Set-Shifting Ability and Schizophrenia: A Marker of Clinical Illness or an Intermediate Phenotype?

Alan E. Ceaser, Terry E. Goldberg, Michael F. Egan, Robert P. McMahon, Daniel R. Weinberger, and James M. Gold

Background: Impairments of executive functioning, such as set-shifting ability, are seen as core deficits of schizophrenia and are of interest as candidate intermediate phenotype markers. The Intradimensional/Extradimensional (ID/ED) shift task offers a differentiated assessment of shifting from previously reinforced stimuli as well as shifting from previously reinforced features and has proven to be sensitive to the impairment seen in patients with schizophrenia.

Methods: We examined ID/ED performance in 147 patients with schizophrenia, 131 of their healthy siblings, and 303 healthy control subjects. Participants were recruited from local and national sources as volunteers for the Clinical Brain Disorders Branch/National Institute of Mental Health “sibling study”.

Results: Nearly all control subjects (87%) finished the task successfully, as did 80% of siblings. In contrast only 54% of patients with schizophrenia were able to complete the task. Despite the apparent similarity of performance across the sibling and healthy comparison group, the two groups differed significantly in terms of the number of stages until failure. This difference, however, was not present at any particular stage or any other measure of performance.

Conclusions: Patients demonstrated robust ID/ED deficits. However, their siblings were minimally impaired, and this impairment did not seem to run in families. These results suggest that impairments on attentional set shifting assessed by ID/ED task are strongly associated with clinical illness, but these impairments are not a promising intermediate phenotype.

Key Words: CANTAB, cognition, executive function, intermediate phenotype, schizophrenia, set-shifting, working memory

Cognitive impairments are attractive intermediate phenotypes for the study of schizophrenia risk genes. There is robust evidence that cognitive impairment is a core feature of the disorder, and it seems likely that the genetic architecture of specific cognitive functions might be simpler than the clinical illness phenotype. Furthermore, there is evidence that the healthy relatives of schizophrenia patients also demonstrate cognitive impairments in many of the same domains observed in ill patients. Specifically, a recent meta-analysis comparing cognitive performance of healthy relatives of schizophrenic patients with that of well control subjects revealed that four of the six largest effect sizes “come from variables having in common executive control functions such as working memory demands, set shifting, and the inhibition of prepotent responses” (1), suggesting that these aspects of cognitive performance might be particularly useful for genetic studies.

Set shifting is a complex construct. In the cognitive experimental literature this construct is often operationalized in the study of task switching (2,3). In this approach, subjects need to shift between performing two tasks on the basis of cues. For example, a cue would inform the subject to make a prosaccade to a target location on some trials, whereas another cue would indicate that an antisaccade is the appropriate responses on others. Performance is typically quantified in terms of error rates on switch as opposed to non-switch trials or reaction time costs on switch trials relative to non-switch trials. Surprisingly, patients with schizophrenia have not demonstrated reliable deficits in several studies examining task switching (4–7), although further study is needed to establish confidence in these results. In the clinical neuropsychological literature, the Wisconsin Card Sorting Test (WCST) is considered a measure of set shifting, and there is well-replicated evidence that patients with schizophrenia demonstrate marked impairments on this task. However, studies of WCST performance in well siblings of patients with schizophrenia have produced mixed results (8–11), and the effect sizes seen on the WCST in the Snitz meta-analysis were not among the largest documented in the cognitive literature. In addition, Kremen et al. (12) recently reported no evidence of heritability of WCST performance in a study of monozygotic and dizygotic twins. These converging data suggest that the WCST might not be a robust intermediate phenotype marker, despite being sensitive to the impairments observed in patients. The cognitive complexity of the task might enhance sensitivity to diagnostic status, because there might be multiple routes to performance failure, although at the same time making it difficult to discern specific genetic effects that impact only some of the cognitive operations involved in the task.

Experimental methods that isolate the different cognitive mechanisms involved in WCST performance might serve to increase power to detect genetic effects. The Intradimensional/Extradimensional (ID/ED) set-shifting task (Downes et al. [13]) was designed to offer a differentiated assessment of many of the processes involved in the WCST. Like the WCST, the ID/ED task requires the subject to learn which stimulus/response choice is correct through the use of feedback. However, the ID/ED is simpler than the WCST, because only two stimuli are presented in every trial, making the feedback less ambiguous—if one stimulus is incorrect, the other must be correct. Thus, the ID/ED task minimizes the role of rule “discovery” that is a demand of the WCST. Studies Of ID/ED performance have consistently found

From the Clinical Brain Disorders Branch (AEC, TEG, MFE, DRW), National Institute of Mental Health, Bethesda; and the University of Maryland School of Medicine (RPM, JMG). Maryland Psychiatric Research Center, Baltimore, Maryland.

Address reprint requests to Daniel Weinberger, M.D., 10 Center Drive, MSC 1379, CRC Bldg. 10, Bethesda, MD 20892-1379; E-mail: weinberd@mail.nih.gov.

Received October 12, 2007; revised May 9, 2008; accepted May 12, 2008.

0006-3223/08/$34.00

doi:10.1016/j.biopsych.2008.05.009

BIOL PSYCHIATRY 2008;64:782–788
Published by Society of Biological Psychiatry
impairments in patients with chronic schizophrenia at the extradimensional stages (14–16), including patients with preserved IQ, with a smaller numbers of patients failing at some of the earlier stages implicating deficits in more elementary processes.

In contrast to chronically ill patients, first-episode patients have been shown to be relatively unimpaired on the ID/ED task (17–19). In fact, Joyce et al. (20,21) and Pantelis et al. (22) have reported that the performance of first-episode patients seems to deteriorate over time on the ID/ED task. Furthermore, there is mixed evidence suggesting that illness duration might correlate with ID/ED performance (23) (but see Tyson et al. [16]). These findings raise the possibility that poor performance on this task is a result of illness chronicity and/or treatment effects but not risk-associated influences.

We thought it was potentially informative to examine ID/ED task performance in patients and in their healthy siblings to determine whether attentional set shifting was a promising intermediate phenotype. In contrast, evidence of intact performance in siblings, coupled with the evidence of relatively intact performance in first-episode patients, would suggest that impairments on the ID/ED task reflect “extra-genetic” influences that are associated with the clinical illness.

Methods and Materials

Subjects

Intradimensional/Extradimensional task performance data were available from 147 patients, 131 unaffected siblings, and 303 healthy comparison subjects who were volunteers for the Clinical Brain Disorders Branch/National Institute of Mental Health (NIMH) “sibling study.” All study procedures were approved by the Institutional Review Board of the NIMH, and all subjects provided written informed consent after complete description of the study. All participants (age 18–60 years) were interviewed by a research psychiatrist (blind to group status) with the Structured Clinical Interview from DSM-IV Axis I disorders (SCID-I) (24) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II) (25). A second and sometimes third psychiatrist independently reviewed all diagnostic data. All subjects received a neurological examination, magnetic resonance imaging scan of the brain, blood laboratory tests, and review of medical history to rule out medical problems that might impact cognitive performance. Participants with alcohol or drug dependence within the last year or more than a 5-year history of alcohol or drug abuse or dependence were excluded. Comparison subjects and siblings who were diagnosed with a current Axis I or Axis II disorders, a history of psychosis, or a history of serious medical/neurological illness that could affect cognitive function were excluded. Participants were required to be free of a history of learning disability and must have been a fluent English speaker by the age of 5 years old. Participants with Wide Range Achievement Test—Revised (WRAT-R) reading scores of < 70 were excluded, to exclude subjects with severe developmental compromise. Healthy volunteers had the additional requirement that they not have a first-degree relative with schizophrenia.

Patients had a mean duration of illness of 12.4 (9.47) years. In some cases, siblings or patients were excluded due to one of the medical or psychiatric exclusions, missing Cambridge Neuropsychological Test Automated Battery (CANTAB) data, or most often, because siblings did not complete the protocol. Of the 147 patients with available CANTAB data, we were able to examine data from one or more siblings for 60 patients. Our sample of patients included 60 with siblings and 87 without; our sample of siblings consisted of 92 siblings with a proband and 39 without. Exclusion factors for the probands of the 39 siblings varied: 17 met medical exclusion criteria (i.e., the proband had a diagnosis of schizophrenia but had an additional diagnosis that complicated interpretation of cognitive data); 8 failed to complete the study but were considered to have a reliable schizophrenia diagnosis; and 14 completed the CANTAB, whereas their ill family member did not. Estimated full-scale IQ from the Wechsler Adult Intelligence Scale, revised edition (WAIS-R) and reading ability with the WRAT-R, a measure of premorbid IQ, (26) are reported for descriptive purposes. The demographic features of the study groups are shown in Table 1.

Neuropsychological Testing

The ID/ED task is a computerized task that consists of nine stages. Stimuli were displayed in outlined boxes to the left and right of center screen as well as centered top and bottom. The boxes served to outline the response area (Figure 1). Failure at a stage occurs if passing criterion, six consecutive correct responses, is not met by the 50th trial, at which point the tests conclude. Performance measures include the percentage of subjects successfully completing each stage, total errors committed at each stage, and the number of stages passed (27). Representative stimuli for stages can be seen in Figure 2.

At the first simple discrimination stage of the test, subjects see two shapes and must determine which one is correct. After making six correct responses in a row, the rule changes so that the other stimulus is correct. This reversal stage assesses a very elementary form of set shifting: can the subject shift from a previously rewarded response choice to a new choice in the face of negative feedback? At the compound discrimination phase,
the subject must again determine which is the correct shape stimulus but is now challenged by the presence of irrelevant line segments that are near the shapes. The second compound discrimination stage consists of a line element superimposed over the shape stimulus, making ignoring the distractor line element more difficult. Thus, the subject must focus attention on the relevant stimuli, and after six correct selections, the reinforcement contingencies reverse. The intradimensional shift stage, the sixth, asks that subjects apply previously learned rules to new shape stimuli. Thus, subjects must learn that a new shape will be rewarded in this stage, an example of an intradimensional shift. Upon successful completion this stage is followed by a reversal stage. At the critical extradimensional shifting stage, subjects need to shift their attention to the previously irrelevant line stimuli and ignore the previously reinforced shapes, precisely the same kind of shifting assessed on the WCST (13,28,29).

Data Analysis
Statistical analysis of demographic data was performed with STATISTICA 7 (version 7.1. 2005, Statsoft, Inc., Tulsa, Oklahoma) and is shown in Table 1 for descriptive purposes. Age differences were analyzed with one-way analyses of variance (ANOVA). Gender ratios were analyzed with $\chi^2$ tests. Both ANOVA and correlations were used to assess the influence of current IQ on ID/ED performance.

Our first analysis included all available patients, siblings, and control subjects with CANTAB data. Outcomes on the CANTAB test were divided into 10 possible ordered categories, on the basis of the stage at which a participant failed the test, with those passing all stages ranked into the highest category (“stage 10”). The SAS PROC FREQ (SAS, Cary, North Carolina) was used to compute Cochran-Mantel-Haenszel (CMH) test statistic (30) for pairwise comparisons of the distribution of stage attained in patients, siblings, and control subjects. This statistic tests the null hypothesis that $Pr(y > x) = .5$, where $x$ and $y$ are the stages at which randomly chosen members of group $x$ and group $y$ (e.g., control subjects and siblings) end the test. The alternative tested is that one group has a higher probability of reaching a later stage than the other. Because few failures occur at earlier stages, the $p$ value was computed from the exact (permutation) distribution of this statistic.

Fisher exact test was used to perform pairwise group comparisons on the stage-specific conditional failure probability (the probability of failing at a given stage, having passed all preceding stages). Westfall and Young’s step-down bootstrap (31) method was used to maintain the Type I error rate at $\alpha = .05$ while...

**Figure 1.** A screenshot of a typical trial from the Compound Discrimination stage.

**Figure 2.** Cambridge Neuropsychological Test Automated Battery stages without reversal stages.

www.sobp.org/journal
performing multiple comparisons (9 stages × 3 sets of pairwise comparisons).

The distribution of errors on the CANTAB battery is highly non-normal. Accordingly, pairwise comparisons of error rates among groups were conducted with the Wilcoxon rank sum test. The p values were adjusted for multiple comparisons with Benjamin and Hochberg’s (32) method for controlling false discovery rates.

Results

Demographic Features
A one-way ANOVA revealed age differences between the three groups [F(2, 578) = 8.68, p = .0002]; post hoc analysis revealed a significant difference between siblings and control subjects. Previous normative studies of ID/ED shift performance have reported that aging effects only emerge after the age of 50 (33,34), suggesting that the mean 4-year age difference between young adult control subjects and siblings was unlikely to impact our results. Correlational analyses revealed that age was not significantly correlated with performance in any diagnostic group.

Gender ratios differed significantly between patients and control subjects [χ² = 35.17, p < .0001] and between patients and siblings [χ² = 25.90, p < .0001]. There was no significant difference in gender ratio observed between siblings and control subjects. Gender did not impact performance in any of our subject groups and was not considered in further analyses.

CANTAB Results
The cumulative percentage of subjects in each group passing the nine stages of the ID/ED task is shown in Figure 3. Nearly all control subjects (87%) finished the task successfully, as did 80% of siblings. In contrast only 54% of patients with schizophrenia were able to complete the task. There was a highly significant difference in the distribution of the number of stages to failure among all three groups (MH = 25.13, p < .0001). This overall analysis was followed by a series of contrasts between pairs of groups. There were also significant group differences in the overall distribution of stages to failure between siblings and control subjects (MH = 8.16, p = .004), between patients and control subjects (MH = 60.84, p < .0001), and between siblings and patients (MH = 14.12, p = .0002). Despite the apparent visual similarity of performance across the sibling and healthy comparison group, the two groups differed significantly in this measure of overall task success. Patients performed significantly worse than both the sibling and the healthy comparison group.

To explore this sibling difference from control subjects, we examined the attrition rates at each stage with Fisher exact tests. None of the stage comparisons were significant. In contrast, patients differed from control subjects at several stages, including simple reversal (SR) (p = .01), compound discrimination (CD) (p < .001), extradimensional shift (EDS) (p < .001), and extradimensional reversal (EDR) (p = .004). Thus, the patient impairment is not limited to the EDS stage; but also includes rather simple reversal learning and the ability to ignore irrelevant distractors. Relative to the sibling group, significant patient impairment was similarly observed at the CD (p = .01), EDS (p = .01), and EDR (p = .04) stages.

We examined error scores at each stage, thinking this might prove to be a more sensitive measure than the simple stage pass/fail data. Post hoc pairwise tests at each stage showed that patients had significantly higher mean errors than control subjects at every stage except SD (p = .28) and IDS (p = .07). Mean errors for siblings were not significantly different from control means at any stage but were different from patient means at every stage except SD (p = .62), CD2 (p = .34), and IDS (p = .22). Note that analysis of error scores on the ID/ED task is limited to those subjects who attempted that stage. Thus, subjects who have already “failed out” of the task are not included in the analysis, perhaps limiting power to detect differences.

Given these largely negative results in the total group, we reasoned that impairment might occur on a familial basis in too small a subset of siblings to be detected in the overall sample. We therefore examined healthy sibling performance as a function of how their affected sibling performed. To do so, we split the patients into three groups: 1) those who completed all stages (PASSALL); 2) those who failed at stage EDS or at the subsequent reversal stage (EDS/EDR); and 3) those patients who were most severely impaired, failing before EDS (EARLYFAIL). Thus, if impaired task performance was indeed familial, we would expect to find more evidence of impairment in the siblings of patients in group 2 and 3 and little evidence of impairment from group 1 siblings. The siblings of patients in the EARLYFAIL group (25) were compared with siblings of the EDS/EDR group (16) and siblings of the PASSALL group (51). In Figure 4, the data strongly contradicted this expectation: siblings of patients who completed all stages had the lowest overall rate (78%) of
successfully completing the task. In contrast, 8% of the siblings of the most impaired patients completed all stages, whereas 87.5% of siblings of patients who failed at or after the EDS stage completed all stages. Thus, the siblings of patients who performed the best seemed to perform the worst. These results are inconsistent with the hypothesized familial effect.

**Exploratory Analysis**

Because there is suggestive evidence that first-episode patients are relatively unimpaired on the ID/ED task and that length of illness might correlate with the extent of impairment in chronic samples, we examined duration of illness effects and found no correlation between overall illness duration and any ID/ED task performance variable in our total patient group. We then compared performance between patients ill < 2 years (mean 1.2 [7], n = 16) with the rest of the patient group (mean 13.9 [9], n = 122) and found no statistically significant differences: 75.6% of the under-2-years-of-illness group completed all task stages, compared with 73.6% of the remaining patients.

To examine the role of general cognitive ability, we correlated estimated full-scale IQ with the number of completed stages and total errors on the ID/ED. In patients, these correlations were: r = .34, p < .05; r = -.44, p < .05; in siblings r = .23, p < .05; r = -.2, p < .05; and in control subjects r = .2, p < .05; r = .23, p < .05, respectively, for stages completed and total errors.

Patients were split into a low-IQ group (IQ < 85, n = 31), a middle-IQ group (IQs = 85–99, n = 68), and a high-IQ group (IQs > 100, n = 47) to explore how IQ influences patient performance. Over 80% of patients in the high-IQ group (Figure 5) passed all stages. Just fewer than 60% of the middle group passed all stages, and only 30% of the low-IQ group was able to complete the task [F(2,144) = 7.65, p < .01]. The low-IQ group performed significantly worse than the other two groups (p < .01). No significant differences were observed between the middle- and high-IQ groups, although the general trend is clear in Figure 5. Overall significant differences in the number of errors committed [F(92,144) = 12.9, p < .01] were also observed among all groups.

**Discussion**

Several key findings emerged from this analysis of ID/ED performance in patients, siblings, and healthy volunteers. First, roughly one-half of our patients were unable to complete the task, indicating a general impairment of set-shifting ability, consistent with prior patient studies (14–16). Our patients demonstrated both the general level and pattern of impairment documented in the literature, with marked difficulties at the compound discrimination, extradimensional shifting, and reversal phases (14,35). This pattern of results suggests that patients with schizophrenia have impairments in the ability to ignore distractors, inhibit prepotent responses, and shift attention from a previously reinforced stimuli dimension to a previously irrelevant dimension. The fact that our patients demonstrate the same type of impairments documented in other samples of chronically ill patients is important in establishing the potential value of examining the performance of their siblings.

Second, we found only subtle evidence for impairment in siblings. Overall, significantly fewer healthy siblings were capable of completing the task compared with control subjects. However, there were no statistically significant differences between siblings and control subjects on the number of errors committed at any stage or evidence of impairment among siblings at a specific stage of the task. Thus, siblings demonstrated no specific set-shifting impairment, and inspection of Figure 3 suggests that the siblings, as a group, performed slightly worse than control subjects, with the subtle difference in cumulative failure rates emerging by stage CD2; thereafter, failure rates at successive stages were closely parallel in the two groups.

The impairment observed in siblings did not seem to be familial: when we examined sibling performance on the basis of how their ill relative performed, we saw little indication of shared impairment. To the contrary, siblings of patients who performed best did worse. These results argue against the role of shared genetic effects impacting ID/ED performance in siblings. The fact that we have found little evidence that performance on the ID/ED task runs in families, coupled with recent evidence that WCST performance might not be heritable, suggests that some aspects of executive function impairment in schizophrenia might, surprisingly, largely implicate non-genetic factors. That is, certain aspects of executive impairment might mark the clinical phenotype, not genetically mediated risk. The success of genetic studies of cognitive intermediate phenotypes is critically dependent on accurate identification and measurement of cognitive deficits that are shared within families, versus those impairments that are associated with clinical illness. At a minimum, we believe our data suggest that the ID/ED task has limited utility as an intermediate phenotype marker.

Is it possible that siblings do not, in fact, have deficits in set shifting, given prior claims in the literature? Clearly, one cannot prove the null hypothesis, and the ID/ED test might offer a limited assessment of set shifting. However, this possibility is worthy of careful study. The two clinical tests often cited as providing evidence of set-shifting impairments, Trailmaking B and the WCST, involve multiple component operations beyond set shifting that might be responsible for the observed impairments. For example, impairments on Trailmaking could result from deficits in processing speed. Sibling impairments have been documented on processing speed measures that do not involve set shifting, such as verbal fluency and digit symbol, bolstering this possibility. Similarly, the WCST impairment could result from limitations in working memory (which have been well documented in the family literature (35,36), the ability to process error information, or in reasoning rather than in set shifting per se. The fact that recent carefully controlled studies of task switching in schizophrenia patients have not produced evidence of set-

---

**Figure 5.** A stage comparison of the percentage of subjects passing by their respective IQ group. The high-IQ patient group did not differ significantly from the control group on any performance measure. Abbreviations as in Figure 3.
shifting impairments adds weight to the possibility that sibling impairments on the WCST and Trailmaking might not be the result of a specific deficit in set shifting. Additional studies examining different aspects of set shifting in patients are needed to guide future family research.

Given the suggestive evidence that first-episode patients demonstrate minimal impairment on the ID/ED, the fact that siblings are unimpaired should be expected. Indeed, the rates of task success in several first-episode samples (75%–80% [16–19]) well surpass our patient group and are closer to the level seen in our control and sibling groups. If patients are unimpaired or minimally impaired at the first onset of illness, it would be surprising if deficits could be detected among well siblings.

We, like Tyson et al. (16), were unable to replicate evidence that increasing duration of illness among chronic patients was associated with increasing degree of impairment as reported by Pantelis et al. (23). Although the evidence suggesting progression of ID/ED impairment from first-episode to chronic illness is based on only a few small studies, the idea that executive impairment is something that emerges over time as a consequence of illness is worthy of further careful study, given the clinical importance of this question.

Indirectly arguing against an illness-specific progressive basis of impairment is the relationship with IQ (which does not progressively decline in schizophrenia [37]). In short, low-IQ patients perform extremely poorly, whereas our average-IQ group performed very similarly to control subjects. It is hard to reconcile this powerful effect of IQ with the suggestive evidence of illness-related deterioration in ID/ED performance, suggesting that additional longitudinal studies examining both IQ and ID/ED performance are needed to address the important question of whether the illness involves progressive compromise of attentional set shifting.

It is possible that some of the impairment in patients is secondary to the impact of antipsychotic medications (38). The fact that we did not observe any impact of duration of illness (and therefore duration of treatment) on task performance suggests that total medication exposure is not likely to be implicated. Available evidence from acute challenge studies has been mixed to date, with some but not all studies reporting mild impairments (39–42). Thus, it seems unlikely that the dense deficit seen in patients can be fully attributed to medication effects, although further study of this issue is clearly needed.

This study has a number of limitations. Like the WCST, the ID/ED task has psychometric limitations. With more than 80% of siblings and control subjects passing the task, it is possible that limited variability in task performance produced a false negative result. That is, this task might not be sufficiently sensitive to detect subtle (and possibly heritable) forms of attentional set-shifting impairment. Also, it is certainly possible that the siblings who volunteer for research participation might not be among the most “affected” members of a family, possibly accounting for the limited evidence of impairment we observed. Despite this possibility, the fact that first-episode patients seem to demonstrate minimal impairment on the ID/ED test bolsters confidence in our basic conclusions that the ID/ED test is: 1) sensitive to the impairments observed in chronically ill patients, and 2) not a robust intermediate phenotype marker.

Alan Ceaser, Drs. Terry Goldberg, Michael Egan, Robert McMahon, and Daniel Weinberger report no biomedical financial interests or potential conflicts of interest.

Dr. Gold receives royalties from the Brief Assessment of Cognition in Schizophrenia and has served as a consultant to Pfizer, Astra Zeneca, and Merck.


www.sobp.org/journal