how to test hypotheses about schizophrenic thought disorder*

Loren J. Chapman and Jean P. Chapman

The research literature on schizophrenic cognition has become voluminous. A number of relevant published experiments have tested each of the dozen or so major competing theories of the nature of schizophrenic cognition. The lack of consensus concerning the nature of schizophrenic thought disorder, after many hundreds of controlled studies, indicates that indiscriminate research is not a magical solution. A major reason for the lack of progress on this problem has been the use of inappropriate research designs. We shall examine the difficulties of the more popular designs as well as suggest better ones.

Traditionally, clinicians have described two aspects of schizophrenic thought disorder: the content of schizophrenic thought, and the form of that thought. A description of the content of schizophrenic thought is a description of the themes of the patients' verbalizations. For example, a schizophrenic patient may show an abnormal preoccupation with dependency and the mother-child relationship. Such a preoccupation might be termed deviant content of thought. In contrast to this deviant content is formal disorder. "Formal thought disorder" refers to the ways in which deviant thought differs in form or structure from that of normal people. For example, if an observer reports that a patient free associates as he speaks, he is reporting a deviancy in the form of the patient's speech, and presumably his thought. Since most theories of schizophrenic thought disorder concentrate on form rather than content, we will also.

The Importance of Defining and Measuring Schizophrenic Thought Disorder

The nature of schizophrenic thought disorder has importance far beyond its curiosity value. Most writers agree that the central defining symptom of schizophrenia is thought disorder, yet different clinicians who use this criterion show low reliability in deciding which newly admitted psychiatric patients should be considered schizophrenic. If one could describe the thought disorder and measure it, the clinician might then be better able to establish objective criteria for diagnosing schizophrenia.

An even more important reason to study the symptoms of schizophrenia is to help locate its cause. Most of the classic issues concerning the nature of schizophrenic thought disorder have implications about the cause of the disorder. Some investigators have built fairly specific theories of organic etiology around specific alleged cognitive deficits in schizophrenia. For example, several investigators who believe they have demonstrated a deficit of attention have inferred a probable defect of the reticular formation. Other investigators point to evidence that the thought disorder varies with the emotional content of the material and with the patient's relation to the examiner as support for the hypothesis that schizophrenia has a motivational origin.

Other inferences about etiology of schizophrenia have been based on the alleged similarities of schizophrenics to some other group. These theories have been built on the similarities of schizophrenic thought to that of children, of drug-induced psychoses, and of various other schizomimetic conditions. Many investigators have noted the similarity of LSD, mescaline, or amphetamine psychosis to schizophrenia and inferred that schizo-

---

*This condensation of chapters 3 and 4 of the authors' 1973 book entitled Disordered Thought in Schizophrenia appears through the courtesy of Prentice-Hall, Inc., Englewood Cliffs, N.J. 07632. Address reprint requests to Dr. Loren J. Chapman, Department of Psychology, University of Wisconsin, Madison, Wis. 53706.
Phrenia has a similar etiology, or at least similar biochemical events. Other clinicians, in turn, have objected that the thought disorder and other symptoms seen in drug-induced psychoses are not truly like those of schizophrenia. Until we can better identify and measure the symptoms, the similarities and differences relevant to these etiological theories cannot be evaluated adequately.

The description and measurement of symptoms is also the first step toward isolating types of schizophrenia. The marked heterogeneity of symptoms of schizophrenia suggests that the disorder is not unitary, but subsumes two or more disorders that have not yet been discovered. The history of progress in psychopathology has been one of successive discoveries of causes of disorders that were first isolated by means of symptom syndromes. This sequence is illustrated by the achievements with respect to general paresis. Clinicians recognized paresis as a disease syndrome in the 19th century. They described the distinguishing symptoms of paresis and were able to use the symptoms to separate many cases of paresis from other psychoses. Research workers sought the cause of paresis by studying groups of patients who were clearly paretic, with the result that Noguchi and Moore isolated the syphilis spirochete in the brains of paretics. Research workers probably would have had difficulty discovering the role of syphilis if they had not first been able to isolate a group of patients with paresis.

Just as the cause of general paresis was found after paretics were first identified by their behavior and isolated from other psychotics, so the causes of types of schizophrenia might be found once the types are isolated. The symptoms that distinguish subvarieties of schizophrenia (if there are such) are more elusive than those of paresis and have not been identified by raw clinical observation.

A reasonable approach to discovering the causes of schizophrenia, therefore, is the attempt to identify the relevant symptoms by experiments and to measure them. Much of our current research has been devoted to this problem of identification and measurement of cognitive symptoms.

The Distinction Between Positive and Negative Symptoms

Hypotheses about the nature of schizophrenic thought disorder are of two broad types—those concerned with negative symptoms, and those concerned with positive symptoms.

**Negative Symptoms**

A hypothesis of a negative symptom of thought disorder is a statement that there is a deficit in an ability. Examples of such hypotheses are that schizophrenics are deficient in ability to interpret proverbs, or that they cannot reason as effectively as normal subjects. A hypothesis of a positive symptom is a statement that some specific kind of thinking replaces normal thinking. It is a statement of what a patient does, rather than what he fails to do. An example might be that schizophrenics respond to proverbs with inappropriately literal interpretations, or that they free-associate excessively, or that they have delusions. Positive symptoms are the most striking features of schizophrenic thought disorder. Schizophrenic conversation and performance is in most cases not an empty inadequacy; often it is, instead, a rich profusion of bizarre behaviors. Positive and negative symptoms are usually, but not always, found together in schizophrenia, and sometimes the observer's choice between calling attention to one or the other seems almost arbitrary. For example, if a patient is unable to define the meaning of a word, he shows a negative symptom, or deficit. If he free-associates to the word, he shows a positive symptom. Positive symptoms always assume a negative symptom, but the reverse is not always true—a patient who does not grasp the meaning of a word might simply fail to respond to the request to define the word, rather than offer a deviant response.

Most clinical descriptions of schizophrenic thought disorder are simple statements of either positive or negative symptoms. Kraepelin's descriptions are a good example. Descriptions of this sort describe how schizophrenics differ from normal people. Yet, by inference, each statement of a specific symptom implies much more than its mere occurrence. The statement of a symptom implies that the symptom is more important than some other possible symptoms for describing schizophrenia. Such an implication is especially clear in the case of negative symptoms. For example, if a psychopathologist states that schizophrenics have lost the ability to form concepts, he implies that they have lost this ability more than some other abilities. In other words, he implies a differential deficit. Because schizophrenics are deficient in almost every behavior that requires a voluntary response, meaningful measurement...
of deficit must always be in terms of differential deficit. Differential deficit can be established by use of a second task that is used to measure an ability the investigator believes is lost less than the ability measured by the experimental task.

The demonstration of a single differential deficit, using a single control task, is a meager achievement and is only the beginning of an adequate description of the schizophrenic loss. The achievement is especially meager when the control task is vocabulary because vocabulary is better retained in schizophrenics than most other abilities. A description of schizophrenic differential deficit might more reasonably rely on several control tasks. Nevertheless, to simplify our discussion we shall focus on research designs that use single control tasks. The principles exemplified in these designs can readily be applied to more complex designs.

Positive Symptoms

Descriptive statements about single positive symptoms in schizophrenia tell more about patients than statements about single negative symptoms. Clinically, schizophrenics do not show every conceivable positive symptom, although they do show almost every conceivable negative symptom. Furthermore, most schizophrenic positive errors do not reflect mere inaccuracy, that is, do not result directly from negative symptoms. If schizophrenics’ positive errors resulted from inaccuracy alone, mental defectives and dull normal people would also show these bizarre symptoms.

Despite such clinical realities, effective measurement of most positive symptoms in schizophrenia has much in common with effective measurement of deficit. Positive symptoms can be separated from the effects of lowered accuracy only in designs that control for effects of lowered accuracy. Consider the following rationale for this view.

Investigators of a positive cognitive symptom almost always choose a task that has a propensity to evoke, from normal subjects, the positive error in question. For example, a test of interpreting proverbs is used to investigate concrete thinking because the proverb test tends to evoke concrete interpretations. As expected, Gorham (1956) found that low-scoring normal subjects made many more concrete interpretations of proverbs than more accurate subjects.

Because the positive errors of interest on most tests increase with inaccuracy both for normal subjects and for schizophrenics, the higher number of positive errors by the schizophrenics probably results to some extent from their lower accuracy. Therefore, measurement of propensity toward a positive error requires control for the effects of lowered accuracy by means of a second measure.

The Measurement of Differential Deficit and Differential Occurrence of Positive Symptoms

The measurement of either differential deficit or differential occurrence of positive symptoms requires at least two measures or scores. Therefore, the investigator must consider the characteristics of the two scores that might affect differential performance of groups.

Reliability, Item Difficulty, and Mean Difference Between Groups

Performance scores on a test are usually viewed as consisting of two components: true score and errors of measurement. The true score is the mean value of a large number of obtained scores that might hypothetically be obtained by an infinite number of parallel tests of the same person. For all types of test reliability, the higher the reliability coefficient, the smaller the errors of measurement and the stronger the relationship between obtained scores and true scores. The correlation of a test score with the hypothetical true score is conventionally considered to be the square root of the reliability coefficient (Gulliksen 1945). The higher the correlation between test scores and true scores, or the higher the reliability, the better the test discriminates between people with different true scores, and the larger the mean difference between groups in obtained score.

The degree of schizophrenic deficit on an ability is inferred, of course, by the mean difference between schizophrenic and normal groups on a test of the ability. It follows that if schizophrenics are as inferior to normal subjects on one ability as on another, but the test that is used to measure one of the abilities is more reliable than the test for the other, a greater deficit will be found on the more reliable measure.

The mean difference between groups is also determined by item difficulty, in part because item difficulty affects task variance: the larger the task variance, the larger the separation of groups of high and low ability. Two tasks that differ in difficulty usually differ in the
mean-difference score between subgroups of subjects who differ in the abilities measured by the tasks. This principle is most immediately obvious for tasks that are either so extremely easy that most subjects receive perfect scores, or so extremely difficult that most subjects achieve zero accuracy. For example, the task of counting one's fingers would not discriminate third graders from fifth graders because almost all students would pass all items. Likewise, a task consisting of problems in advanced calculus would fail to distinguish the two grades because almost all students would fail all items. A task that distinguished the groups better would be one of intermediate difficulty, such as adding columns of numbers. On extremely difficult tasks, errors of all subjects are near a ceiling of 100 percent errors, so differences in ability cannot be detected. Similarly, extremely easy tasks would not detect differences in ability, because all subjects are near a floor of no errors.

Effects of a ceiling or floor of errors on separating of groups are more widely understood by investigators than are differences in difficulty level of less extreme degree. On a dichotomously scored free-response task that cannot be answered correctly by guessing and is composed of items that tap the same ability, items that yield in the neighborhood of 50 percent errors distinguish among ability levels better than either more difficult or easier items. This is one of the most firmly established principles of psychometric theory and practice. Although most writers who discuss the principle attribute it to Lord (1952b), who provided the definitive proof, Lord acknowledged that similar conclusions had been reached by a number of earlier writers. Lord showed that 50 percent accuracy for the entire group tested is the most discriminating level of item difficulty across a wide range of mean interitem correlations.

If one of the two groups is larger than the other, the difficulty level that yields the largest difference between the two groups will not coincide precisely with the 50 percent level. Such discrepancies are, however, small. In the usual study comparing schizophrenic and normal subjects, even if one group is two or three times as large as the other, the point of maximal separation of groups will differ by only a few percentage points from 50 percent difficulty for all the subjects grouped together.

The effect of difficulty on separation of groups by mean score can be illustrated by comparing two age groups of children on vocabulary items that differ in difficulty. Figure 1 shows data on accuracy of a group of third-grade children and a group of fifth-grade children on the vocabulary items of the Stanford-Binet Intelligence Scale. The abscissa of figure 1 presents six sequential groups of vocabulary items, each representing a different level of difficulty (as measured by percentage accuracy for third and fifth grades combined). Each point on the abscissa represents several vocabulary items of a given level of difficulty. Against this abscissa is plotted the percentage of third-grade children and the percentage of fifth-grade children passing each set of items. (If the two grades were combined, the resultant curve would represent a variable plotted against itself.) The curves for the two grades tend to join at the extreme ends of the dimension of difficulty, and they attain greatest separation for items at about the 50 percent level of difficulty for the two groups combined. This relationship between difficulty and difference between groups may be generalized to any two groups that differ in accuracy, including schizophrenic and normal subjects.

Tasks must be matched on both difficulty and reliability to measure differential deficit.

Matching Tasks on Reliability

Reliability has come to have a variety of related meanings but, in general, refers to the extent to which measurements can be repeated with the same results. In classic mental test theory, reliability is the correlation between scores on two parallel tests that measure the same thing. In practice, parallel tests are very difficult to construct, so reliability is usually estimated by other methods. One estimate of reliability is test-retest reliability, which indicates the extent to which subjects maintain the same relative positions in the group on two occasions. A second type of estimate of reliability is split-half reliability, which is the correlation of subjects' scores on one-half of a test with their scores on the other half, with a correction for double length. This coefficient is a measure of consistency of items; that is, the extent to which the various items of a test tend to measure the same trait or ability. A disadvantage of using a measure of split-half reliability is that the items can be divided into a large number of different halves, with each division of items yielding a different value. For this reason, internal consistency of items is usually measured by a statistic called "coefficient alpha," which is often referred to as Kuder-Richardson Formula 20 or...
the Hoyt reliability coefficient. Coefficient alpha is the average value of all possible split-half reliability coefficients for any given set of scores. Coefficient alpha is usually the best available estimate of the correlation of the test with a second parallel test if the test is homogeneous in content. However, coefficient alpha underestimates reliability in a test that is factorially heterogeneous. It overestimates reliability for a speeded test in which a significant portion of the variance is attributable to some subjects' not finishing the test. One might reasonably use coefficient alpha for matching on reliability if tests are unspeeded and are equivalent in their factorial homogeneity.

**Matching Tasks on Difficulty**

One should match tasks on the distributions of difficulty as well as on the mean value. To take an extreme case, if half the items were so extremely easy that everyone passed them and the other half were so extremely difficult that everyone failed them, the task would not distinguish groups of differing ability although the mean number of errors per item would be about 50 percent. Tasks yield highest overall discrimination between subjects of different levels of ability when all items are of uniform 50 percent difficulty (Cronbach and Warrington 1952, Horst 1966, and Lord

---

Figure 1. Performance of third- and fifth-grade children on Stanford-Binet vocabulary items plotted against item difficulty for the combined groups.
1952b). If, however, a test is to be used for subjects both at low levels of ability and at very high levels of ability, a test's usefulness at those particular levels is improved by use of items that are at 50 percent difficulty at each level. In other words, a range of difficulty of items is required. Applications of this principle are seen in standard intelligence tests.

The matching should be not just on mean item difficulty but also on the variance and the shape of the distribution of item difficulty. Otherwise, the tests might, in effect, remain unmatched for high-scoring subjects or for low-scoring subjects.

In multiple-choice items, guessing raises accuracy, so that the item accuracy level that yields the largest difference between groups is somewhat higher than 50 percent. No exact values can be designated because they are affected slightly by the intercorrelations among items and by nonrandom patterns of guessing by subjects who do not know the answer. Nevertheless, in constructing multiple-choice tests, just as in constructing free-response tests, one may obtain differing degrees of discrimination between groups by varying the level of item difficulty.

The standardization sample for the matching of tests should consist of normal subjects only. It would not be desirable to match tasks using the two groups combined, because any true differential deficit due to pathology will affect the amount of variance on the two tasks, and this in turn will affect their reliabilities. Thus, if normal subjects score equally well on two tasks but schizophrenics score much lower on one of the tasks, the task with the greater schizophrenic deficit will have the larger variance for the two groups combined, and hence the largest difference between groups. Matching for the schizophrenic and normal subjects combined would eliminate some or all evidence of differential deficit in ability. Therefore, tasks should be matched using only normal subjects of a wide range of ability.

One might object that people at different levels of general ability have different patterns of specific abilities. Retardates, for example, have differential deficits, that is, a greater inferiority to normal people on some abilities than on others. To match two tests on difficulty using a sample that includes such subjects, one must select items so as to compensate for this differential deficit. Some readers might be uneasy about using such tests for measuring schizophrenic differential deficit because the tests have been matched by adjusting them for another differential deficit. Our view is that tests that are matched by adjusting for differential deficits of nonpsychotics of low intelligence are exactly what are needed. Since the goal is to identify a specific deficit that will distinguish between schizophrenics and others who tend, in general, to be low scorers, tasks matched on difficulty for nonpsychotic subjects of low ability are required.

It might be reasoned that finding a sample of normal subjects who score as low as the schizophrenics would make the whole enterprise of matching tests unnecessary, since the purpose of matching tests is to avoid mistaking generalized deficit for specific differential deficit in ability. Ultimately, however, one must conclude that finding low-scoring normal subjects is not a reasonable alternative to matching tests, for the reasons presented below in a discussion of matching schizophrenic and normal groups.

**Practical Steps to Achieve Matched Tasks**

The job of matching on reliability as well as on mean, variance of item difficulty, and shape of the distribution of item difficulty is not prohibitively laborious. Matching on these variables can be achieved for most pairs of tasks by giving large, and preferably equal, numbers of items of the two types to normal subjects of widely ranging ability and then selecting pairs of items of the same difficulty and item-scale correlation. The matching should then be cross-validated on a second sample of normal subjects.

If, despite close matching, tasks differ very slightly on mean or variance for normal subjects, scores may be converted to standard scores on the basis of the performance of the normal standardization sample. This conversion removes differences between the tests in mean and variance and facilitates comparison. It does not, of course, remove differences between tests in power to distinguish the more able from the less able subjects. For example, the subtests of the Wechsler Adult Intelligence Scale, although expressed in standard score units, differ greatly in power to distinguish between the more able and the less able subjects.

**Illustrative Example**

The need for equivalent reliability of tasks has recently been illustrated by the present writers...
SCHIZOPHRENIA BULLETIN

The test used was a 114-item multiple-choice analogies test with items of gradated difficulty, cast in the following format:

Fire is to red as
Banana is to
1. yellow
2. blue
3. soft
4. green

The test was given to 49 severely disturbed schizophrenics and 206 normal subjects. From the pool of 114 items, various subsets of items were drawn to compare schizophrenic and normal performance. These items were then arbitrarily divided into two forms. Form A consisted of the 62 items for which the first word of the analogy began with a letter between A and L, whereas form M consisted of the remaining 52 items for which the first letter of the word fell between M and Z. As expected, normal subjects scored about the same on the two forms, and the schizophrenics showed about the same performance deficit on one form as on the other.

A demonstration of the effects of difference in reliability on performance deficit was then made by drawing subtests from the two forms with identical means but different reliabilities for normal subjects. Coefficient alpha (Kuder-Richardson Formula 20) was used as the measure of reliability. As can be seen in table 1, a pair of subtests was drawn with coefficient alpha values of .80, one subtest from each form. Another pair of subtests was drawn with lower coefficient alpha values, .47 for the subtests from form A and .49 for the subtest from form M. All four subtests were closely matched on mean accuracy. As expected, the more reliable subtests had the lower variances of item difficulty and the higher test variances. Table 1 gives the mean accuracy of the schizophrenic and normal subjects for each subtest. As predicted, the schizophrenics showed a greater performance deficit on the subtests with the higher reliabilities, regardless of the form used. A comparison of the schizophrenic performance on the less reliable form M subtest with the more reliable form A subtest yielded poorer schizophrenic accuracy on form A than on form M ($t = 4.89; p < .001$). In contrast, the more reliable form M subtest and the less reliable form A subtest yielded the opposite finding, with the schizophrenics scoring lower on form M than on form A ($t = 4.32, p < .001$). These results demonstrate the ease with which differences in reliability may artifactually yield an apparent differential deficit in ability. If, as in the usual study of differential deficit, the two forms were designed to measure different abilities, the investigator could easily conclude from performance on either one of these pairs of subtests that he had demonstrated a differential deficit in ability.

In the same article, we also demonstrated that differential difficulty of items can produce differences in power of tests to distinguish between differing ability levels despite equivalent reliabilities (both difficulty and reliability again measured by normal performance alone). In this study, a free-response vocabulary test was given to 52 schizophrenics and 56 normal control subjects, some of the items being affect laden.

### Table 1. Performance of normal and schizophrenic subjects on tests of high and low reliability.

<table>
<thead>
<tr>
<th></th>
<th>High reliability</th>
<th>Low reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Form A</td>
<td>Form M</td>
</tr>
<tr>
<td></td>
<td>$\bar{x}$</td>
<td>$s$</td>
</tr>
<tr>
<td>Normal control group</td>
<td>6.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Schizophrenic group</td>
<td>3.0</td>
<td>2.8</td>
</tr>
<tr>
<td>Number of items</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Coefficient alpha</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>Variance of item difficulty</td>
<td>.02</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Performance of normal and schizophrenic subjects on tests of medium and high difficulty.

<table>
<thead>
<tr>
<th></th>
<th>Medium difficulty</th>
<th>High difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Emotional</td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td>$\bar{x}$</td>
<td>$s$</td>
</tr>
<tr>
<td>Normal control group</td>
<td>6.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Schizophrenic group</td>
<td>3.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Number of items</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Coefficient alpha</td>
<td>.68</td>
<td>.04</td>
</tr>
<tr>
<td>Variance of item difficulty</td>
<td>.04</td>
<td></td>
</tr>
</tbody>
</table>

(emotional) and others being affectively neutral. Table 2 shows the performance of both schizophrenic and normal subjects on two 10-item emotional and two 10-item neutral subtests with coefficient alpha values around .70. Two subtests, one emotional and one neutral, were chosen from very difficult items, and two others, one emotional and one neutral, from items of medium difficulty. The very difficult items produced less than 10 percent accuracy both for the normal subjects alone and for schizophrenic subjects alone, whereas the medium-difficulty items produced accuracy for normal subjects of about 65 percent and accuracy for the two groups combined of 51 percent. The schizophrenics, as expected, did not differ significantly from the normal subjects on the very difficult subtests ($t = .32$ for emotional and .72 for neutral), but did show a larger deficit on the medium-difficulty subtests ($t = 4.76$ for emotional and 5.84 for neutral; $p < .001$ for both). Thus, if subtests of differing levels of difficulty are used, the investigator who compares groups on one emotional and one neutral test might conclude that he had found a greater schizophrenic deficit on emotional than on neutral vocabulary, or vice versa, depending on the different levels of difficulty employed in the study.

Methodological Problems Resulting from Matching Schizophrenic and Normal Groups

A number of commonly used designs involve matching schizophrenic and normal groups on one or another variable. This matching is done for one of two similar purposes: to rule out a variable as a possible source of spurious results, or to establish differential deficit. For either purpose, investigators match the groups on a control task and then compare the groups on an experimental task. If the schizophrenics perform worse than normals on the second task, the investigator usually concludes either that 1) the schizophrenics have lost the ability measured by the experimental task, and this loss cannot be attributed to loss of the ability measured by the control task, or 2) the schizophrenics have lost the ability of the experimental task more than they have lost the ability of the control task. The control task is typically an IQ test, a vocabulary test, or some other task of special interest to the investigator. This procedure is plagued by several methodological difficulties and a major problem in matching subgroups is statistical regression.

Statistical Regression as a Hazard in Matching Groups

Disturbed schizophrenics, as a group, perform less accurately than normal subjects on almost any task that requires a voluntary response. Nevertheless, the performance scores of schizophrenic and normal groups almost always overlap. It is possible, therefore, to match subsamples on accuracy by choosing atypical subjects—that is, the best-scoring schizophrenics and the poorest-scoring normal subjects. The extreme scorers on one task will not, however, always be the extreme scorers on the second task. The lack of perfect correlation of two tasks means that subgroups of schizo-
phrenics and normal subjects matched on one task will probably differ on a second. Each subgroup will have scores on the second task closer to the mean of their own group. This phenomenon is called "statistical regression." Sir Francis Galton (1885) coined the term "regression" from the fact that the scores of each subgroup "regress" toward the mean of the total group from which it was drawn. Statistical regression produces artifactual findings which lead an investigator to conclude that there is a greater schizophrenic deficit on an experimental task than on a control task. Statistical regression is likely to occur not only in comparing matched groups of schizophrenic and normal subjects but also in comparing any two matched subgroups drawn from larger groups of subjects who differ in general performance level. Regression effects depend on errors of measurement as well as on imperfect correlation of true abilities.

Selection of Atypical Subjects as a Hazard in Matching Groups

Meehl (1970 and 1971) has pointed out that matching groups on one variable may systematically "unmatch" the groups on other variables. This systematic unmatching is especially a problem for studies in which schizophrenic and normal groups are used if schizophrenia affects the variable on which the experimenter matches. Matching on current functioning IQ, for example, unmatches on premorbid IQ. If groups are matched on current functioning IQ, the schizophrenic group will have a higher premorbid IQ than the current IQ of the normal group. This unmatching on premorbid IQ is a serious defect for testing many hypotheses, because true effects of schizophrenia may be obscured.

Another difficulty caused by matching groups of schizophrenic and normal subjects on current test score is that matching eliminates those schizophrenics who show the most thought disorder. One finds, for almost any measure of schizophrenic thought disorder, that some of the schizophrenics perform at a normal level. The schizophrenic subjects who score in the normal range on one test include more than a chance proportion of patients who show little thought disorder on other tests. Such selection may lead to an erroneous conclusion that schizophrenics do not show some hypothesized cognitive defect. It should be noted that elimination of disturbed subjects tends to obscure true differences between groups, while statistical regression creates spurious differences.

One cannot simultaneously match normal subjects and disturbed schizophrenics on current functioning IQ and match the normal subjects' present functioning IQ with the schizophrenics' premorbid IQ. If an investigator believes he has done so, he must either have a poor estimate of premorbid IQ or have "schizophrenics" who are not deeply disturbed. Yet, both present functioning and premorbid IQ must be controlled. The solution is to test for the effects of each separately.

Analysis of Covariance as an Alternative to Matching Groups

Occasionally, instead of selecting subgroups matched on relevant variables such as on IQ score, investigators have used analysis of covariance in an attempt to remove the effects of differences in scores on one test from scores on a second test. In other words, they wish to discover whether the two groups would differ on the second test if they did not differ on the first test. Such a question cannot be answered by any form of statistical analysis. The only legitimate use of analysis of covariance is for reducing variability of scores in groups that vary randomly. Its use is invalid for preexisting disparate groups that differ on the variable to be covaried out. Schizophrenic and normal groups are disparate groups. Recently, Lord (1967) has forcefully pointed out that if one group differs from a second group on each of two variables, covarying out the effects of one variable in order to obtain predicted scores for each group on the second variable will not yield equivalent "corrected" scores for the two groups.

Research Designs for Identifying Differential Deficit and Differential Occurrence of Positive Symptoms

To identify and measure schizophrenic cognitive symptoms, investigators use research designs in which schizophrenics are compared with normal subjects. Most investigators have used one or another of a very few designs.

A researcher uses a design to justify his making the leap from a finding of some schizophrenic deviant performance to a statement of a specific negative or positive cognitive symptom. The most widely used
designs are not the best ones. They yield valid inferences only in certain unusual circumstances, which will be described below.

We shall use published studies to illustrate the applications of designs and, whenever possible, the same studies to point out interrelations among designs. The same study may be used to illustrate more than one design, since different designs are often only alternative modes of analysis of the same data.

The Design of the Unmatched Control Task

Negative Symptoms

Recognizing the deficiencies inherent in using a single task to demonstrate differential deficit, investigators have added a control task which, they believe, measures some ability that is less impaired than the ability measured by the experimental task. Because of the generalized deficit characteristic of schizophrenics, the investigator using this design seldom expects schizophrenic subjects to perform as well as normal subjects on the control task, but he assumes that their generalized deficit affects the experimental and the control tasks equally. He further reasons that if the ability measured by the experimental task is lost to the same degree as the ability measured by the control task, the magnitude of the difference between schizophrenic and normal subjects will be the same for the two tasks. The investigator computes for each subject a score of the difference in accuracy between the two tasks—that is, accuracy score on the control task minus accuracy score on the experimental task—and he compares the normal and schizophrenic groups on this difference score. He predicts a larger difference between schizophrenic and normal subjects on the experimental task than on the control task. If he finds the hypothesized difference, he concludes that he has demonstrated differential deficit in ability.

The conclusion is seldom justified. In most published studies using this design, the investigator chose easy items for his control task and harder items for his experimental task. The most common finding of such studies is that normal subjects make more errors on the experimental task than on the control task, but the difference is greater for the schizophrenics. The investigator invariably concludes from his finding that he has demonstrated a differential loss of ability. As discussed above, however, two tasks that differ in difficulty for normal subjects may usually be expected also to differ in the mean difference score by which they distinguish two groups that differ on both abilities. The investigator could obtain the very opposite results—that is, a greater deficit on the control task than on the experimental task—by selecting easy items for his experimental task and the items for his control task from the middle range of difficulty.

The danger of this artifact may be illustrated by a study of Truscott (1970), who compared schizophrenic and normal subjects on four tasks, including one that required recall of random word strings and one that required recall of normal sentences. The groups differed less on recall of random word strings, on which both groups achieved less than 20 percent accuracy, than on recall of normal sentences, on which the two groups’ combined accuracy was much closer to 50 percent. Truscott believed that the difference in performance on the two tasks reflected a differential loss of abilities and concluded that the schizophrenics showed a lack of facilitation of recall by normal syntactic constraints. Her results, however, may instead have been at least in part an artifact of the levels of difficulty of the items she chose for her two tasks. If Truscott had used shorter word sequences, she would presumably have sampled items at the easy end of the difficulty dimension and might well have found the opposite result—that is, a greater deficit in recalling random word strings than in recalling sentences.

For some tasks, adequacy of performance is defined by criteria other than number of errors. Prominent examples of other criteria for difficulty are speed of response (as in reaction time tasks) or trials to criterion (as in verbal learning). When performance is evaluated by such criteria, the probable direction of the artifact resulting from differences in difficulty is often ambiguous, but one may nevertheless be sure that differences in difficulty produce differences in mean score between the more able and the less able subjects.

There is one situation in which unmatched tasks may give legitimate evidence of differential deficit in ability. If the control task would be expected, on psychometric grounds, to yield the larger differences between groups of differing ability but the experimental task, nevertheless, yields the greater difference between schizophrenic and normal performance, one must attribute the differential performance to the effects of schizophrenia...
rather than to characteristics of the tasks. For example, the Stanford-Binet or the Wechsler Adult Intelligence Scale, or even one of the more reliable subtests of a Wechsler scale, will usually be more reliable than a more casually constructed experimental task of similar length.

**Positive Symptoms**

Investigators often seek to demonstrate that schizophrenics make one positive error more often than a second. The two errors may be evoked by two different tasks, but more often by a single task. For example, several investigators have compared the frequencies of two positive errors on an object-sorting test. The logic of the design is similar whether the two scores are obtained from two tasks or one task.

The use of an unmatched control task suffers from the same defects for demonstrating positive symptoms as for negative symptoms. If task materials are constructed so that the tasks’ propensities to evoke two positive errors differ in strength, the schizophrenic group, being less accurate, can be expected to be more vulnerable to the stronger of the two positive errors. An example is seen in one of the studies reported by Chapman, Chapman, and Miller (1964). They presented both schizophrenic and normal subjects with names of conceptual categories and names of objects that might fall in the categories. One subtask, for example, required the subject to sort a series of cards into two boxes, putting cards with names of things that have legs into one box and cards with names of things that do not have legs into the other box. The cards were of three types: names of animate creatures that have legs (such as horse and lion), names of inanimate objects that have legs (such as table and sofa), and irrelevant items. Both schizophrenic and normal subjects excluded names of things corresponding to a rare meaning of the instructed concept (e.g., inanimate objects with legs) more than those corresponding to a common meaning (e.g., animate objects with legs), but schizophrenics showed this error pattern to a more marked degree. This apparent demonstration of differential propensity toward a positive symptom might be an artifact of the schizophrenics’ lowered accuracy together with the differential propensity of the task to evoke one rather than the other of the two kinds of errors.

This kind of alternative interpretation potentially applies to measures of any pair of positive symptoms. Even everyday conversation can be viewed as a task that puts demands on the subject and tends to evoke from normal subjects some kinds of errors more than others. For example, normal subjects show associative intrusions based on similarity of ideas more often than they show associative intrusions based on similarity of sound. If schizophrenics show an accentuation of these same error propensities, their greater propensity toward one of these types of associative intrusion rather than the other may be an artifact of their lower ability to converse appropriately.

**The Design of Manipulating a Single Variable**

**Negative Symptoms**

A design closely analogous to that of the unmatched control task is one in which the investigator manipulates an independent variable that appears to tap the hypothesized loss. The experimenter constructs two or more tasks that are alike in every respect but one. The difference between the tasks is the independent variable. For example, Chapman (1958) used this design inappropriately in a test of the familiar idea that schizophrenics have a deficit in conceptual thinking. In a conceptual card-sorting task, he manipulated the difficulty of the correct concept. The hypothesis was that the increase in difficulty of concepts would produce an increase in errors for both normal and schizophrenic subjects, but that the increase would be greater for schizophrenics. To put it another way, the schizophrenic performance deficit would increase in response to the variable of difficulty. This hypothesis was confirmed. Curves representing the finding are shown in figure 2.

The performance data showed a greater schizophrenic deficit in dealing with hard concepts than easy concepts. Such a difference in performance does not, however, warrant an inference that schizophrenics have lost the ability to interpret hard concepts any more than they have lost the ability to interpret easy ones. The design is susceptible to artifactual findings due to an arbitrary choice from among the many possible points on the dimension of difficulty. Most often the investigator chooses points that vary from low to moderate difficulty, as in Chapman’s study discussed above. For such tasks, a divergence such as that shown in figure 2 would be expected for any two groups that differ in level of accuracy. If the investigator instead manipulated the
variable of difficulty within the extreme high end of the range of difficulty, the error curves would, of course, be expected to converge instead of diverge with the increase in difficulty, as shown in figure 1. Such a choice of points on the dimension of difficulty would yield the opposite finding—that is, a greater schizophrenic performance deficit on the easier items than on the harder ones. Independent variables that produce an increase in difficulty for normal subjects will usually produce a successive divergence and convergence of the performance curves for schizophrenic and normal groups.

For these reasons, most studies in which the design of the single variable is used leave unsettled the question whether the variable truly taps a schizophrenic differential deficit.

Positive Symptoms

The design suffers from the same difficulties for positive symptoms as for negative symptoms. The manipulation of the opportunity for giving a positive error should produce more positive errors for schizophrenics than for normal subjects simply because of schizophrenics' lowered accuracy. Chapman (1958), in one part of the conceptual card-sorting study already described, manipulated propensity of the materials to evoke an inappropriate associate response. The independent variable was the strength of associative connection between an associative distracter and a stimulus word. Such a tendency of the task to elicit positive errors cannot be separated from difficulty. Normal

1 From Chapman (1958).
subjects, as well as schizophrenics, produce more associative errors on items that have strong incorrect associative alternatives than on items that have weak ones. This means that the propensity of materials to evoke an associative response is a source of difficulty for normal subjects.

**The Design of Matching Groups on a Control Task**

**Negative and Positive Symptoms**

Sometimes an investigator selects, from the available schizophrenic and normal subjects, subsamples that are matched on performance on his control task. He then compares these selected schizophrenic and normal subjects on his experimental task. The hypothesis is that despite the matching on the control task, the subsamples will differ on the experimental task.

The defects in this design are the same for positive and negative symptoms. The design fails to remedy a lack of matching of tasks. If schizophrenic and normal subjects are matched on their performance on a task with low power to distinguish between different ability levels, they may still differ on a test with greater power to distinguish groups.

This design also suffers from the hazards of any design in which matched groups are used—the hazards of statistical regression and the elimination of the more disturbed patients. The smaller the proportion of subjects who qualify for the matched subgroups, the greater the hazard of statistical regression. For example, Whiteman (1954) tested 31 normal and 31 schizophrenic subjects on both a social task (a conceptual task of interpreting interpersonal interactions) and two non-social tests of conceptual reasoning (a verbal analogies and a pictorial concept-formation test). He found that the schizophrenics made more errors than the normal subjects on both tasks. Nevertheless, the two distributions of error scores overlapped. In one of his analyses, Whiteman selected 12 normal and 12 schizophrenic subjects matched on mean score on the two nonsocial tasks, and found that the subgroup of 12 schizophrenics scored more poorly on the social task than the 12 matched normal subjects. With selection of such atypically scoring subgroups for matching purposes, statistical regression surely contributed to the finding. If Whiteman had used the reverse procedure of selecting subgroups matched on his social task, he probably would have found the schizophrenic subgroup inferior to the normal subgroup on the nonsocial task. Whiteman’s type of analysis would then have yielded the opposite conclusion.

Investigators are most likely to resort to this design when they suspect that their tasks may be unmatched in ways that produce differences in performance, yet the design fails to remedy this defect.

**The Design of the Matched Control Task**

**Negative Symptoms**

One highly appropriate design for investigating differential deficit is that in which the investigator matches an experimental task and a control task on psychometric characteristics that affect the difference in mean score between normal subjects of differing ability levels. As shown previously, tests must be matched on reliability, mean-item difficulty, variance of item difficulty, and shape of the distribution of item difficulty for normal subjects.

The choice of an appropriate nonschizophrenic standardization group is a crucial step in this design. The standardization group should have a wide enough range of ability, and it must contain a substantial number of subjects who score as poorly as schizophrenics. The standardization sample should not, however, include subjects such as brain-impaired patients, who might themselves show psychopathological differential deficit. The most suitable normal standardization group for such a study is one comprising normal and dull normal adults who do not suffer from brain damage or other psychopathology. The equivalence of the two tests must be verified for any particular group of schizophrenic subjects by recomputing the reliability and other characteristics for a normal sample that scores at the level of the schizophrenic group. With such equivalence, any difference between the two test scores for schizophrenic subjects can be attributed to genuine differential deficit in ability.

One must be careful, when matching tests, not to capitalize on errors of measurement. If one achieves the matching by selecting some items and dropping others, errors of measurement will very likely result in spurious matches. The equivalence of tests will not be genuine. This problem is reduced by having good-sized normative
samples for the selection of items, but it is better handled by cross-checking the equivalence of tests on a new group of normal subjects.

We have used this design in several recent studies (Chapman and Chapman 1975, Chapman, Chapman, and Daut 1974, and Rattan and Chapman 1973). For example, Rattan and Chapman (1973) matched two multiple-choice vocabulary subtests, in only one of which the incorrect alternatives included a word that had a strong associative connection with the word to be defined. Using a normal standardization group of a wide range of ability, the investigators matched their “with-associates” and “without-associates” subtests on coefficient alpha, and on mean, variance, and, roughly, shape of the distribution of item difficulty. Despite this matching for normal subjects, the schizophrenics made more errors on the with-associates subtests than on the without-associates subtest. The design made possible a conclusion of differential deficit without any possible artifact from generalized deficit.

Positive Symptoms

The design of the matched control task has apparently not been used in any published investigation of positive symptoms. The design would, however, be valuable for determining which of two positive symptoms occurs more often in schizophrenia. If two such error scores were matched, and if a pathological group differed from a normal group on the difference between the two error scores, one could then conclude that the patients have a deviant differential propensity toward that one of the two positive errors.

The Design of Manipulating Two Matched Variables

Negative and Positive Symptoms

This sophisticated design is an elaboration of the design of the matched control task. The investigator matches not only the two entire tasks but, also, paired subsets of items at various points on the two dimensions of difficulty. The matching of paired tasks on difficulty is accompanied by matching on reliability. The investigator thus manipulates two matched variables and compares their effects on errors made by normal and schizophrenic subjects.

The investigator selects items for the two variables so that the two accuracy curves of normal subjects are equivalent. One task should be as reliable as the other at any given level of accuracy. The tasks are given to a new normal control group as well as to a schizophrenic group to test the hypothesis that an increase in the variable of the experimental task will produce a faster increase in errors for schizophrenics than will an increase in the variable of the control task despite the similar difficulty and reliability for normal subjects. Such a finding is legitimate evidence for either differential deficit in ability or differential susceptibility to one of two positive symptoms. Figure 3 shows the hypothesized result.

In figure 3, the percentage of errors of the normal standardization group is taken as normative data and is plotted on the abscissa. The percentages of errors of both the normal experimental group and the schizophrenics are plotted against it on the ordinate. If the standardization is successful, the slope for the normal experimental group should be 1.0 for both variables. The hypothesis is supported if the schizophrenics show a steeper slope in response to the experimental variable than the control variable.

Rosenbaum, Ebner, and Ritzler (1966) used a design very close to this in a study of proprioceptive deficit in schizophrenia. Their experimental task required the subject to lift small weights in sequence, and to report which of two weights was the heavier. They presented this proprioceptive discrimination task at two levels of difficulty, the more difficult task being with lighter weights. The matched control variable was discrimination of loudness of tones. The subject was presented with two tones and was asked to report which was louder. Rosenbaum, Ebner, and Ritzler used the data of Stevens, Mack, and Stevens (1960) to select the proprioceptive and auditory items so that the two variables produced an equivalent rate of increase in errors for normal subjects. They found that, despite this matching of the two variables for normal subjects, the experimental task variable (difficulty in proprioceptive discrimination) produced a greater increase in errors for the schizophrenics than did the control task variable (difficulty in auditory discrimination). They concluded that schizophrenics have a greater deficit in proprioceptive discrimination than in auditory discrimination. Rosenbaum, Ebner, and Ritzler did not report reliability values for any of the measures, but these were
Figure 3. The predictions in the design of manipulating two matched task variables.

Using Measures of Thought Disorder to Discover Types of Schizophrenia

Most psychopathologists believe that if more than one subdisorder is hidden within schizophrenia, these subdisorders have not yet been discovered. A reasonable approach to separating out such subdisorders might start with measures of thought disorder. Thought disorder is a central feature of schizophrenia; nevertheless, schizophrenics appear heterogeneous in their thought disorder. Any finding of schizophrenic thought disorder in an objective study is almost invariably attributable only to a portion of the schizophrenic group. Many schizophrenics score in the normal range even though they are cognitively disordered by other criteria.

Comment

The most appropriate design for measuring cognitive deficit is almost always that of matched tasks or its variant, the design of matched variables. It is unfortunate that these very useful designs have seldom been used. Most of the available research evidence on theories of schizophrenic thought disorder must be obtained from the more hazardous design of matched groups or from the exceptional cases in which other designs yield valid conclusions. Knowledge in this field will advance much more rapidly when investigators adopt the more appropriate designs.

very possibly equivalent because of the matching on difficulty and the other similarities of the two tasks.
**Factor Analysis**

The most widely used approach to seeking varieties of schizophrenia has been factor analysis, which is a method of isolating sources of covariation among test scores. In this method, matrices of correlations between test scores are analyzed to discover factors that account for the patterns of correlation. Tests that have a high correlation with one another tend to be accounted for by the same factor. People who score high on different factors are viewed as belonging to different types.

A difficulty in using factor analysis with extant measures of thought disorder is that the correlation of tests with one another tends to increase with increases in their reliability. A test with low reliability will have little loading on any factor. Published measures of thought disorder vary greatly on reliability because they have been constructed with little attention to their psychometric properties. Factor-analytic studies of schizophrenic cognition are also extremely difficult to conduct, because they require scores on several tests from a large number of patients.

**Seeking Correlates of Thought Disorder**

A simpler approach than factor analysis for seeking varieties of schizophrenia is to look for specific relationships that, if found, would seem likely to indicate more than one kind of schizophrenia. An outstanding example of this approach is the study of Willner and Struve (1970), who suggested, on the basis of prior research, that poor performance on analogies might indicate minimal brain damage. They hypothesized that such patients might have slightly abnormal electroencephalograms. These investigators worked with a group of newly admitted patients with varied nonorganic diagnoses including schizophrenia. They found abnormal electroencephalograms more often in patients with poor than in patients with adequate analogy performance. They interpreted these results as indicating that a subset of these patients may have developed their psychiatric problems in response to minimal brain injury. The investigators suggested that the level of analogy performance that should be considered pathological should vary with verbal IQ. This suggestion amounts to saying that analogy performance should be measured as differential deficit and that verbal IQ should be the control task. The methods of matching tests that we have described here should facilitate development of such a measure.

**Seeking a Bimodality on a Measure of Thought Disorder**

Another simple approach to seeking subgroups of patients is to use a measure of a single aspect of thought disorder, thus distinguishing patients who have that symptom from patients who do not. The methodological requirements of a measure of thought disorder for such a study are the same as for testing a hypothesis about the nature of the thought disorder. Matched tasks are required. The rationale is as follows: A single task is seldom adequate for separating patients who have a cognitive symptom from those who do not, because scores on measures of thought disorder are usually correlated with general performance level. Therefore, two tasks are needed, with the difference between the two scores being the measure of the thought disorder. For example, one might choose a proverb-interpretation task and a vocabulary task, with the expectation that some schizophrenics, but not others, will have a greater deficit on proverbs than on vocabulary. Each subject receives a difference score of accuracy on proverbs minus accuracy on vocabulary.

The two tasks must be matched so that the difference score will be independent of general performance level. As discussed above, this means matching on reliability, mean, variance, and shape of the distribution of item difficulty, and on the distribution of test scores. With tests having all of these characteristics, normal subjects as well as those schizophrenics who do not have the kind of thought disorder being measured will have difference scores that are normally distributed around zero. Schizophrenics who do have the thought disorder measured by the difference score should have a distribution of scores around a point other than zero. Thus, if some schizophrenics have that kind of thought disorder and other schizophrenics do not, one might reasonably hope to find a bimodal distribution of the difference scores. Such a bimodality, if it is clear-cut, might correspond to two kinds of schizophrenics. One might look for other correlates of which mode the score falls in, such as biochemical differences, differences in case history variables, or differences in genetic history of psychosis.

Unfortunately, there is no statistical test for bimodality. One must judge its presence or absence by
inspection, and such judgment requires more subjects than are used in most studies of the presence or absence of a symptom.

Also, a bimodality does not always emerge even when two distinct groups with different scores lie within the subjects tested. Murphy (1964) has pointed out that the likelihood of finding a bimodality in such a case depends on the difference between means of the two groups, the variances of the two groups, and the groups' relative sizes. Murphy shows that for equal-size groups of the same variance, the two groups' means must differ by three to four standard deviations for a clear bimodality to emerge. Such a finding could appear for schizophrenics only for a schizophrenic symptom which sharply distinguishes schizophrenic from normal subjects and which is measured both reliably and in a way that excludes the effects of generalized deficit. The development of such measures should be readily achievable by the methods described above.

One must take care not to be misled by artificial bimodalities that may be created by one's method of selecting a sample of schizophrenics. For example, an artificial bimodality might occur in a sample that includes both medicated and drug-free patients, or both newly admitted and long-term chronic patients. In general, if patients are selected so as to fall into a dichotomy or bimodal distribution on some other related variable, this selection may lead to an artifactual bimodality on thought disorder.

Conclusions

The most potentially fruitful approach to resolving the heterogeneity of schizophrenic cognition is probably that of discovering new subvarieties of schizophrenia with different kinds of thought disorder. Almost nothing has been done using this approach, because the measures of thought disorder have not been adequate to distinguish specific symptoms from generalized deficit. The use of adequately matched tasks will make this distinction possible.

References


Rosebaum, G.; Ebner, E.; and Ritzler, B. "Effects of Stimulus Intensity on Auditory and Weight Discrimination of Normals and Schizophrenics." Presented at the


---

**The Authors**

Loren J. Chapman, Ph.D., is Professor, and Jean P. Chapman, Ph.D., is Lecturer, Department of Psychology, University of Wisconsin, Madison, Wis. 53706.

---

**available from nimh**

*It's Good to Know about Mental Health,* a brochure recently released by the National Institute of Mental Health, attempts to define mental health and describes some of the causes and warning signs of mental disorders, such as prolonged anxiety, depression, tension-caused physical symptoms, and perfectionism. The booklet gives a number of simple suggestions for keeping or achieving good mental health. It also describes the kinds of mental illness and where and how the mentally ill can obtain help. Single courtesy copies are available from Public Inquiries, Room 9C-05, National Institute of Mental Health, 5600 Fishers Lane, Rockville, Md. 20852.

*Definition of Terms in Mental Health, Alcohol Abuse, Drug Abuse, and Mental Retardation,* a revised, updated edition of a manual defining basic terms in mental health and related areas of concern, has been published by NIMH. The 123-page manual provides a common language for the exchange of information among concerned disciplines. The lack of commonly understood definitions has made it difficult to compare programs in terms of work done, costs, facilities, staff, and persons served. The manual, if widely used, should make possible enhanced sharing of information on program planning and evaluation. It is designed to be useful to program planners and developers, program administrators, statisticians, budget officers, and others involved in funding and cost accounting. The manual, NIMH Mental Health Statistics Series C, No. 8, DHEW Publication No. (ADM) 74-38, is available from the Survey and Reports Section, Biometry Branch, Room 18C-25, National Institute of Mental Health, 5600 Fishers Lane, Rockville, Md. 20852.