intelligence tests and neuropsychological tests. However, differences between patients and siblings did not result from one or few individuals with extreme scores.
IQ rise of about 6 points, since test norms were collected in the early 1980s. If evaluated with more recent test norms these children would probably have average results.

Generally, it is often suggested that emotional, non-organic distress influences the test results. However, such an effect is very unlikely given the normal outcome. Even measures of attention and memory, known to be sensitive for emotional distress, did not differ between patients and siblings.

Conclusion
The present data strongly suggest that patients do not suffer from neuropsychological deficits related to acute disease or early treatment. In the future, patients’ baseline scores can be used to discriminate between possible adverse sequelae of disease and/or treatment and eventually, to optimise treatment protocols compromising between high cure rate and good quality of life. Ideally, neuropsychological assessment early after hospitalisation also selects patients who need early intervention for mental or academic deficits, but this was not the aim of this study.

Neuropsychological assessment of children with ALL shortly after diagnosis with sibling controls is feasible and essential to discriminate between adverse sequelae of treatment. Prospective, longitudinal study designs should become the standard for evaluating possible treatment effects.

ACKNOWLEDGEMENTS
This study is a collaborative multicentre investigation, which is coordinated by the Paediatric Oncology Centre, Groningen University Hospital. Participating centres: Wilhelmina Children’s Hospital, Utrecht; St Radboud University Medical Centre, Department of Paediatrics, Nijmegen; Vrije Universiteit University Medical Centre, Department of Paediatrics, Amsterdam; University Medical Centre, Emma Children’s Hospital, Amsterdam; University Medical Centre, Willem-Alexander Children and Youth Centre, Leiden.

The authors would like to thank the children and families who participated in this study, and dr A Jennekens-Schinkel for reviewing this paper.

Authors’ affiliations
N C Jansen, A Kingma, R I van Dommelen, W A Kamps, Departments of Paediatric Haematology Oncology, Groningen University Hospital, Groningen, Netherlands
P Tellegen, A Bouna, Department of Psychology, University of Groningen, Groningen, Netherlands
A Veerman, Dutch Childhood Leukemia Study Group, The Hague, Netherlands

Contract grant sponsor: Dutch Cancer Society
Competing interests: none declared

Parts of this manuscript were presented at the XXXIV Meeting of the International Society of Pediatric Oncology (SIOP), 18–21 September 2002, at Porto, Portugal

REFERENCES