Neurocognitive Deficit in Schizophrenia:
A Quantitative Review of the Evidence

R. Walter Heinrichs and Konstantine K. Zakzanis
York University

The neurocognitive literature on test performance in schizophrenia is reviewed quantitatively. The authors report 22 mean effect sizes from 204 studies to index schizophrenia versus control differences in global and selective verbal memory, nonverbal memory, bilateral and unilateral motor performance, visual and auditory attention, general intelligence, spatial ability, executive function, language, and interhemispheric tactile-transfer test performance. Moderate to large raw effect sizes (d > .60) were obtained for all 22 neurocognitive test variables, and none of the associated confidence intervals included zero. The results indicate that schizophrenia is characterized by a broadly based cognitive impairment, with varying degrees of deficit in all ability domains measured by standard clinical tests.

It is expected that schizophrenia will be understood in the near future as a kind of brain disease. Substantive and technical advances in the neurosciences are being brought to bear on the problem of tracking the disordered cognition and behavior of schizophrenia to dysfunctional brain structures and systems. Evidence from neuropsychological, neuropsychopathological, and neuroimaging studies has revived interest in schizophrenia as a neuropsychiatric disease with a putatively identifiable neuropathophysiology (Andreasen et al., 1994; Heinrichs, 1993; Levin, Yurgelun-Todd, & Craft, 1989; Randolph, Goldberg, & Weinberger, 1993).

From a behavioral perspective there is broad agreement that schizophrenia produces impairment of neuropsychological function. Individuals with the illness are likely to exhibit impairment on a wide assortment of neuropsychological tasks (Blanchard & Neale, 1994). Recent findings include deficits in attention (Braff, 1993; Cornblatt & Keilp, 1994), executive function (Gruzelier, Seymour, Wilson, Jolley, & Hirsch, 1988; Heinrichs, 1990; Van der Does & Van den Bosch, 1992; Weinberger, Wagner, & Wyatt, 1986), motor and tactile dexterity (Goldstein & Zubin, 1990; Schwartz, Carr, Munich, & Bartuch, 1990), spatial abilities (Green & Walker, 1985; Raine, 1992; Stuss et al., 1983), affect recognition (Andreasen & Ehrhardt, 1982; Ennis & Whelton, 1994), intellectual ability (Goldberg, Gold, Greenberg, & Griffin, 1993), language functions (Crawford, Obonsawin, & Bremner, 1993), and memory (Pauksen et al., 1995; Randolph et al., 1994; Saykin et al., 1991).

Although a number of narrative reviews have described and integrated neurocognitive findings on schizophrenia (e.g., Heaton & Crowley, 1981; Heinrichs, 1993; Levin et al., 1989; Randolph et al., 1993), there has been no comprehensive quantitative evaluation of this literature. In view of the increasing use of neurocognitive measures in schizophrenia research and the accumulation of empirical evidence, we undertook such an evaluation. A number of questions were formulated to guide our review:

1. Does neurocognitive testing provide reliable evidence of impairment in schizophrenia and what is the average magnitude of difference between patients and healthy controls?
2. Do tests of specific neurocognitive functions (e.g., memory, language, attention) reveal similar magnitudes of difference between patients and controls or are some aspects of neurocognitive performance spared in the illness?
3. Are there relationships between neurocognitive impairment and clinical and demographic attributes of patients and controls?
4. Are there relationships between specific neurocognitive functions and more general tests of ability (e.g., IQ)?

Accordingly, the goal of our research was to estimate the consistency, strength, and selectivity of neurocognitive deficits in schizophrenia. To accomplish this goal we compiled the results of multiple studies using meta-analytic techniques. In addition to solving potential problems with traditional literature reviews (see Wolf, 1986), meta-analysis provides tools for the analysis of magnitude and consistency of evidence (i.e., effect size d, distribution nonoverlap percentage, confidence intervals) as well as tools to assess variables that may mediate the magnitude and consistency of evidence (i.e., moderator variable analysis). The effect size estimate d measures the degree to which the phenomenon is present in the population (Cohen, 1988). In mathematical terms, d is the difference between patient and control means on a variable of interest calibrated in pooled standard deviation units. Effect size analysis allows the pitfalls of null hypothesis–statistical significance testing to be avoided, including faulty conclusions about a hypothesis that are
based on a count of significant and nonsignificant studies (see Bakan, 1966; Cohen, 1994; Schmidt, 1996). Variability of effect sizes across studies can be indexed by common statistics like standard deviations and confidence intervals but also by formal tests of homogeneity (see Hedges, 1994; Hedges & Olkin, 1985; Rosenthal, 1991, 1995). This variability in effects can be explored and articulated through moderator variable analysis if it is due to study differences in sample or design features.

With these considerations in mind we organized 22 neurocognitive test variables into meta-analyses to summarize the magnitude of schizophrenia-control discrimination in the published literature (see Table 1). These variables reflected memory, motor, attention, intelligence, spatial function, executive ability, language, and interhemispheric transfer processes. Effect sizes were calculated for each of the individual dependent variables. Attribute variables (e.g., age, education, hospitalizations, etc.) were recorded and, when numbers warranted, were correlated with neurocognitive effect sizes to explore the contribution of moderators to effect size variation.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Recorded test variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Global Verbal Memory</strong></td>
<td>CVLT: Trials I–V recall&lt;br&gt;RAVLT: Trials I–V recall&lt;br&gt;Portland Paragraph: immediate recall&lt;br&gt;Selective Reminding Test: sum recall&lt;br&gt;LNINB: memory index&lt;br&gt;WMS: logical memory score&lt;br&gt;WMS-R: verbal memory index</td>
</tr>
<tr>
<td><strong>Selective Verbal Memory</strong></td>
<td>CVLT: long-delay free recall, percentage words retained, number of intrusions, recognition trial hits&lt;br&gt;WMS-R: logical memory II&lt;br&gt;RAVLT: recognition trial hits&lt;br&gt;RMT: memory for words</td>
</tr>
<tr>
<td><strong>Nonverbal Memory</strong></td>
<td>ROCFT: recall trials score&lt;br&gt;WMS, WMS-R: visual reproduction trials&lt;br&gt;BVRT: recall trials&lt;br&gt;RMT: memory for faces</td>
</tr>
<tr>
<td><strong>Motor</strong></td>
<td>Purdue Pegboard: dominant hand score, bilateral score&lt;br&gt;Grooved Pegboard: dominant hand score&lt;br&gt;Motor Tapping Test: dominant hand score</td>
</tr>
<tr>
<td><strong>Attention</strong></td>
<td>WAIS-R: Digit Span&lt;br&gt;Trail Making Test: Parts A and B&lt;br&gt;CPT: hits, discrimination score&lt;br&gt;Sstroop Test: interference scores</td>
</tr>
<tr>
<td><strong>General Intelligence</strong></td>
<td>WAIS-R: Full Scale, Verbal, and Performance IQ&lt;br&gt;Non-WAIS-R: NART IQ, ILS, Quick Test, Mill Hill&lt;br&gt;Verbal IQ, and Wonderlic Personnel Test</td>
</tr>
<tr>
<td><strong>Spatial Ability</strong></td>
<td>Benton Line Orientation Test: number correct&lt;br&gt;Benton Facial Recognition Test: corrected score&lt;br&gt;WAIS-R: block design score</td>
</tr>
<tr>
<td><strong>Executive Function</strong></td>
<td>WCST: categories achieved, number of perseverative responses, number of perseverative errors, percent perseverative errors</td>
</tr>
<tr>
<td><strong>Language Function</strong></td>
<td>WCST: categories achieved, number of perseverative responses, number of perseverative errors, percent perseverative errors</td>
</tr>
<tr>
<td><strong>Tactile-Transfer</strong></td>
<td>Contralateral finger localization, tactile recognition scores</td>
</tr>
</tbody>
</table>

**Note.** CVLT = California Verbal Learning Test (Delis et al., 1987); RAVLT = Rey Auditory Verbal Learning Test (Lezak, 1995, pp. 438–442); Portland Paragraph (Lezak, 1995, p. 462); Selective Reminding Test (Buschke & Fuld, 1974); LNINB = Luria-Nebraska Neuropsychological Battery (Golden et al., 1985); WMS = Wechsler Memory Scale (Wechsler, 1945); WMS-R = Wechsler Memory Scale—Revised (Wechsler, 1987); RMT = Recognition Memory Test (Warrington, 1984); ROCFT = Rey-Osterrieth Complex Figure Test (Lezak, 1995, pp. 475–480); BVRT = Benton Visual Retention Test (Benton, 1974); Purdue Pegboard (Tiffin, 1968); Grooved Pegboard (Klove, 1963); Finger Tapping Test (Reitan & Wolfson, 1993); WAIS-R = Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981); Trail Making Test (Reitan, 1958); CPT = Continuous Performance Test (Nuechterlein & Dawson, 1984); Stroop Test (Golden, 1978; Stroop, 1933); NART = National Adult Reading Test (Nelson, 1982); ILS = Institute of Living Scale (Shipley, 1946); Quick Test (Ammons & Ammons, 1962); Mill Hill Scale (Raven, 1976); Benton Line Orientation and Facial Recognition Tests (Benton et al., 1994); WCST = Wisconsin Card Sorting Test (Heaton et al., 1993); COWAT = Controlled Oral Word Association Test (Benton & Hamsher, 1983); Chicago WFT = Chicago Word Fluency Test (Milner, 1964); Token Test (De Renzi & Vignolo, 1962); Peabody Picture Vocabulary Test (Dunn & Dunn, 1981).
Method

Literature Search

The PsycINFO and Medline databases were used to find studies for inclusion in the meta-analyses. The key words used were included in the Appendix. The articles located by the computer search were limited to published English language articles. According to Green and Hall (1984), meta-analysts doing a literature search should also page through the volumes of pertinent journals for a topic by year. This was done with every issue for the two relevant specialty journals, Schizophrenia Research and Schizophrenia Bulletin, as well as for six psychiatric and psychological journals that publish a high volume of studies on schizophrenia (American Journal of Psychiatry, Archives of General Psychiatry, Biological Psychiatry, British Journal of Psychiatry, Journal of Abnormal Psychology, and Journal of Nervous and Mental Disease) and five journals in neuroscience and behavior that publish schizophrenia research but at a lower frequency than the psychiatric journals (Brain, Journal of the International Neuropsychological Society, Neuropsychology, Neuroscience, and Neuropsychopharmacology). This proved to be an effective supplementary way to search for articles, given the small and focused set of key words in any particular abstract or title that a computer database uses to search for key words. Despite its apparent cost in time, the reasoning behind using the individual journal technique was to avoid missing a useful article that lies outside the database’s regular purview. Although absolute coverage of the entire relevant literature is hard to achieve, this serves to reduce the likelihood that bias was involved in the search outcome (White, 1994).

In the course of the literature search, the frequency with which any given neurocognitive variable was reported across studies dictated suitability for meta-analytic synthesis. If the variable was reported in two or more primary studies, the variable was tracked for meta-analytic synthesis. In this way, studies were compiled for 22 meta-analyses of different neurocognitive test measures.

Criteria for Inclusion

Articles were included if they met the following criteria: (a) publication between 1980 and 1997, (b) research designs with a control group comprising healthy participants, and (c) study statistics convertible to effect size $d$ (e.g., means, standard deviations, $F$, $t$, $X^2$; see Wolf, 1986). Studies reporting only significance levels with no descriptive or inferential statistics were excluded. The year 1980 was chosen as a year of publication criterion because it corresponded roughly to the introduction and use of more systematic and reliable diagnostic criteria for schizophrenia, such as the Diagnostic and Statistical Manual of Mental Disorders (3rd ed.; DSM-III; American Psychiatric Association, 1987). The year 1997 was chosen as an upper year limit to ensure maximal coverage of the literature by the computer-based journal databases. However, there is typically a delay of a few months between journal publication and availability in the database, and our article search took place in mid-1997. Therefore, coverage of 1997 articles was incomplete.

If these criteria were met, the article was then assessed for two additional criteria. First, schizophrenic patients must have met diagnostic criteria for either the DSM-III (or later) or the International Classification of Diseases (ICD-9 or 10) classification systems and have satisfied these criteria on the basis of a structured clinical interview. Second, neurocognitive tests must have been administered by an examiner trained in standardized testing procedures and supervised by a psychologist. This was important to track in order to ensure that the quality of the neurocognitive assessment administered in each study was held relatively constant and did not influence the findings. In the case of neurocognitive data published in drug trials, only studies reporting pretreatment baseline data were included.

If the research article met the above criteria, its content variable(s) was included in one (or more) of the meta-analyses. In the case of separately published studies that used the same participant samples, the decision rule was adopted to treat these studies as a single study with multiple independent variables (Hedges & Olkin, 1985). The $d$ statistic (Cohen, 1988) was calculated for each comparison as the difference between schizophrenia and control group means normalized by the pooled standard deviation. Whenever means and standard deviations were reported, these were used to derive effects. When inferential statistics were reported without central tendency and dispersion data, the effects were calculated from these statistics based on formulas provided by Wolf (1986).

Recorded Variables

Recorded variables for every article used in each of the meta-analyses included the journal name, title, author(s), and date of publication of the article. Recorded potential moderator variables included: interest of published mean and education in control and patient samples; for the patient samples, the chlorpromazine-equivalent medication dose (equivalent neuroleptic daily dosage in mg), percentage of male participants, onset age (defined as the age at first psychiatric admission) and duration of illness, percentage of patients on medication, and number of psychiatric hospitalizations were recorded. These study characteristics were used to describe the study set retrieved and for moderator variable analysis.

Organizing the myriad of neurocognitive test variables reported in the literature into a coherent classification was a major challenge. Several strategies exist in the literature for organizing diverse tests into categories of neurocognitive function, and each of these has advantages and disadvantages. First, there are a priori approaches such as Lezak’s (1995) classification, which are influenced by theoretical and practice-related considerations about the test measures and their putative underlying processes. For example, Lezak includes motor and executive ability tasks in the same chapter, presumably on the basis of a common substrate in the frontal brain or some other assumed link. Such classifications have quantitative statistical underpinning, and even advocates of this approach admit to an element of arbitrariness in test organization (see Lezak, pp. 333–334). A second approach is based on factor-analytic studies of neuropsychological test batteries (see Goldstein, 1984, pp. 184–210). Factor analysis provides a quantitative description that relates different tests to a smaller number of underlying abilities. However, the validity of this approach as a general strategy for organizing tests in a meta-analysis depends in part on the availability of factor analyses that include all of the tests in the literature on neuropsychology and schizophrenia. In practice it is the standard test batteries like the Halstead-Reitan (e.g., Ernst, Warner, Hochberg & Townes, 1988) and Luria-Nebraska (e.g., McKay & Golden, 1981) that tend to be studied factor analytically. In the present case, we were unable to find factor-analytic studies of schizophrenic samples that included all, or even most, of the neurocognitive test variables reported in the literature. Finally, it is possible to avoid constructs altogether and simply compile effects for individual tests. This approach incorporates the fewest assumptions about the data; however, it is unwieldy in view of the dozens of tests in common use and the inconsistency with which different scores from the same test are reported in the literature (e.g., categories vs. perseverative errors, Wisconsin Card Sorting Test; see Heaton, Chelune, Talley, Kay, & Curtiss, 1993).

A further consideration is that regardless of classification method,
many neuropsychological tests are probably influenced by several component processes. Thus, scores on the Vocabulary subtest of the Wechsler Adult Intelligence Scale—Revised (WAIS—R; Wechsler, 1981) may reflect both language abilities and general intelligence; scores on the Trail Making Test (Reitan, 1958) may reflect visual scanning and perception but also motor speed, hand–eye coordination, and attention (Lezak, 1995, pp. 382–384). Hence it may be misleading to categorize tests on the basis of a faulty assumption that test performance is determined by only one process.

Accordingly, whenever possible we tried to avoid aggregating different tests and their effect sizes into hypothetical categories and adopted a strategy, presented in Table 1, whereby effect sizes were calculated and reported on individual tests. In most cases effect sizes were grouped into a priori categories for presentation only and not for statistical purposes. However, this was not feasible for memory test variables because of the diversity of tests and related variables reported in the literature. Hence memory test variables were selected and aggregated into three effect size categories: global verbal, selective verbal, and nonverbal. The global verbal memory category included summary indexes of performance like total words recalled across learning trials and verbal memory quotients. Selective memory variables included delayed-recall scores, savings over time, and intrusion error rates. These selective variables are believed to be more sensitive than global scores to memory disorders involving the medial temporal–diencephalic system (see Delis, Kramer, Kaplan & ObrÃ¡, 1987, pp. 2–3; Lezak, 1995, pp. 147–150). Non-verbal memory variables included recall scores for visual stimuli such as shapes and patterns that did not involve language. The distinction between verbal and nonverbal memory has a long history in neuropsychology and reflects the existence of distinct, or “material-specific,” memory disorders in neurological patients (Callev, Edelst, Kugelmass, & Lerer, 1991). Similarly, tests of unilateral manual dexterity and finger oscillation were collapsed into a unilateral motor category. This arrangement is supported by factor-analytic studies showing the emergence of a distinct motor factor in psychiatric populations (e.g., Goldstein & Shelly, 1972). These tests included Finger Tapping (Reitan & Wolfson, 1993), the Grooved Pegboard (Klove, 1963), and the Purdue Pegboard (Tiffin, 1968).

The executive category comprised several test variables from the Wisconsin Card Sorting Test (WCST; Heaton et al., 1993). This is by far the most widely used test of executive ability in schizophrenia studies (see Van der Does & Van den Bosch, 1992). Like Saykin et al. (1991), we combined perseveration (including perseverative responses, errors, and percentage of perseverative errors) and categories scores into an overall index of WCST performance. In the event that studies reported data on attempts to teach schizophrenia patients how to do the test, only preinstruction baseline data were used.

Attention is a basic neuropsychological function involved in some degree in all test performances. Nonetheless, several tests make relatively modest demands on higher-level cognition and are regarded primarily as measures of attention, selection, concentration, and vigilance. They include two forms of digit span tasks: one with concurrent distraction (Oltmanns, 1978) and one without (WAIS—R, Digit Span subtest; Wechsler, 1981). Digit Span meta-analysis was composed of test scores from both of these paradigms. In addition, the Continuous Performance Test (CPT) is a class of attentional tasks that all involve a requirement to respond to target and ignore non-target stimuli over a period of time. Scores like target “hits,” errors, and sensitivity measures (”d”) based on signal detection theory are available for the test. In their analysis of the CPT, Van den Bosch, Rombouts, and van Asma (1996) concluded that motor speed along with attentional mechanisms contribute to successful performance. The same almost certainly applies to the Trail Making Test (Reitan, 1958), which Lezak (1995, pp. 382–384) considers an attentional task with perceptual and motor components. In addition, the Trail Making Test has two forms that differ in complexity, namely, Part A and Part B. In contrast, the Stroop Test (Golden, 1978; Stroop, 1935) does not require a manual response. The strength of the interference trial effect, where participants have to name perceptually incongruent color words, has long been regarded as a measure of selective attention in both clinical and experimental psychology. Completion time or errors were coded for Part A and Part B of the Trail Making Test. Continuous Performance Test scores included hits, “d,” and errors. Interference scores, error, or reaction time in the incongruent color–word condition were collected for the Stroop Test.

Full-scale IQ scores were indexed from articles using the WAIS—R, the Quick Test, the National Adult Reading Test (NART), the Wonderlic Personnel Test (Wonderlic, 1978), and the Peabody Picture Vocabulary Test (see Table 1). Intelligence quotients derived from the WAIS—R were treated separately from those derived from alternative tests like the NART. Only WAIS—R performance IQs were gathered for the performance IQ meta-analytic synthesis. The verbal IQ meta-analysis also included test scores from the WAIS—R only.

Spatial abilities were assessed with the Block Design subtest (WAIS—R), and two tests from Benton’s (Benton, Sivan, Hamsher, Varney, & Spreen, 1994) laboratory, Facial Recognition and Line Orientation (see Table 1).

In the domain of language, word fluency test scores collected included data from the Chicago Word Fluency Test (Milner, 1964) and the Controlled Oral Word Association Test (Benton & Hamsher, 1983), as well as from generically named word fluency tasks. Recorded vocabulary scores were taken from the WAIS—R, Shipley, and Mill Hill scales (see Table 1). The Token Test (De Renzi & Vignolo, 1962) was used to index receptive language.

Finally, the interhemispheric transfer-task meta-analysis was made up of scores from finger localization tests and transfer-task test scores (see Bellini et al., 1991; Carr, 1980; Ditchfield & Hemsley, 1990; Straube & Oades, 1992, pp. 339–341).

Results and Discussion

Data from each primary study were entered into the Statistical Package for the Social Sciences (SPSS, 1990) and the Software for the Meta-Analytic Review of Research Literature (DSTAT; Johnson, 1989) statistical software packages to produce the descriptive and inferential meta-analytic results. Descriptive data for the published studies are presented first, followed by mean effect size summaries and related statistical analyses.

Descriptive Data

Two hundred and four studies published between 1980 and mid-1997, yielding 509 effect sizes, met criteria for inclusion in the present analysis. Descriptive statistics for the study set are shown in Table 2. The table shows how many studies were used in each of the 22 meta-analyses, as well as the number of individual effect sizes that were incorporated into each mean effect size. The number of schizophrenia patients and normal controls for each meta-analysis is also included. In total, neurocognitive test results from 7,420 schizophrenia patients and 5,865 normal healthy controls were recorded across meta-analyses.

Table 2 also includes a “vote count” statistic in which the schizophrenia—control comparison from each primary study was tallied as significant (p < .05) or nonsignificant
between schizophrenia and normal control groups. The highest proportion of significant test score differences occurred on the Trail Making Test—Part B, and word fluency tests. Published literature showed significant differences in global verbal memory, nonverbal memory, and attention. Accordingly, in the published literature, 33 schizophrenia-control comparisons were reported as statistically significant and three were nonsignificant (the vote count). Thus, for example, there were 33 schizophrenia-control comparisons with global verbal memory data drawn from 31 independent studies. Thirty of these comparisons were reported as statistically significant and three were nonsignificant (the vote count). Accordingly, in the published literature, global verbal memory, nonverbal memory, Trail Making Test—Part B, and word fluency tests yielded the highest proportion of significant test score differences between schizophrenia and normal control groups.

(p > .05). Thus, for example, there were 33 schizophrenia-control comparisons with global verbal memory data drawn from 31 independent studies. Thirty of these comparisons were reported as statistically significant and three were nonsignificant (the vote count). Accordingly, in the published literature, global verbal memory, nonverbal memory, Trail Making Test—Part B, and word fluency tests yielded the highest proportion of significant test score differences between schizophrenia and normal control groups.

Descriptive data available for demographic and clinical variables are presented in Table 3. In terms of gender composition, the published literature reflects samples that are disproportionately male (82.4%) and fairly chronic in terms of the length of illness history. Most patients underwent initial psychiatric hospitalization in their early twenties and almost 78% were medicated at the time of neurocognitive testing. The percentage of schizophrenia patients on medication is presented in Table 3. This statistic is more

Table 2
Sample-Size Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. of Studies</th>
<th>n d</th>
<th>Schizophrenia (n)</th>
<th>Controls (n)</th>
<th>Vote count* (SIG/NSIG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Verbal</td>
<td>31</td>
<td>33</td>
<td>1,088</td>
<td>1,187</td>
<td>30/3</td>
</tr>
<tr>
<td>Selective Verbal</td>
<td>7</td>
<td>9</td>
<td>559</td>
<td>733</td>
<td>6/3</td>
</tr>
<tr>
<td>Nonverbal</td>
<td>14</td>
<td>16</td>
<td>379</td>
<td>577</td>
<td>13/3</td>
</tr>
<tr>
<td>Motor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unilateral Skill</td>
<td>6</td>
<td>16</td>
<td>232</td>
<td>179</td>
<td>11/5</td>
</tr>
<tr>
<td>Bilateral Skill</td>
<td>5</td>
<td>5</td>
<td>237</td>
<td>249</td>
<td>3/2</td>
</tr>
<tr>
<td>Attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>18</td>
<td>18</td>
<td>440</td>
<td>401</td>
<td>12/6</td>
</tr>
<tr>
<td>Trail Making Test—Part A</td>
<td>12</td>
<td>12</td>
<td>1,204</td>
<td>596</td>
<td>9/3</td>
</tr>
<tr>
<td>Trail Making Test—Part B</td>
<td>15</td>
<td>15</td>
<td>1,372</td>
<td>805</td>
<td>13/2</td>
</tr>
<tr>
<td>Continuous Performance</td>
<td>14</td>
<td>15</td>
<td>417</td>
<td>335</td>
<td>10/5</td>
</tr>
<tr>
<td>Stroop</td>
<td>6</td>
<td>6</td>
<td>179</td>
<td>130</td>
<td>3/3</td>
</tr>
<tr>
<td>General Intelligence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIS-R IQ</td>
<td>35</td>
<td>35</td>
<td>1,018</td>
<td>1,048</td>
<td>18/17</td>
</tr>
<tr>
<td>Non-WAIS-R IQ</td>
<td>43</td>
<td>43</td>
<td>1,069</td>
<td>1,233</td>
<td>25/18</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>17</td>
<td>17</td>
<td>717</td>
<td>480</td>
<td>14/3</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>27</td>
<td>27</td>
<td>995</td>
<td>658</td>
<td>14/13</td>
</tr>
<tr>
<td>Spatial Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block Design</td>
<td>12</td>
<td>13</td>
<td>1,166</td>
<td>479</td>
<td>7/6</td>
</tr>
<tr>
<td>Line Orientation</td>
<td>4</td>
<td>4</td>
<td>225</td>
<td>202</td>
<td>3/1</td>
</tr>
<tr>
<td>Facial Recognition</td>
<td>8</td>
<td>8</td>
<td>197</td>
<td>157</td>
<td>6/2</td>
</tr>
<tr>
<td>Executive Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin Card Sort</td>
<td>43</td>
<td>104</td>
<td>1,387</td>
<td>1,153</td>
<td>73/31</td>
</tr>
<tr>
<td>Language Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Fluency</td>
<td>29</td>
<td>36</td>
<td>1,020</td>
<td>899</td>
<td>30/6</td>
</tr>
<tr>
<td>Token</td>
<td>7</td>
<td>7</td>
<td>290</td>
<td>239</td>
<td>6/1</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>38</td>
<td>38</td>
<td>2,046</td>
<td>1,186</td>
<td>13/25</td>
</tr>
<tr>
<td>Tactile-Transfer</td>
<td>12</td>
<td>16</td>
<td>297</td>
<td>287</td>
<td>11/5</td>
</tr>
</tbody>
</table>

Note. n d = number of effect sizes (d) in each meta-analysis; WAIS-R = Wechsler Adult Intelligence Scale—Revised.

*a Vote count refers to the number of statistically significant (SIG) effects in published studies relative to the number of nonsignificant (NSIG) effects.

Table 3
Schizophrenia Sample Characteristics for Studies Used in the Meta-Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mdn</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>23.5</td>
<td>36.1</td>
<td>30.2</td>
<td>5.0-819.0</td>
<td>204</td>
</tr>
<tr>
<td>Patient age (years)</td>
<td>31.0</td>
<td>34.4</td>
<td>10.0</td>
<td>18.1-56.5</td>
<td>200</td>
</tr>
<tr>
<td>Age of onset (years)</td>
<td>21.9</td>
<td>22.2</td>
<td>3.2</td>
<td>17.5-28.0</td>
<td>40</td>
</tr>
<tr>
<td>Length of illness (years)</td>
<td>12.3</td>
<td>12.7</td>
<td>7.6</td>
<td>0.0-29.6</td>
<td>43</td>
</tr>
<tr>
<td>Hospital admissions (a)</td>
<td>3.6</td>
<td>3.9</td>
<td>7.4</td>
<td>0.0-9.5</td>
<td>29</td>
</tr>
<tr>
<td>Male (%)</td>
<td>77.0</td>
<td>82.4</td>
<td>63.0</td>
<td>25.0-100.0</td>
<td>168</td>
</tr>
<tr>
<td>Patient education (years)</td>
<td>12.0</td>
<td>12.0</td>
<td>1.1</td>
<td>9.0-15.3</td>
<td>134</td>
</tr>
<tr>
<td>Medication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% on neuroleptics</td>
<td>100.0</td>
<td>77.6</td>
<td>36.6</td>
<td>0.0-100.0</td>
<td>100</td>
</tr>
<tr>
<td>Cpz.-eq. daily dose (mg)*</td>
<td>577.3</td>
<td>582.1</td>
<td>340.3</td>
<td>0.0-1,771.9</td>
<td>59</td>
</tr>
</tbody>
</table>

*a The cpz.-eq. daily dose is the mean chlorpromazine-equivalent (cpz.-eq.) dose in schizophrenia samples as reported in 59 studies. Studies that recruited medication-free or medication-naive patients were entered as a cpz.-eq. dose of 0. The remaining 145 studies included in the meta-analyses did not report medication data in cpz.-eq. units.
Informative than the chlorpromazine-equivalent daily dose because of infrequent neuroleptic dose-reporting in the published literature. In addition, clozapine and risperidone dosages are not convertible into chlorpromazine doses. However, all anti-psychotic medication is reflected in the binary medication variable. Of 204 studies, only 13 (6%) reported data on unmedicated patient samples. The possible influence of medication on neurocognitive test performance has been recognized for many years (see Spohn & Strauss, 1989), but convincing evidence for this influence has been lacking and clinicians face ethical constraints in removing patients from treatment for research purposes. Thus, apart from some exceptions (e.g., Cleghorn, Kaplan, et al., 1990), researchers relied largely on medicated patient samples during the period between 1980 and 1997.

**Neurocognitive Tests**

Effect size summaries are presented for each neurocognitive test or test category in Tables 4—9 and 11. Two kinds of ds are reported. Mean ds presented in the second column of Tables 4—9 are raw effect sizes in absolute value, and those presented in Table 11 are corrected for sample size (see Johnson, 1989), percentage overlap (O/L%), and 95% confidence interval (for raw ds only) are presented in Tables 4 and 11. The results show that defective total verbal learning over trials, or similar summary scores, is a reliable finding in the schizophrenia literature, with a schizophrenia-control distribution overlap of somewhat more than 25%. More specific studies, using variables like rate of forgetting, savings after a delay interval, or other indexes sensitive to organic amnesic disorders, also reveal a substantial effect, although relatively few studies reported these variables. At the same time, there is no suggestion of a disproportionate impairment on selective measures relative to more global indexes. Moreover, the nonverbal results are comparable in magnitude to the verbal memory data. However, nonverbal memory shows more heterogeneity of effects across studies and appears to be a less reliable deficit than verbal memory deficits.

In terms of possible moderator variables, there were no significant product-moment correlations between basic demographic and clinical variables and the memory-related effect sizes. However, with a limited number of studies reporting, there were nonsignificant correlation trends with several variables. These included relationships between schizophrenia sample age and global verbal memory effect sizes, $r(33) = -.25$, $p > .05$, two-tailed. In addition, there were suggestive trends in the selective verbal memory data with percentage of male patients, $r(9) = .58$, $p > .05$, two-tailed, and with years of education, $r(9) = -.41$, $p > .05$, two-tailed. These relations between verbal memory, age, and gender also occur in normal participants (e.g., Delis et al., 1987, pp. 35–37; Elwood, 1995). The nonverbal effect sizes correlated modestly with percentage of male patients, $r(16) = .31$, $p > .05$, two-tailed. Age of illness onset and hospitalization data were reported too infrequently to allow for calculation of correlations.

**Executive, motor, and tactile-transfer.** Effect sizes for the WCST, based on a summary of its constituent variables, including perseverative responses, errors, and categories, are presented in Table 5. These results show a moderately large and reliable impairment of WCST performance in schizophrenia samples, with roughly half of the patients separated from normal control participants. The large number of studies reporting WCST results allowed for analysis of moderator variables, but only the correlation with number of hospitalizations approached significance, $r(29) = .36$, $p > .05$.

### Table 4

**Mean Effect Sizes ($M_d$) for Memory Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n_d$</th>
<th>$M_d$</th>
<th>$SD_d$</th>
<th>O/L %</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Verbal</td>
<td>33</td>
<td>1.53</td>
<td>0.66</td>
<td>29.3</td>
<td>1.27–1.86</td>
</tr>
<tr>
<td>Selective Verbal</td>
<td>9</td>
<td>1.11</td>
<td>0.64</td>
<td>41.1</td>
<td>0.62–1.60</td>
</tr>
<tr>
<td>Nonverbal</td>
<td>16</td>
<td>1.42</td>
<td>2.33</td>
<td>31.9</td>
<td>0.18–2.67</td>
</tr>
</tbody>
</table>

**Note.** $n_d$ = number of effect sizes (Cohen's $d$); $SD_d$ = standard deviation of effect sizes; O/L % = percentage overlap between test score distributions for schizophrenia patients and normal controls; 95% CI = 95% confidence interval.

### Table 5

**Mean Effect Sizes ($M_d$) for Executive, Motor, and Tactile-Transfer Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n_d$</th>
<th>$M_d$</th>
<th>$SD_d$</th>
<th>O/L %</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wisconsin Card Sort</td>
<td>104</td>
<td>0.95</td>
<td>0.44</td>
<td>46.5</td>
<td>0.85–1.04</td>
</tr>
<tr>
<td>Unilateral Motor</td>
<td>16</td>
<td>0.89</td>
<td>0.42</td>
<td>48.4</td>
<td>0.66–1.11</td>
</tr>
<tr>
<td>Bilateral Motor</td>
<td>5</td>
<td>1.42</td>
<td>0.40</td>
<td>31.9</td>
<td>1.01–1.65</td>
</tr>
<tr>
<td>Tactile-Transfer</td>
<td>16</td>
<td>1.39</td>
<td>1.90</td>
<td>31.9</td>
<td>0.38–2.41</td>
</tr>
</tbody>
</table>

**Note.** $n_d$ = number of effect sizes (Cohen's $d$); $SD_d$ = standard deviation of effect sizes; O/L % = percentage overlap between test score distributions for schizophrenia patients and normal controls; 95% CI = 95% confidence interval.
two-tailed. In addition, a large number of studies \((n = 33)\) reported WCST results along with WAIS–R IQ results, thereby permitting an examination of possible relationships between general intellectual levels in schizophrenia samples and the WCST effect size in the same sample. There was a significant relationship between WCST effect size and IQ score, \(r(33) = -.54, p < .01\), and between WCST effects and IQ effects, \(r(33) = .56, p < .005\). These relationships imply that patient samples are likely to differ from control samples on the WCST if they also differ in terms of IQ. Thus, impaired WCST scores in schizophrenia may be, in part, a reflection of low general intellectual abilities.

Unilateral motor tests produced effect sizes that were similar in magnitude to the WCST results, especially when the effects were corrected for sample size (see also Table 11). There was a nonsignificant association with the proportion of male patients in the study sample, \(r(14) = -.36, p > .05\), two-tailed. The number of studies reporting bilateral motor results was very small, making general conclusions difficult.

Results of tests where patients identify and discriminate tactile stimuli under conditions that require transfer of information across the corpus callosum showed a large mean effect size, even after correction for sample size. At the same time, the results were very heterogeneous, indicating the existence of patient samples with extensive overlap, as well as samples with very little overlap, with control distributions. There were no significant moderator correlations in the tactile-transfer results, but there were suggestive trends with patient sample age, \(r(15) = .47, p > .05\), two-tailed; education, \(r(10) = -.68, p > .05\), two-tailed; and percentage of male patients, \(r(15) = .33, p > .05\), two-tailed.

**Attention.** Results for attention variables are presented in Table 6. The most frequently reported, attention-related test variable was digit span. Eighteen studies reported data, with 3 of these reporting results with the distraction version (Oltmanns, 1978) and the rest with the WAIS–R version without distraction. The average effect size for the subset of distraction digit spans was .81, but the number of studies was too small to allow for a statistical comparison. The overall effect size of .62 represents a hypothetical overlap with control distributions of about 60%, suggesting a fairly modest degree of discrimination. Larger effects were obtained in studies using versions of the Continuous Performance Test and the Stroop Test, but the number of articles reporting Stroop results is small. Overall, the Continuous Performance and Trail Making tests seem to be about equally sensitive to schizophrenia, yielding large effect sizes that separate more than half of the patient and control distributions. Nevertheless, Trail Making Test results are based on smaller sample sizes, which causes shrinkage in the corrected effect size compared with the Continuous Performance data (see Table 11 presented later).

The two effect sizes for the Trail Making Test are also informative because they represent a situation where a simple and a more complex and demanding version of a neurocognitive task can be studied. If task difficulty and task complexity, rather than specific neurocognitive functions like visual scanning and sequencing, contribute to schizophrenia–control differences, more demanding versions of the same basic test should yield significantly larger effect sizes than less demanding versions. Thus Part B of the Trail Making Test should produce a larger effect size than Part A when applied to the same sample of patients and controls. Part A requires rapid sequential connection of numbers in a paper-and-pencil format, whereas Part B requires alternation between alphabetic and numeric sequences. Fossum, Holmberg, and Reinvang (1992) have shown that the spatial arrangement and dual symbol systems of Part B underlie the slower Part B completion times.

Accordingly, we carried out a paired-difference \(t\) test comparing effect sizes from the two forms. There was no significant difference between the two parts of the Trail Making Test, \(t(25) = -1.3, p > .05\), two-tailed. Hence, it appears likely that task difficulty and complexity are less important than the more specific attention and scanning-related neurocognitive processes elicited by this particular test. Alternatively, because both alphabetic and numeric sequences are probably overlearned information, the differences in processing demands between Parts A and B of the Trail Making Test may be minimal.

**General intellectual ability.** Effect sizes for intelligence test results are presented in Table 7. Full scale IQ estimates were obtained from a variety of intelligence tests in addition to the WAIS–R (see Table 1). To assess whether WAIS–R and non-WAIS–R-derived IQs differed in terms of their ability to separate schizophrenia and control distributions, we calculated a \(t\) test on the respective mean \(d\) values. This calculation indicated substantially larger average effect sizes in patient–control comparisons that used the WAIS–R, \(t(76) = 2.81, p < .01\), two-tailed. Whether this indicates that the WAIS–R is more difficult for patients to complete than

### Table 6

**Mean Effect Sizes \(M_d\) for Attention Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>(n_d)</th>
<th>(M_d)</th>
<th>SD(d)</th>
<th>O/L %</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digit Span</td>
<td>18</td>
<td>0.62</td>
<td>0.51</td>
<td>61.8</td>
<td>0.35–0.96</td>
</tr>
<tr>
<td>Continuous Performance</td>
<td>15</td>
<td>1.18</td>
<td>0.49</td>
<td>37.8</td>
<td>0.94–1.50</td>
</tr>
<tr>
<td>Stroop</td>
<td>6</td>
<td>1.22</td>
<td>0.63</td>
<td>41.1</td>
<td>0.23–2.11</td>
</tr>
<tr>
<td>Trail Making—Part A</td>
<td>12</td>
<td>0.95</td>
<td>0.32</td>
<td>48.4</td>
<td>0.73–1.16</td>
</tr>
<tr>
<td>Trail Making—Part B</td>
<td>15</td>
<td>1.07</td>
<td>0.52</td>
<td>41.1</td>
<td>0.80–1.33</td>
</tr>
</tbody>
</table>

*Note. \(n_d\) = number of effect sizes (Cohen’s \(d\)); SD\(d\) = standard deviation of effect sizes; O/L % = percentage overlap between test score distributions for schizophrenia patients and normal controls; 95% CI = 95% confidence interval.*

### Table 7

**Mean Effect Sizes \(M_d\) for General Intelligence Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>(n_d)</th>
<th>(M_d)</th>
<th>SD(d)</th>
<th>O/L %</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIS–R IQ</td>
<td>35</td>
<td>1.24</td>
<td>0.89</td>
<td>37.8</td>
<td>0.88–1.56</td>
</tr>
<tr>
<td>Non-WAIS–R IQ</td>
<td>43</td>
<td>0.63</td>
<td>0.60</td>
<td>61.9</td>
<td>0.35–0.92</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>27</td>
<td>0.98</td>
<td>0.68</td>
<td>44.6</td>
<td>0.70–1.20</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>17</td>
<td>1.46</td>
<td>0.89</td>
<td>29.3</td>
<td>1.07–1.90</td>
</tr>
</tbody>
</table>

*Note. \(n_d\) = number of effect sizes (Cohen’s \(d\)); SD\(d\) = standard deviation of effect sizes; O/L % = percentage overlap between test score distributions for schizophrenia patients and normal controls; 95% CI = 95% confidence interval; WAIS–R = Wechsler Adult Intelligence Scale—Revised.*
measures like the Shipley, Quick, and NART is unclear. The non-WAIS–R measures tend to be briefer and less comprehensive in scope than the WAIS–R. At the same time, some of these alternative measures, like the NART, are regarded as measures of premorbid rather than morbid intelligence. We extracted the NART data from the study set and found a mean $d$ of .42 ($SD = .68$) for 19 studies. Hence, smaller differences between patients and controls on such tests may reflect a robustness of performance in the presence of schizophrenic illness. However, there is a considerable degree of dispersion around the NART mean, and individual study effects ranged from $-1.85$ to $+0.91$, which raises questions of reliability in the use of this measure to index premorbid ability levels. In any case, the data suggest that putative tests of general intellectual function are not interchangeable and can be expected to vary in sensitivity to schizophrenic illness.

Intelligence measured with the WAIS–R yields large effect sizes in schizophrenia–control comparisons, with distribution overlaps that are well below 50%. In addition, intellectual differences are reliable, with none of the confidence intervals including zero. Finally, no significant relationships were found between IQ effect sizes and potential moderators including age, education, neuroleptic dose, sample gender composition, or age of illness onset.

Spatial ability. Results pertaining to the relatively small number of patient–control comparisons that used tests of visuospatial skill are presented in Table 8. These tests are not used frequently in the schizophrenia literature, but the available data suggest there are moderately large effect sizes, with less than 50% overlap between schizophrenia and control distributions. However, the $d$ values shrink considerably when corrected for sample size (see Table 11 presented later), indicating that most studies using spatial ability tests are based on small samples of patients and controls. With this correction the effect sizes correspond to more moderate effect sizes reported in Tables 4–9 to 0.1 or smaller. Accordingly, we regard the available data as suggesting that spatial ability is a robust and changeable performance area in schizophrenia that is not likely to be influenced by confounders that are generally associated with the distribution of schizophrenia patients and normal controls. Nonetheless, any conclusion in this regard should be tempered by the fact that the data reported here are based on a relatively small number of studies.

Language. The effect size summary for language-related tests is presented in Table 9. This shows large effects for tests of word generation such as the Controlled Oral Word Association Test and the Token Test (the latter is primary an auditory comprehension measure; see Table 1). In both cases, these effect sizes represent only about one-third overlap between schizophrenia and control distributions. However, this evidence is qualified by a relatively small number of contributing studies in the case of the Token Test. In addition, the relatively large standard deviations imply considerable dispersion and hence heterogeneity of effect sizes in both the fluency and Token Test data.

In contrast, a more moderate average effect size was obtained for vocabulary tests. This is another aggregate of measures, with some derived from the WAIS–R Vocabulary subtest and others based on scales such as the Mill Hill and Shipley scales (see Table 1). Nonetheless, there was no significant difference between the average effect size generated by WAIS–R and non-WAIS–R-based vocabulary tests, $t(36) = 1.71, p > .05$.

Overall, expressive and receptive language tests appear to be fairly powerful and moderately reliable discriminators of schizophrenia and control populations. Some of the effect sizes correlated significantly with potential demographic or clinical moderators. However, with 13 studies reporting both fluency and chlorpromazine dose data, there was a strong trend for highly medicated samples to yield smaller word fluency differences relative to control samples, $r(13) = -.75, p > .05$.

Inferential data. It is unlikely that a literature review will uncover every study of a hypothesis that has been conducted. Rosenthal (1979) has called this the "file drawer problem" because of the tendency for studies of nonsignificant results to remain unpublished, perhaps languishing in forgotten file drawers (see also Wolf, 1986, p. 37). Kraemer and Andrews (1982) note that "published research studies tend to be biased toward positive findings. A study is often abandoned if it is apparent that statistically significant findings will not be forthcoming. Reports of nonsignificant findings are generally unpublished even when they are replications of earlier studies reporting significant results" (p. 405). To deal with the file drawer problem in the present set of meta-analyses, Orwin’s (1983) fail-safe $N$ procedure was carried out. Table 10 shows the estimated number of unpublished or unretrieved studies, each with a trivial effect size of $d = 0.1$ ($92%$ distribution overlap), needed to reduce the effect sizes reported in Tables 4–9 to 0.1 or smaller. Although the actual existence of such countervailing evidence cannot be entirely ruled out, it seems unlikely that such large numbers of studies are actually present in neglected file drawers.

Summary and Conclusions

In our quantitative review, we found neurocognitive deficit to be a reliable finding in schizophrenia. In addition, it is clear that the deficit exists in relation to most neurocogni-

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n_d$</th>
<th>$M_d$</th>
<th>$SD_d$</th>
<th>O/L %</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Fluency</td>
<td>36</td>
<td>1.39</td>
<td>1.20</td>
<td>31.9</td>
<td>0.91–1.80</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>38</td>
<td>0.69</td>
<td>0.48</td>
<td>57.0</td>
<td>0.50–0.86</td>
</tr>
<tr>
<td>Token Test</td>
<td>7</td>
<td>1.35</td>
<td>0.94</td>
<td>34.7</td>
<td>0.98–1.70</td>
</tr>
</tbody>
</table>

Note. $n_d =$ number of effect sizes (Cohen’s $d$); $SD_d =$ standard deviation of effect sizes; O/L % = percentage overlap between test score distributions for schizophrenia patients and normal controls; 95% CI = 95% confidence interval.
Table 10
The File Drawer Problem and the Fail-Safe N

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated number of unpublished studies needed to achieve $d = 0.1$ (92% overlap)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td></td>
</tr>
<tr>
<td>Global Verbal</td>
<td>472</td>
</tr>
<tr>
<td>Selective Verbal</td>
<td>91</td>
</tr>
<tr>
<td>Nonverbal</td>
<td>211</td>
</tr>
<tr>
<td>Motor</td>
<td></td>
</tr>
<tr>
<td>Bilateral Motor</td>
<td>66</td>
</tr>
<tr>
<td>Unilateral Motor</td>
<td>121</td>
</tr>
<tr>
<td>Attention</td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>94</td>
</tr>
<tr>
<td>Continuous Performance</td>
<td>162</td>
</tr>
<tr>
<td>Stroop</td>
<td>67</td>
</tr>
<tr>
<td>Trail Making Test—Part A</td>
<td>102</td>
</tr>
<tr>
<td>Trail Making Test—Part B</td>
<td>146</td>
</tr>
<tr>
<td>General Intelligence</td>
<td></td>
</tr>
<tr>
<td>WAIS-R IQ</td>
<td>399</td>
</tr>
<tr>
<td>Non-WAIS-R IQ</td>
<td>228</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>240</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>228</td>
</tr>
<tr>
<td>Spatial Ability</td>
<td></td>
</tr>
<tr>
<td>Line Orientation</td>
<td>20</td>
</tr>
<tr>
<td>Block Design</td>
<td>82</td>
</tr>
<tr>
<td>Facial Recognition</td>
<td>62</td>
</tr>
<tr>
<td>Executive Function</td>
<td></td>
</tr>
<tr>
<td>Wisconsin Card Sort</td>
<td>884</td>
</tr>
<tr>
<td>Language Function</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>224</td>
</tr>
<tr>
<td>Word Fluency</td>
<td>464</td>
</tr>
<tr>
<td>Token</td>
<td>62</td>
</tr>
<tr>
<td>Tactile-Transfer</td>
<td>141</td>
</tr>
</tbody>
</table>

Note. WAIS-R = Wechsler Adult Intelligence Scale—Revised.

The value of 0.1 was arbitrarily chosen as a small or trivial effect size (see Wolf, 1986, pp. 37–39).

Table 11
Mean Neurocognitive Effect Sizes Ordered by Magnitude and Corrected for Sample Size

<table>
<thead>
<tr>
<th>Test or construct</th>
<th>$M_d$</th>
<th>$SD$</th>
<th>$n$</th>
<th>Patients below $Mdn$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Verbal Memory</td>
<td>1.41</td>
<td>0.59</td>
<td>31</td>
<td>78</td>
</tr>
<tr>
<td>Bilateral Motor Skill</td>
<td>1.30</td>
<td>0.38</td>
<td>5</td>
<td>77</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>1.26</td>
<td>1.00</td>
<td>17</td>
<td>77</td>
</tr>
<tr>
<td>Continuous Performance</td>
<td>1.16</td>
<td>0.49</td>
<td>14</td>
<td>75</td>
</tr>
<tr>
<td>Word Fluency</td>
<td>1.15</td>
<td>1.00</td>
<td>29</td>
<td>75</td>
</tr>
<tr>
<td>Stroop Test</td>
<td>1.11</td>
<td>0.49</td>
<td>6</td>
<td>74</td>
</tr>
<tr>
<td>WAIS-R IQ</td>
<td>1.10</td>
<td>0.72</td>
<td>35</td>
<td>74</td>
</tr>
<tr>
<td>Token</td>
<td>0.98</td>
<td>0.49</td>
<td>7</td>
<td>71</td>
</tr>
<tr>
<td>Tactile-Transfer</td>
<td>0.98</td>
<td>1.71</td>
<td>12</td>
<td>71</td>
</tr>
<tr>
<td>Selective Verbal Memory</td>
<td>0.90</td>
<td>0.62</td>
<td>7</td>
<td>70</td>
</tr>
<tr>
<td>Wisconsin Card Sort</td>
<td>0.88</td>
<td>0.41</td>
<td>43</td>
<td>69</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>0.88</td>
<td>0.66</td>
<td>27</td>
<td>69</td>
</tr>
<tr>
<td>Unilateral Motor Skill</td>
<td>0.86</td>
<td>0.39</td>
<td>6</td>
<td>69</td>
</tr>
<tr>
<td>Trail Making—Part B</td>
<td>0.80</td>
<td>0.50</td>
<td>15</td>
<td>68</td>
</tr>
<tr>
<td>Nonverbal Memory</td>
<td>0.74</td>
<td>1.98</td>
<td>14</td>
<td>67</td>
</tr>
<tr>
<td>Trail Making—Part A</td>
<td>0.70</td>
<td>0.36</td>
<td>12</td>
<td>66</td>
</tr>
<tr>
<td>Facial Recognition</td>
<td>0.61</td>
<td>0.36</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>Digit Span</td>
<td>0.61</td>
<td>0.43</td>
<td>18</td>
<td>64</td>
</tr>
<tr>
<td>Line Orientation</td>
<td>0.60</td>
<td>0.63</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>Non-WAIS-R IQ</td>
<td>0.59</td>
<td>0.51</td>
<td>43</td>
<td>64</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.53</td>
<td>0.21</td>
<td>38</td>
<td>62</td>
</tr>
<tr>
<td>Block Design</td>
<td>0.46</td>
<td>0.39</td>
<td>12</td>
<td>61</td>
</tr>
</tbody>
</table>

Note. The $M_d$ refers to the mean effect size for the test variable indicated, corrected for sample size. The $n$ refers to the number of independent studies in the database for each mean effect size. The percentage of patients below the median refers to an estimate of the proportion of patients scoring below the median of the joint patient–control aggregate sample on the test variable in question. This estimate is derived from Rosenthal and Rubin’s (1979, 1982) binomial effect size display. WAIS-R = Wechsler Adult Intelligence Scale—Revised.
NEUROCOGNITIVE DEFICIT IN SCHIZOPHRENIA

At the mild end, a patient's neurocognitive dysfunction is relative to his or her own potential and may overlap with levels of function obtained by many healthy individuals. At the severe extreme, a patient may have cognitive and neurological abnormalities that are completely distinguished from normal control values.

The limitations of this kind of explanation include its testability. There are practical difficulties involved in accurately assessing premorbid neurocognitive potential in large numbers of schizophrenia patients. Moreover, the continuum argument must assume that patients who show average neurocognitive function—roughly half of those with the illness—have nevertheless undergone a relative decline. Therefore, the same proportion of patients must have been slightly above average by normative performance standards before illness onset. This seems a rather difficult prediction to test. Finally, although discordant monozygotic twins provide an important perspective on schizophrenia, they cannot be regarded as representative of either the healthy or patient populations as a whole.

Another explanation for the 61–78% neurocognitive deficit rate in our quantitative review is that such deficit is secondary, peripheral rather than central to the illness. Most diseases have peripheral features that are tied only modestly and indirectly to the primary pathology. For example, overall brain volume reductions are less prevalent in Alzheimer's disease than volume reductions in a specific structure, the hippocampus (Seab et al., 1988). Presumably, this reflects the primary locus of pathology in the hippocampus and other specific structures and the less direct relation of this focal pathology to general changes in brain volume. The functions measured by standard neurocognitive tests in schizophrenia patients may also be relatively peripheral to the still unknown primary dysfunction.

At the same time, it may not be reasonable to expect extremely high rates of control–patient discrimination given the vagaries inherent in psychological measurement. For a comparison, consider the normative data on the WCST provided by Heaton et al. (1993). The effect size for categories achieved on the WCST is \( -1.04 \) SD units when neurological patients with focal frontal lesions are compared with healthy controls. When “categories” is extracted as a separate score from our own meta-analytic data on schizophrenia, the mean effect size is \( -1.05 \). Given the traditional difficulty of discriminating schizophrenia from more conventional neurobehavioral disorders, it may well be that at least some neurocognitive tests are as sensitive to schizophrenia as they are to neurological conditions in general (Heaton et al., 1994). Moreover, the sensitivity of neurocognitive tests to schizophrenia compares favorably with structural and functional neuroimaging methods, which tend to yield effect sizes less than 1.0 (Raz and Raz, 1990; Zakzanis & Heinrichs, 1997).

Of equal theoretical interest for understanding schizophrenia is the vexed question of heterogeneity and whether substantial but not absolute (e.g., \( d > 4.0 \), nonoverlap = 97%) effect sizes reflect the existence of cognitively intact and impaired subpopulations of patients. In addition, several tests and constructs are characterized by large dispersion of individual effects around their means (e.g., Performance IQ, word fluency, tactile-transfer tests, nonverbal memory tests; see Table 11). Such dispersion also implies the existence of impaired and normal subgroups of patients. This is inimical to the continuum view of neurocognitive deficit, which argues for a single disease process with variable expression. In contrast, a strong form of the heterogeneity argument is the contention that multiple disease processes underpin schizophrenia (see Carpenter, Buchanan, Kirkpatrick, & Tamminga, 1993; Heinrichs, 1993; Liddle & Barnes, 1990). Advocates of the single disease-continuum model can point to brain features like ventricle size to show that schizophrenia patients are normally distributed on such measures, with little evidence of the kind of bimodality that might indicate subpopulations (David, Goldberg, Gibbon, & Weinberger, 1991). On the other hand, there is clear evidence of bimodal distributions in schizophrenia on some psychophysiological measures, like smooth-pursuit eye movements (Clementz, Grove, Iacono, & Sweeney, 1992; Iacono, Moreau, Beiser, Heming, & Liu, 1992). Our meta-analytic evidence cannot resolve this debate, but it forcefully underscores the need to develop testable illness models that take heterogeneity into account and extend the search for evidence of multiple disease processes.

Despite the fact that all tests and neurocognitive categories yielded at least moderate and often large effects, some tests appear to produce greater differences between patients and controls than other tests. The order of effect magnitudes in Table 11 bears a very rough similarity to Saykin et al.'s (1991) study of a schizophrenia sample that showed the greatest difference from controls on a group of memory functions, followed by differences in motor–attentional and language functions. In addition, a recent meta-analysis by Woodruff, McManus, and David (1995) supports the idea that corpus callosum abnormalities exist in schizophrenia. This is consistent with defective recognition of tactile information on interhemispheric transfer tasks. Nevertheless, our meta-analysis, based on 204 studies, 7,420 schizophrenia patients, and 5,865 normal controls, confirms previous suggestions that any selectivity of deficit in schizophrenia occurs in the context of a background of a very general impairment. Even the less-sensitive tasks yielded effect sizes between 0.5 and 1.0 on average, suggesting that any differential sensitivity of specific tests is relatively mild and coexists with a more broadly based impairment in cognitive brain function. Moreover, it is not known to what extent differences in task complexity, difficulty, floor or ceiling effects, and other psychometric properties may have contributed to the appearance of more or less powerful effect sizes across selected neurocognitive measures (see Chapman & Chapman, 1973, 1978). Data from the Trail Making Test suggest that more complex tests do not necessarily increase the disadvantage patients show relative to controls on cognitive tasks. However, a comparison of WAIS–R- and non-WAIS–R-derived IQ estimates showed that different measures of the same cognitive function may yield markedly different degrees of separation between schizophrenia and control samples.
Hence, a cautious interpretation of our meta-analytic data would be that all areas of neurocognitive function are compromised in a large proportion of schizophrenia patients. Long-standing arguments concerning "core" or selective deficits against a background of general impairment are unlikely to be resolved until researchers routinely calibrate the difficulty and complexity levels of their tests before applying them to clinical populations.

Moderator variable analysis is an essential part of meta-analytic investigation because it assesses the contribution of study differences in sample attributes, design, and instrument features to effect size heterogeneity and replication. However, moderator analysis is dependent on the individual reporting practices of investigators and presumably the expectations of journal editors and reviewers. In terms of moderator variables, the published literature on neurocognition in schizophrenia is surprisingly limited and often inadequate. We found that even basic participant attributes, such as age, education, and gender composition, were not reported by all studies. Clinical variables of special relevance to schizophrenia, like neuroleptic medication dose, age of illness onset, duration of illness, and frequency of hospitalization were so underreported that we were able to conduct only a limited correlational study of potential moderators, with many correlations missing statistical significance because of inadequate power.

In the moderator variable analysis there were nonsignificant trends that imply possible links between neurocognitive deficits and some demographic and clinical aspects of the patient samples. Thus, higher neuroleptic doses associate with smaller differences between patients and controls on expressive language tests (see Spohn & Strauss, 1989). With a very limited number of studies available for analysis, there is little to suggest an association between chronicity of illness and neurocognitive deficit. A moderate, nonsignificant trend is suggested between frequent hospitalization and executive deficits measured by the WCST. More consistent reporting of these sample attributes by researchers in individual studies will increase statistical power for meta-analytic purposes and facilitate future efforts at understanding how neurocognitive and clinical variables are related in the published literature.

Although many studies did not report IQ along with other neurocognitive variables of interest, there is evidence that at least one commonly used neurocognitive measure in schizophrenia research is tied to general intellectual ability. Differences between patients and controls in terms of WCST performance were correlated with differences in IQ scores and in IQ effect sizes, which is consistent with earlier studies suggesting a relationship between intelligence and WCST performance in neuropsychiatric patients (e.g., Heinrichs, 1990).

The finding that patient samples that differ in general intelligence from their control samples are also likely to differ in WCST performance reinforces the question of whether general ability factors underpin apparently selective neurocognitive deficits in schizophrenia. In other words, the illness may bring with it a generalized cognitive deficit that manifests itself in putatively different tests that nonetheless depend heavily on general ability for successful performance. Although defective WCST scores may coexist with average IQs in many neurological disorders, this does not appear to be the case in schizophrenia (see Heaton et al., 1993).

It is also possible that functions other than intelligence contribute to patient performance on a broad variety of tasks. Neurocognitive tests probably measure several component processes at the same time. These processes may be weighted differentially in terms of particular task demands. Hence, the final score or performance value may reflect this complexity rather than just the preservation or impairment of a supposedly "pure" function. For example, verbal memory tasks, such as the California Verbal Learning Test (Delis et al., 1987), require not only intact verbal memory for success but also basic language comprehension, auditory attention, and expressive language. Similarly, many tasks in addition to those subsumed under the label of "attention" require attention-related skills. It is not known to what extent shared functions contribute to neurocognitive test performance and give rise to spurious impressions of selective impairment. However, concurrent reporting of general and more specific neurocognitive test scores, along with data on task difficulty, will make it easier to assess the credibility of putatively selective deficits (Chapman & Chapman, 1978, 1989).

Neuropsychological research has traditionally used matching strategies to demonstrate selectivity of deficit. In this way, a lesion group and a control group are matched in terms of general intellectual ability, education, and other attributes that may confound the demonstration of specific deficit. In the case of schizophrenia, however, the nature of the illness and its onset in late adolescence or early adulthood prematurely curtail academic achievement. Hence, the use of educational attainment or IQ for matching purposes can lead to spuriously low estimates of general ability in schizophrenia populations. With this problem in mind, some investigators have argued for the use of measures for matching purposes, such as oral reading or spelling tests, that are less sensitive to the illness (Kremen et al., 1996). This is an excellent suggestion for future research on neurocognitive aspects of schizophrenia. In our review we found that the overwhelming majority of studies matched patients and controls on age (98%) and gender composition (82%). Only 19 studies even reported data from the NART, a measure of putative premorbid intellectual ability that may lend itself to matching purposes. Such tasks are clearly not represented widely in matching paradigms in the published literature. This relative neglect of matching issues from a neuropsychological standpoint may reflect the fact that most studies employ neurocognitive measures within a framework that includes neuroimaging techniques, symptom scales, and psychophysiological measures. The psychometric requirements of a valid patient-control comparison in neurocognitive terms may be overlooked in the concern for the needs of a valid neurobiological comparison. Nevertheless, future research on the neuropsychology of schizophrenia will be more informative to the extent that the vagaries of Meehl's (1970) "matching fallacy" are addressed.
In conclusion, we addressed the strength and consistency of neurocognitive findings in schizophrenia in our quantitative review. This literature has accumulated in tandem with the view that schizophrenia is a neurological disorder that manifests itself in behavior. Our quantitative synthesis confirms that a large proportion of this patient population is impaired on standard neurocognitive tests, and the evidence suggests that any selective deficits in functions like verbal memory are relative and exist against a background of general dysfunction. This picture is complicated by issues of test equivalence within and between neurocognitive functions, by the need to improve reporting of potential moderator variables, and by the need to develop testable illness models that can account for why substantial numbers of patients are indistinguishable from controls on many tests.

References

References marked with an asterisk indicate studies included in the meta-analyses.


NEUROCOGNITIVE DEFICIT IN SCHIZOPHRENIA


NEUROCOGNITIVE DEFICIT IN SCHIZOPHRENIA


The National Adult Reading Test (NART) test


Oltmanns, T. F. (1978). Selective attention in schizophrenia and


*Oltmanns, T. F., Murphy, R., Berenbaum, H., & Dunlop, S. R. (1986). Lexical cohesion and


O’Carrol, R. (1992). Selecting controls for schizophrenia research studies: The use of the National Adult Reading Test (NART) is a measure of premorbid ability. Schizophrenia Research, 8, 137–141.


NEUROCOGNITIVE DEFICIT IN SCHIZOPHRENIA


Stroop, J. R. (1935). Studies of interference in serial verbal
Schizophrenia: Empirical
Chicago: Author.
SPSS reference guide.
Sonobe, N. (1993). Eye movements, verbal material recall and


**Appendix**

**Key Words Used in Computer Database Search**

- Schizophrenia Memory
- Schizophrenia California Verbal Learning
- Schizophrenia Wechsler
- Schizophrenia Nonverbal Memory
- Schizophrenia Selective Memory
- Schizophrenia Global Memory
- Schizophrenia Verbal Memory
- Schizophrenia Complex Figure
- Schizophrenia Face Recognition
- Schizophrenia Spatial Perception
- Schizophrenia Line Orientation
- Schizophrenia Transfer
- Schizophrenia Purdue
- Schizophrenia Motor Dexterity
- Schizophrenia Finger Tapping
- Schizophrenia Motor Disorder
- Schizophrenia Stroop
- Schizophrenia Digit Span
- Schizophrenia Continuous Performance
- Schizophrenia Backward Masking
- Schizophrenia Trails
- Schizophrenia Trail Making
- Schizophrenia Wisconsin Card Sorting Test
- Schizophrenia IQ
- Schizophrenia Verbal IQ
- Schizophrenia Performance IQ
- Schizophrenia Vocabulary
- Schizophrenia Block Design
- Schizophrenia Word Fluency
- Schizophrenia Token
- Schizophrenia Comprehension
- Schizophrenia Affect Recognition
- Schizophrenia Dichotic Listening

Received May 5, 1997
Revision received October 15, 1997
Accepted October 20, 1997