Title
Cognitive development in schizophrenia: Follow-back from the first episode

Journal
Journal of Clinical and Experimental Neuropsychology, 28(2)

ISSN
1380-3395

Authors
Bilder, Robert M
Reiter, Gail
Bates, Jay
et al.

Publication Date
2006-02-01

Peer reviewed

Cognitive Development in Schizophrenia:

Follow-Back from the First Episode

Robert M. Bilder[1], Gail Reiter[2], Jay Bates[3], Todd Lencz[2], Philip Szeszko[2], Robert S. Goldman[4], Delbert Robinson[2], Jeffrey A. Lieberman[5], and John M. Kane[2]

[1]UCLA Neuropsychiatric Institute

[2]Zucker Hillside Hospital Division of North Shore – Long Island Jewish Health System

[3]Bristol Myers Squibb

[4]Pfizer, Inc.

[5]University of North Carolina – Chapel Hill

Correspondence to:

Robert M. Bilder, Ph.D.
UCLA Neuropsychiatric Institute, Room C8-849
Los Angeles, CA 90095
310-85-9474
rbilder@mednet.ucla.edu

Acknowledgements. This work was supported by grants from the National Institutes of Health (MH060575, MH060374, and RR020750).

Abstract

Despite consensus that schizophrenia is a neurodevelopmental disorder characterized by cognitive deficits, objective data documenting the course of cognitive development remain sparse. We conducted a "follow-back" study to examine premorbid cognitive ability in individuals who later went on to develop schizophrenia or schizoaffective disorder, and a group of demographically matched healthy volunteers. We obtained school records containing standardized achievement test scores from the 1st through 12th grades, and scholastic aptitude test results from the 11th and 12th grades, and examined the developmental trajectories of cognitive performance with respect to prospective examinations conducted following participants' enrollment in our study of first episode psychosis. We found significant differences in academic achievement tests as early as the first grade, with scores from participants who would later develop schizophrenia lagging behind their peers by 0.8 to 1.1 grade equivalents. This gap widened resulting in a difference between groups of 1.5 to 1.8 grade equivalents by the 12th grade. In the subset of patients for whom SAT scores were available, we found that WAIS-R Full Scale IQ was 11.5 points lower than predicted from earlier SAT scores, suggesting a substantial decline in cognitive ability accompanying the initial episode of illness. These findings suggest that schizophrenia is marked by substantial cognitive deficits in the first grade, that there may be additional subtle decline preceding the overt onset of psychotic symptoms, and that the initial episode of illness is marked by additional decline. These observations may help advance concepts of premorbid cognitive ability in the schizophrenia syndrome and constrain models of pathophysiology.

Keywords: schizophrenia, cognition, neuropsychology, development, intelligence

Introduction

Neuropsychological methods are often used to define psychometric deficits in patients relative to an identified healthy population. These methods may also consider whether current deficits reflect decrements from "premorbid" cognitive attainments. One challenge in understanding cognitive deficits in schizophrenia has been the determination of when – and if – a genuine *premorbid* period exists. There is now widespread recognition that neurodevelopmental anomalies, marked by a broad range of cognitive, affective, and neuromotor disturbances, are apparent well before symptom onset and probably as early as anyone can accurately measure integrative neural systems function (Cornblatt, Obuchowski, Roberts, Pollack, & Erlenmeyer-Kimling, 1999; Erlenmeyer-Kimling et al., 1995; Erlenmeyer-Kimling et al., 1997; Erlenmeyer-Kimling et al., 2000; Walker, Diforio, & Baum, 1999; Walker & Levine, 1990). There is consensus that deficits exist early, and that by the time of the first episode these deficits are large, with generalized deficit approximating 1.5 standard deviations compared to healthy comparison groups (Bilder et al., 2000). But it has remained controversial precisely how much of this deficit is present from the beginning of life, and how much may reflect a process of cognitive decline preceding the overt onset of symptoms. Resolving these questions may have major implications for narrowing the range of possible pathological processes responsible for schizophrenia.

There are sparse empirical data documenting the trajectory of cognitive development followed by people who go on to develop schizophrenia. One approach takes population samples for which cognitive measurements were conducted for some other reason, and looks for differences between the individuals who went on to develop illness and the rest of the sample. For example, Davidson and colleagues found impairments on national draft tests taken at ages 16-17 in individuals who later developed schizophrenia (Davidson et al., 1999; Reichenberg et

al., 2002). In another example, Cannon and colleagues detected cognitive impairments that remained stable from ages 4 to 7 among the sub-sample of a large birth cohort who went on to develop schizophrenia (Cannon et al., 2000). Limitations of this approach include the limited scope and timing of data that were collected for other purposes, and the need for large samples to yield sufficient numbers of cases who go on to develop schizophrenia. It is also possible to conduct more detailed prospective assessments of "high risk" samples. Both the "genetic high risk" strategy (e.g., studying family members of people with schizophrenia), and the "behavioral high risk" strategy (e.g., studying people who show early signs of illness) have been used to enrich the yield of samples to include more individuals likely to develop schizophrenia, but each of these approaches is subject to its own biases.

There are several alternate approaches to study developmental course of cognitive function. One is to use estimates of premorbid function, either based on cognitive test scores considered insensitive to deterioration, or to compute demographically based predictions, and examine post-illness deviations from these estimates (Bilder et al., 1992; Dalby & Williams, 1986; Kremen, Seidman, Faraone, & Tsuang, 2001; Weickert et al., 2000). A shortcoming of these methods is that they rely on estimates derived from post-illness performance or history of academic attainment, either of which may be affected by schizophrenia, resulting in downwardly biased estimates of premorbid ability. Premorbid indices that rely on family reports or home movies (Walker et al., 1990) may demonstrate early impairments dramatically, but family reports are subject to retrospective biases, and the movies were not originally collected in a systematic manner, thus enabling little insight into the longitudinal course of cognitive function.

Finally, it is possible to gather information about the premorbid functioning of people who are already diagnosed with schizophrenia. This is referred to as the "follow-back" strategy.

This approach, if focused on patients ascertained at the time of the first episode, may benefit from more representative sampling, without confounds introduced by genetic or behavioral high-risk strategies. A disadvantage shared with the large birth cohort or draft registry studies is that the available measures may be suboptimal. One study using this approach analyzed test scores from the Iowa Tests of Basic Skills and Iowa Tests of Educational Development acquired at grades 4, 8 and 11, in a group of patients who later went on to develop schizophrenia (Fuller et al., 2002). While there was no control group, the patients showed a pattern of increasing impairment from grades 8 to 11 and by grade 11 were significantly impaired relative to state norms. We report here application of similar methods, using the follow-back approach to characterize the trajectories of cognitive development among individuals ascertained during the first episode of schizophrenia, and relating these findings to more comprehensive cognitive characterization ~6 months following treatment for the initial episode of illness, in comparison to healthy, demographically matched volunteers from the community.

Methods

Participants

The overall methods for sample ascertainment, diagnostic assessment, symptom ratings, and neuropsychological assessment have been described elsewhere, as have other characteristics of brain structure and clinical course (Bilder et al., 1995; Bilder et al., 2000; Bogerts et al., 1993; Chakos et al., 1994; Chakos, Lieberman, Alvir, Bilder, & Ashtari, 1995; Degreef et al., 1992; Lieberman et al., 1992; Lieberman et al., 2001; Robinson et al., 1999a; Robinson et al., 1999b; Robinson et al., 2002; Snyder, Bilder, Wu, Bogerts, & Lieberman, 1995; Strous et al., 2004;

Szeszko et al., 1999; Szeszko et al., 2002; Szeszko, Bilder, Dunlop, Walder, & Lieberman, 1999). The present report describes results from 94 patients and 36 healthy volunteers described by Bilder et al. (2000). In brief, patients admitted to the inpatient service for a first episode of psychotic illness and who had less than 12 prior weeks of cumulative lifetime neuroleptic treatment were recruited. Patients satisfied Research Diagnostic Criteria (RDC)(Spitzer, Endicott, & Robins, 1977; Spitzer, Endicott, & Robins, 1978) for schizophrenia or schizoaffective disorder, based on structured interviews with the Schedule for Affective Disorders and Schizophrenia (SADS)(Endicott & Spitzer, 1978) and reviews of patients' histories. Patients with current or past serious neurological or endocrine disorder were excluded. After a complete description of the study, we obtained written informed consent. The healthy comparison group was recruited through local announcements and advertisements. These participants were selected to be similar to the patients on distributions of sex and age. They were free of RDC mental disorders and other major illnesses, as determined using the SADS Lifetime Version interview, physical examination, and urinalysis. None of the subjects had a current substance use disorder or a history of substance dependence, chronic neurological or medical illness, or drug treatment known to affect the brain.

Following consent to participate in the parent study, subjects were offered the opportunity to participate also in our "follow-back" study, which was explained to them, and if they agreed and provided informed consent for this, we then requested school records from institutions at which the participant indicated prior attendance. These procedures were approved by both the Hillside Hospital Division of North Shore – Long Island Jewish Medical Center, and the New York Board of Education, Research Office.

Facilitating our work was a New York State regulation requiring that academic data be

maintained for a period of 50 years following high school graduation or early termination from public school. Complicating our scientific goals, there was not uniformity of testing across schools for the time frame of interest to our study (which spanned several decades). Given the considerable heterogeneity of tests actually administered, each test score was entered as a separate record. We report here analyses of two different types of follow-back test data: (1) achievement test scores from grades 1-12; and (2) Scholastic Aptitude Test (SAT) scores.

Achievement Test Scores

We received transcripts from 59 of our 94 patients and 26 of our 36 healthy volunteers, yielding a total of 3,729 individual test scores (since each test score or grade reflected a unique record, there were approximately 44 records per individual) spanning a 45-year period from 1954 to 1999. Grade equivalent (GE) scores were available for 2,335 of these records. Despite attempts to gather information regarding normative samples for the different tests, publishers did not provide sufficient information for us to determine the equivalency of GE scores across tests. We therefore collapsed GE scores across tests, under the assumption that there was unlikely to be systematic bias in the scoring of tests administered to each group.

There were a substantial number of records with national percentile equivalency (NPE) scores (n = 2,146), and a sufficient number of cases had both GE and NPE scores (n=1,324) to assess their correlation. Multiple regression predicting GE scores from subjects' grade level and NPE scores yielded a multiple R of .93 ($F(2, 1321) = 4358$, $p < .001$). Thus we estimated GE scores for records that had NPE's but missing GE's, yielding a total of 3,157 records with actual or estimated GE scores. These scores came from a total of 63 different tests. The most frequently available scores came from the Stanford Achievement Test (454 records), Iowa Test of Basic Skills (396 records), the Metropolitan Achievement Test (308 records), Comprehensive

Test of Basic Skills (279 records), and California Achievement Test (277 records). All other tests contributed fewer than 150 records (or less than 5%) to the total number of test records. These scores were used in subsequent analyses.

Scholastic Aptitude Test Scores

The College Entrance Examination Board (CEEB) generates the most widely used tests for college-bound individuals in the United States, namely the Scholastic Aptitude Test (SAT) and the Preliminary Scholastic Aptitude Test (PSAT). Both SAT and PSAT scores are coded on a scale from 200 to 800. While initially conceived as a scale with mean = 500 and SD = 100, actual scores have deviated from this due to changes in population characteristics. Since our comparisons focused on differences between patients and our own healthy comparison group rather than the national normative sample, and the healthy volunteers were age-matched to the patients, we did not attempt to "correct" these scores for putative population changes over time. We had SAT and/or PSAT results from 39 patients and 24 Healthy volunteers.

We also include here selected results from cognitive testing conducted ~6 months following the initiation of treatment for the first episode of illness (when symptoms of the first episode had generally reached asymptotic levels of improvement), along with other clinical and historical measures as previously published (Bilder et al., 2000).

Our analysis plan aimed to answer several basic questions:

(1) Did patients and healthy volunteers differ in their academic achievement test scores? If there was a difference:

(a) How early could this difference be detected?

(b) Did the difference increase over the years preceding illness?

(2) Given that the SAT/PSAT, administered in the 11^{th} and 12^{th} grades, correlates highly with

other tests of ability, it may serve as a "proxy" measure of cognitive ability at that time. It is noteworthy also that the administration of the SAT/PSAT in 11th/12th grade usually precedes the onset of the first episode by several years. Thus we aimed to ask:

(a) Was there a difference between patients and healthy volunteers in SAT/PSAT scores? In other words, did these groups differ in ability, even considering that these are unique "college bound" subgroups of the original samples?

(b) if SAT/PSAT scores generate robust estimates of cognitive ability (Full Scale IQ) years later in healthy volunteers, do patients' post-onset scores deviate from levels predicted by their SAT/PSAT scores? In other words, do patients show deterioration in general cognitive ability between the time that they took the SAT/PSAT and the time of testing following the first episode of illness?

Results

Demographic Characteristics of Subsamples

Table 1 shows the basic demographic characteristics of patients and healthy volunteers in the "achievement testing" and "SAT/PSAT" sub-samples.

Insert Table 1 about here

Patients and healthy volunteers who had achievement test scores did not differ significantly ($p<.05$, two-tailed) in age, sex, or handedness but they did differ in ethnicity (with more non-white patients); education, parental social class, and IQ were also significantly lower among patients compared to healthy volunteers. The same pattern generally applied for the sub-sample of patients and healthy volunteers for whom we had SAT/PSAT scores, but in this subgroup the patients and healthy volunteers did not differ in education, which is highly unusual for samples of

patients and healthy individuals.

We examined subject characteristics for these sub-samples compared to other individuals in the larger sample (from Bilder et al., 2000) who did not contribute achievement test scores. The group contributing achievement test scores were similar on all demographic and clinical characteristics described above (see Table 1), except their education was higher ($t = 2.4$, $df = 92$, $p=.016$; specifically patients who did not provide achievement tests had mean ± SD years of education = 12.3 ± 2.5). The group contributing SAT/PSAT scores were also similar to other patients from the original sample in all respects except education ($t = 4.5$, $df = 92$, $p<.001$; non-SAT/PSAT patient had only 12.3 ± 2.1 years of education) and Full Scale IQ ($t = 3.1$, $df = 92$, $p = .003$; non-SAT/PSAT patients had IQ of only 82.9 ± 12.3).

Achievement Test Scores

The number of records available at each grade in the patient and healthy volunteer groups, along with descriptive statistics for these, are shown in **Table 2**. These grade equivalent (GE) scores were used as dependent variables in a two-way ANOVA with grade (1 through 12) and group (patient, Healthy volunteer) as independent variables. **Figure 1** illustrates the developmental trajectory in grade equivalent performance of patients and Healthy volunteers. The main effects of grade ($F(11,3133) = 227$, $p < .001$) and group ($F(1,3133) = 141$, $p < .001$) were both significant, but the group by grade interaction effect was non-significant ($F(11,291)=1.69$, $p = .07$).

Insert Table 2 and Figure 1 about here

Tests for the group by grade interaction effect in this ANOVA model might be overly conservative since each exam is considered independent of the others, when in fact there are clear relations within subjects across testing occasions. Further, the precision of estimates is poorer in

those grades where the number of observations is lower (e.g., for the 12th grade there were only 9 records for healthy volunteers). Our primary goal was to determine whether the overall increment in test scores as a function of grade might differ between groups. To do this, we applied curve fitting procedures, including linear, quadratic and growth curve analytic models (SPSS release 11.5.0) to each group (patients, healthy volunteers) separately, modeling grade equivalent score as a function of grade. For the patients, the linear component of the regression model using only grade explained 58% of variance in grade equivalent score, and the slope of the regression was .889 (standard error = 0.018; 95% confidence intervals for slope = .854 to .924). Adding a quadratic term did not contribute significantly to this model. For healthy volunteers the linear component of the regression model using grade explained 60% of variance in grade equivalent score, and the slope of the regression was .943 (standard error = 0.021; 95% confidence intervals for slope = .901 to .985). While the addition of quadratic term contributed significantly to this model for the healthy volunteers at $p < .05$ (by t-statistic), the additional term contributed less than 1% variance in the overall model and was not considered further (B = -0.015; standard error of B = .007). The linear regression slope was thus significantly steeper for healthy volunteers, suggesting an increasing discrepancy over grade. There was also a significant difference in the intercepts, with patients' first grade scores estimated at a grade equivalent score of ~1.3 and healthy volunteers' first grade scores estimated more than a grade higher at a grade equivalent score of ~2.4. The regression model estimates, given the difference in slopes, that the difference between groups in grade 1 (~1.1 grade equivalents) increased by grade 12 (to ~1.8 grade equivalents). This degree of increasing discrepancy (~ 0.7 grade equivalents) is generally concordant with our observed data, where the difference in grade equivalent scores increased from ~0.8 to ~1.5 grade equivalents from the 1st to the 12th grade (see Table 2).

Insert Table 3 about here

Scholastic Aptitude Test Scores

Analysis was also conducted for 39 patients and 24 healthy volunteers for whom we had results of the Scholastic Aptitude Test (SAT) examinations. Because this examination is administered selectively to "college-bound" students these participants would be expected have higher functioning than other members of our sample. The fact that we received these records for only 39 of 94 patients, compared to 24 of 36 healthy volunteers (chi-square=6.6, df = 1, p<.01 (two-sided)), suggests that fewer patients were considered college-bound and took these tests. As noted above, the sub-sample who provided SAT or PSAT scores also had higher education and Full Scale IQ compared to the remaining patients who did not provide records.

We obtained a total of 268 records (161 from patients, 107 from healthy volunteers) from grades 11 and 12. Patients were more likely to have records from grade 12 compared to healthy volunteers (chi-square = 7.6, df = 1, p=.006). ANOVA on these scores with group and grade revealed a significant main effect of group (F=23.2, df=1,263, p<.001), but the effect of grade (F=.08, df = 1,263), and the group by grade interaction (F=3.65, df=1,263, p=.057) were non-significant. Given the possible bias in test-taking behavior (with patients more likely to take the test again in grade 12), and the non-significant trend for the difference in scores to decrease in grade 12, we decided to examine both the mean and the maximum scores obtained over grades 11 and 12. While we found that the math scores were higher than verbal scores by about 50 points (estimated marginal mean for math = 485, 95% CI = 465 to 506; estimated marginal mean for verbal = 431, 95% CI = 410 to 451; main effect of subject: F=13.8, df=1,259, p<.001), there were no interactions of subject with group, grade, or group by grade (all F<1, df=1,259). We similarly found there was a difference between PSAT and SAT scores, with SAT scores higher

(estimated marginal mean for PSAT = 428, 95% CI = 404 to 451; estimated marginal mean for SAT = 476, 95% CI = 459 to 493; main effect of test type: F=10.7, df=1,263, p<.001), but there was no interaction with group (F=1.76, df=1,263, p=.185)(all PSAT's were given in grade 11 so there was no interaction term reflecting grade in this analysis).

We therefore collapsed across both subject (math, verbal) and test type (PSAT, SAT) for further analyses, and for simplicity refer to these averages as "mean SAT" scores. **Table 4** provides descriptive statistics for the mean SAT scores including scores separately at grades 11 and 12. Overall, however, it can be seen that there is an 80-point difference between patients and healthy volunteers, reflecting an effect size d = .67 (pooled SD = 120). Analyzing the mean and maximum scores for each subject corroborated these results. ANOVA with one score for each subject (see descriptive statistics in the bottom of Table 4) revealed significant group differences of 88 and 102 points (95% CI's for these difference scores: 33 to 142, and 43 to 162) for the mean and maximum SAT scores, respectively (F's = 10.3 and 11.8, df = 1,61, p's<.002). These results suggest that the actual discrepancy between patients and controls may be even larger (d=.78 for mean considering pooled SD of 113, d=.82 for maximum considering pooled SD of 125).

Insert Table 4 about here

Our next aim was to develop a regression equation for the healthy individuals, using their SAT scores from the 11^{th} and 12^{th} grades to "predict" their WAIS-R Full Scale IQ obtained approximately 8 years later. We examined regressions using both the mean and maximum SAT score. Since the correlation was slightly higher for the mean (r=.82) compared to the maximum (r=.77), we used the equation for the mean (FSIQ = 62.9 + [.097*SAT]). We then compared the FSIQ estimate based on the SAT to the obtained FSIQ in each group. These results are shown in

Table 5.

Insert Table 5 about here

Table 5 shows that patients and healthy volunteers differed only by approximately 8 points in "predicted IQ" (mirroring their differences in SAT scores on which these predictions were based), but that the observed IQ scores of patients were ~20 points lower than the healthy volunteers. Subtraction of observed from predicted IQ scores suggests a decline of ~11.5 points in Full Scale IQ between the time that the SAT's were administered, and the time of IQ testing ~6 months following the patients' initial enrollment and treatment in our study.

A possible concern about our findings with respect to "premorbid" functioning is that for some patients, clinical signs of illness may already have been detectable, and if so "illness" could have influenced our results. We used three different methods to examine possible early signs of illness, namely: (1) age at first signs of any significant behavior change noted by family members ("age of first symptoms"); (2) age at first signs of psychotic symptoms ("age of first psychotic symptoms"); and (3) age at which the first treatment was received for any behavior problem ("age of first psychiatric treatment"). Since our period of study included only the grades through high school (age 18), we examined the scores of all patients who had "early" onset (less than or equal to age 18) to determine if this may have impacted our findings. We found 13 cases had early symptom onset, 8 had early onset of psychotic symptoms, and 8 had early treatment. We compared these "early onset" cases to all other cases in our sample using t-tests on all the grade equivalent scores from grades 1 through 12, and on the mean and maximum SAT scores. The only significant effects ($p<.05$, two-tailed, uncorrected for multiple tests) detected in these 52 independent t-tests were observed on SAT scores for the effect of early treatment. Mean ± SD SAT score in the late vs. early treatment cases was 455 ± 110 vs. 379 ± 109 ($t=2.01$, $df = 61$, $p<$

.05)); corresponding values for maximum SAT were: 522 ± 122 vs. 413 ± 104 ($t = 2.40$, $df = 61$, $p < .02$).

Discussion

The results demonstrate that objective test scores obtained from the academic records of individuals who would later go on to develop schizophrenia were significantly lower than those of their peers who did not develop mental illness. These differences were apparent already in the first grade, which is usually the first time that individuals receive standardized tests. The effect does not appear to be subtle, with the difference of approximately 1 grade equivalent in the 1st grade likely reflecting approximately 1 standard deviation (SD) deficit. In comparison, the cognitive deficit in patients after onset of illness was -1.7 SD, as measured prospectively using the WAIS-R Full Scale IQ. To the extent that it is appropriate to generalize from these metrics, it would be estimated that roughly 60% of the global cognitive deficit in people who go on to develop schizophrenia are apparent by the first grade.

The near parallel curves in grade equivalent scores between patients and healthy volunteers suggest that the significant differences in achievement are maintained and may widen slightly over the years through high school. The gap between patients and healthy volunteers appeared to increase by approximately 0.7 grade equivalents from grade 1 to grade 12. Because the variance of scores increases in the more advanced grades, it is not clear whether this represents a larger discrepancy in the true ability difference between groups, or the same underlying difference in ability measured with different precision at this later period. The possibility that there is a genuine but subtle decline in functioning preceding the onset of overt psychotic symptoms is consistent both with our data, and the results of a similar study using

standardized tests in a follow-back design (Fuller et al., 2002). That study revealed a deficit relative to a state-wide population for individuals who would go on to develop schizophrenia, but this deficit was only statistically significant in the 11th grade, not earlier in their academic careers.

The analysis of PSAT and SAT scores suggested that the change in ability accompanying the onset of schizophrenia approximates 11.5 IQ-equivalent points, reflecting an effect size of approximately .77 SD. This apparent drop in cognitive ability needs to be considered in light of several factors. First, the sub-groups of both patients and healthy volunteers who take the PSAT and/or the SAT are "college-bound" students who on average have higher ability than those who do not take these tests. It is conceivable, and perhaps likely, that those with higher ability prior to the onset of psychosis have greater decrements in function accompanying the onset of schizophrenia. This suggestion was offered previously to help explain larger "deterioration" index scores among cases with higher estimated premorbid ability (Bilder, 1985; Bilder et al., 1988; Bilder et al., 1992). Second, it should be recognized that the patients' "current" IQ scores were obtained at a time when they had stabilized from the initial illness (i.e., approximately 6 months after the start of treatment for the first episode), but some residual effects of acute illness cannot be ruled out definitively. Our own follow-up studies suggest a high degree of stability and at best modest improvement in IQ scores (e.g., about 3 points) over subsequent years of treatment with conventional antipsychotics (Goldman et al., 1999), and the effects of newer antipsychotics may offer at best limited additional benefit (probably less than 0.5 SD in effect size terms, or less than 7 points in IQ-equivalent terms). These observations are sobering and suggest that new therapeutic approaches will likely be needed to reverse the combined effects of pre-onset cognitive deficits together with additional compromise that may accompany onset.

There are multiple limitations to the current findings. First, the test scores we used reflect

a complex mixture of ability and achievement measures, and while these tend to correlate highly with other measures of cognitive ability, inferences drawn from such scores must necessarily consider that we may be measuring different psychological constructs. Second, our data on grade equivalent scores was based on a broad diversity of tests, and there is no way to validate our assumption that the grade equivalent and national percentile measures derived from these tests are generally equivalent. Finally, there may be biases in our data reflecting selective attrition of records at different grade levels (for example, the increased frequency of missing data in later school years could reflect migration of students into "tracks" that differed in their requirements for tests using grade equivalent or national percentile scoring systems). Despite these limitations, our results fit well with existing literature, and further provide plausible estimates of the degree of cognitive deficit that is apparent as early as this can be measured, and the degree of deficit that may accompany the onset of overt symptoms of schizophrenia.

References

Bilder, R. M. (1985). *Subtyping in chronic schizophrenia: Clinical, neuropsychological, and structural indices of deterioration*. Ann Arbor, MI: University Microfilms.

Bilder, R. M., Bogerts, B., Ashtari, M., Wu, H., Alvir, J. Ma., Jody, D. et al. (1995). Anterior hippocampal volume reductions predict "frontal lobe" dysfunction in first episode schizophrenia. *Schizophrenia Research, 17,* 47-58.

Bilder, R. M., Degreef, G., Pandurangi, A. K., Rieder, R. O., Sackeim, H. A., & Mukherjee, S. (1988). Neuropsychological deterioration and CT-scan findings in chronic schizophrenia. *Schizophrenia Research, 1,* 37-45.

Bilder, R. M., Goldman, R. S., Robinson, D., Reiter, G., Bell, L., Bates, J. A. et al. (2000). Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates. *American Journal of Psychiatry, 157,* 549-559.

Bilder, R. M., Lipschutz-Broch, L., Reiter, G., Geisler, S. H., Mayerhoff, D. I., & Lieberman, J. A. (1992). Intellectual deficits in first-episode schizophrenia: Evidence for progressive deterioration. *Schizophrenia Research, 18,* 437-448.

Bogerts, B., Lieberman, J. A., Ashtari, M., Bilder, R. M., Degreef, G., Lerner, G. S. et al. (1993). Hippocampus-amygdala volumes and psychopathology in chronic schizophrenia. *Biological Psychiatry, 33,* 236-246.

Cannon, T. D., Bearden, C. E., Hollister, J. M., Rosso, I. M., Sanchez, L. E., & Hadley, T. (2000). Childhood cognitive functioning in schizophrenia patients and their unaffected siblings: a prospective cohort study. *Schizophrenia Bulletin, 26,* 379-393.

Chakos, M. H., Lieberman, J. A., Alvir, J., Bilder, R., & Ashtari, M. (1995). Caudate nuclei volumes in schizophrenic patients treated with typical antipsychotics or clozapine. *Lancet, 345,* 456.

Chakos, M. H., Lieberman, J. A., Bilder, R. M., Borenstein, M., Lerner, G., Bogerts, B. et al. (1994). Increase in caudate nuclei volumes of first-episode schizophrenic patients taking antipsychotic drugs. *Am.J.Psychiatry, 151,* 1430-1436.

Cornblatt, B., Obuchowski, M., Roberts, S., Pollack, S., & Erlenmeyer-Kimling, L. (1999). Cognitive and behavioral precursors of schizophrenia. *Dev.Psychopathol., 11,* 487-508.

Dalby, J. T. & Williams, R. (1986). Preserved reading and spelling ability in psychotic disorders. *Psychol Med, 16,* 171-175.

Davidson, M., Reichenberg, A., Rabinowitz, J., Weiser, M., Kaplan, Z., & Mark, M.

(1999). Behavioral and intellectual markers for schizophrenia in apparently healthy male adolescents. *American Journal of Psychiatry, 156,* 1328-1335.

Degreef, G., Ashtari, M., Bogerts, B., Bilder, R. M., Jody, D. N., Alvir, J. Ma. J. et al. (1992). Volumes of ventricular system subdivisions measured from magnetic resonance images in first-episode schizophrenic patients. *Arch.Gen.Psychiatry, 49,* 531-537.

Endicott, J. & Spitzer, R. L. (1978). Schedule for affective disorders and schizophrenia. *Arch.Gen.Psychiatry, 35,* 837-844.

Erlenmeyer-Kimling, L., Adamo, U. H., Rock, D., Roberts, S. A., Bassett, A. S., Squires-Wheeler, E. et al. (1997). The New York High-Risk Project. Prevalence and comorbidity of axis I disorders in offspring of schizophrenic parents at 25-year follow-up. *Arch.Gen.Psychiatry, 54,* 1096-1102.

Erlenmeyer-Kimling, L., Rock, D., Roberts, S. A., Janal, M., Kestenbaum, C., Cornblatt, B. et al. (2000). Attention, memory, and motor skills as childhood predictors of schizophrenia-related psychoses: the New York High-Risk Project. *Am.J.Psychiatry, 157,* 1416-1422.

Erlenmeyer-Kimling, L., Squires-Wheeler, E., Adamo, U. H., Bassett, A. S., Cornblatt, B. A., Kestenbaum, C. J. et al. (1995). The New York High-Risk Project. Psychoses and cluster A personality disorders in offspring of schizophrenic parents at 23 years of follow-up. *Arch.Gen.Psychiatry, 52,* 857-865.

Fuller, R., Nopoulos, P., Arndt, S., O'Leary, D., Ho, B. C., & Andreasen, N. C. (2002). Longitudinal assessment of premorbid cognitive functioning in patients with schizophrenia through examination of standardized scholastic test performance. *American Journal of Psychiatry, 159,* 1183-1189.

Goldman, R. S., Conley, J., Bates, J., Reiter, G., Robinson, D., Bilder, R. et al. (1999). Stability of deficits in executive function and memory in first-episode schizophrenia. *Biological Psychiatry, 45,* 38S.

Kremen, W. S., Seidman, L. J., Faraone, S. V., & Tsuang, M. T. (2001). Intelligence quotient and neuropsychological profiles in patients with schizophrenia and in normal volunteers. *Biological Psychiatry, 50,* 453-462.

Lieberman, J. A., Alvir, J. M. J., Woerner, M., Degreef, G., Bilder, R. M., Ashtari, M. et al. (1992). Prospective study of psychobiology in first-episode schizophrenia at Hillside hospital. *Schizophrenia Bulletin, 18,* 351-371.

Lieberman, J. A., Chakos, M., Gerig, G., Wu, H., Alvir, J., Hoffman, E. et al. (2001). Longitudinal treatment-related changes to cortex in schizophrenia. *Biological Psychiatry, 49,* 75S-76S.

Reichenberg, A., Weiser, M., Rabinowitz, J., Caspi, A., Schmeidler, J., Mark, M. et al. (2002). A population-based cohort study of premorbid intellectual, language, and behavioral

functioning in patients with schizophrenia, schizoaffective disorder, and nonpsychotic bipolar disorder. *Am J Psychiatry, 159,* 2027-2035.

Robinson, D., Woerner, M. G., Alvir, J. M., Bilder, R., Goldman, R., Geisler, S. et al. (1999a). Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Arch.Gen.Psychiatry, 56,* 241-247.

Robinson, D. G., Woerner, M. G., Alvir, J. M., Bilder, R. M., Hinrichsen, G. A., & Lieberman, J. A. (2002). Predictors of medication discontinuation by patients with first-episode schizophrenia and schizoaffective disorder. *Schizophr.Res., 57,* 209-219.

Robinson, D. G., Woerner, M. G., Alvir, J. M. J., Geisler, S., Koreen, A., Sheitman, B. et al. (1999b). Predictors of treatment response from a first episode of schizophrenia or schizoaffective disorder. *American Journal of Psychiatry, 156,* 544-549.

Snyder, P. J., Bilder, R. M., Wu, H., Bogerts, B., & Lieberman, J. A. (1995). Cerebellar volume asymmetries are related to handedness: a quantitative MRI study. *Neuropsychologia, 33,* 407-419.

Spitzer, R. L., Endicott, J., & Robins, E. (1977). *Research Diagnostic Criteria (RDC) for a Selected Group of Functional Disorders*. New York: New York Biometrics Research Division.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria: Rationale and reliability. *Arch.Gen.Psychiatry, 35,* 773-782.

Strous, R. D., Alvir, J. M., Robinson, D., Gal, G., Sheitman, B., Chakos, M. et al. (2004). Premorbid functioning in schizophrenia: relation to baseline symptoms, treatment response, and medication side effects. *Schizophrenia Bulletin, 30,* 265-278.

Szeszko, P. R., Bilder, R. M., Dunlop, J. A., Walder, D. J., & Lieberman, J. A. (1999). Longitudinal assessment of methylphenidate effects on oral word production and symptoms in first-episode schizophrenia at acute and stabilized phases. *Biological Psychiatry, 45,* 680-686.

Szeszko, P. R., Bilder, R. M., Lencz, T., Pollack, S., Alvir, J. M., Ashtari, M. et al. (1999). Investigation of frontal lobe subregions in first-episode schizophrenia. *Psychiatry Res., 90,* 1-15.

Szeszko, P. R., Strous, R. D., Goldman, R. S., Ashtari, M., Knuth, K. H., Lieberman, J. A. et al. (2002). Neuropsychological correlates of hippocampal volumes in patients experiencing a first episode of schizophrenia. *American Journal of Psychiatry, 159,* 217-226.

Walker, E. F., Diforio, D., & Baum, K. (1999). Developmental neuropathology and the precursors of schizophrenia. *Acta Psychiatr.Scand.Suppl., 395,* 12-19.

Walker, E. F. & Levine, R. J. (1990). Prediction of adult-onset schizophrenia from childhood home movies of the patients. *American Journal of Psychiatry, 147,* 1052-1056.

Weickert, T. W., Goldberg, T. E., Gold, J. M., Bigelow, L. B., Egan, M. F., & Weinberger, D. R. (2000). Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect. *Arch.Gen Psychiatry, 57,* 907-913.

Table 1. Characteristics of patients and healthy volunteers				
	Achievement		**SAT/PSAT**	
	Patients	**Healthy volunteers**	**Patients**	**Healthy volunteers**
N	59	26	39	24
Sex (M, F)	36, 23	17, 9	22, 17	16, 8
Hand preference	38, 21	21, 5	29, 10	20, 4
Age at time of NP exam (years)	25.5 ± 6.1	24.3 ± 6.5	26.5 ± 6.7	23.1 ± 5.1
Ethnicity (White, African-American, Hispanic, Asian, Other)	29, 20, 7, 3, 0	22, 1, 0, 1, 2[1]	24, 10, 2, 3, 0	20, 1, 0, 1, 2[5]
Education (years)	13.5 ± 2.0	14.7 ± 1.5[2]	14.2 ± 2.0	14.8 ± 1.6
Parental Social Class	3.2 ± 1.3	2.4 ± 1.0[3]	3.0 ± 1.4	2.3 ± 0.9[6]
WAIS-R FSIQ	86.6 ± 13.9	110.4 ± 14.0[4]	91.2 ± 13.8	111.4 ± 13.9[7]
Age at first behavior change noted by family	21.5 ± 6.4	NA	22.6 ± 6.5	NA
Age at first psychotic symptoms	23.0 ± 6.5	NA	24.3 ± 6.5	NA
Age at first treatment for psychiatric illness	23.2 ± 5.8	NA	23.7 ± 6.3	NA
RDC Diagnosis (Schizophrenia, Schizoaffective)	46, 13	NA	29, 10	NA

Note. Unless indicated otherwise, values are mean ± standard deviation.
[1]Patients and Healthy volunteers differ: chi-square = 18.1, df=4, p=.001
[2]Patients and Healthy volunteers differ: t = 2.8, df = 83, p=.006
[3]Patients and Healthy volunteers differ: t = 2.6, df = 81, p=.01
[4]Patients and Healthy volunteers differ: t = 7.2, df = 83, p=.001
[5]Patients and Healthy volunteers differ: chi-square = 9.7, df=4, p=.046
[6]Patients and Healthy volunteers differ: t = 2.1, df=59, p=.036
[7]Patients and Healthy volunteers differ: t = 5.6, df=61, p=.001

Table 2. Regression models for patients and healthy volunteers, predicting grade equivalent score from grade				
	Patient		**Healthy volunteer**	
	Parameter	95% Confidence Intervals (lower, upper)	Parameter	95% Confidence Intervals (lower, upper)
Regression Slope	.889	.854, .924	.943	.901, .985
Intercept	1.276	1.037, 1.514	2.392	2.133, 2.651

Table 3. Descriptive statistics for mean Scholastic Aptitude Test scores						
Grade	Group	N of records	Mean	SD	95% Confidence Intervals	
					Lower	Upper
11	patient	89	405	106	382	429
	healthy volunteer	76	507	124	481	532
12	patient	72	438	102	412	464
	healthy volunteer	30	482	126	442	523
11 and 12	patient	161	420	106	402	436
	healthy volunteer	106	500	125	478	522
		N of Subjects				
Mean SAT	Patient	39	411	97	377	445
	Healthy volunteer	24	499	117	456	542
Max SAT	Patient	39	469	106	432	505
	Healthy volunteer	24	571	128	524	618

Table 5. FSIQ as predicted from SAT scores in grades 11 and 12 compared to actual FSIQ					
	Patient	**Healthy volunteer**	**t-test**	**DF**	**P (two-tailed)**
Predicted FSIQ	103 ± 9.5	111 ± 11.4	3.2	61	.002
Actual FSIQ	91.2 ± 13.8	111 ± 13.9	5.6	61	.001
"Drop" (predicted – actual)	11.5 ± 9.9	-0.1 ± 7.9	4.9	61	.001

Figure 1. **Grade Equivalent Scores By Grade**. Grade equivalent scores for patients (triangles) and healthy volunteers (squares) are plotted as a function of grade for grades 1 through 12; error bars represent 95% confidence intervals around the mean at each grade for each group. The lines are linear regression functions of the data series for these groups. While the curves are nearly parallel, the slope is significantly steeper for healthy volunteers relative to patients, reflecting greater improvement in grade equivalent scores over successive grades (see text for details).