One of the major social programs of the 1960s was the development of community mental health centers. As with most early attempts at evaluation, the results were pessimistic. This article reanalyzes one of the earliest, and best-known, evaluations of a community-based treatment facility. Following the conceptual framework of Campbell and his associates, it was found that the various threats to the validity of the findings indicate a consistent and systematic bias against detecting a positive effect for the new mental health center. In light of recent federal legislation mandating formal evaluations, appropriate procedures are discussed.

TWO REGIONAL MENTAL HEALTH TREATMENT FACILITIES

A Reanalysis of Evaluation of Services

A. JOHN McSWEENY
West Virginia University
PAUL M. WORTMAN
University of Michigan

In the 1960s a major change in focus took place within the mental health field. This change was caused by several factors, including dissatisfaction with the expense and the effectiveness of long-term custodial treatment, the recognition of the rights of the poor to adequate health

---

AUTHORS' NOTE: *Preparation of this article was supported by a postdoctoral fellowship from the National Institute of Mental Health, Grant No. MH 00180, to the first author and Contract No. C-74-00115 from the National Institute of Education to the second author. The authors are greatly indebted to William Smith of the Rockford School of Medicine and Jeffrey Kaplan and Imat Amidjaya of the H. Douglas Singer Mental Health Center for their generosity in supplying and helping interpret the data. An earlier version of this paper was presented at the 85th Annual*

treatment, and the development of effective psychotropic medications. This resulted in a major decline in the population of mental hospitals and an increased emphasis on treating persons in their natural environments. The term usually applied to this "bold new approach" (as it was described by President Kennedy) is the "community mental health movement."

In one Midwestern state, the community mental health movement initially took the form of two new types of treatment facilities. The first was the community mental health center: a community-based treatment facility that usually provided outcare counseling, day-care programs, emergency hospitalization, and consultation/educational services. Local community groups were encouraged to build and manage these facilities through revised state tax laws and through state and federal matching funds. The second type of facility was a state-financed and managed regional treatment center. The regional treatment center performed three major functions: (a) consultation and other forms of assistance to local community groups in the process of constructing community treatment centers; (b) transitional services to communities that did not have such a center; and (c) direct incare and outcare services to psychiatric patients in the region it served. It is this last service that the remainder of this article will be concerned with.

The direct clinical services that the regional treatment facilities offered were analogous to those provided by a traditional state hospital. However, the regional treatment center differed from the traditional state hospital in several ways. First, the regional treatment center was designed to be physically attractive and allow patients a reasonable degree of comfort, freedom, and privacy. Semiprivate sleeping quarters, low ceilings, the absence of locked doors, and mixed-sex treatment units minimized the institutional aspects of the treatment environment. Second, the incare treatment program was designed to provide brief, but intensive, treatment. As Smith (1975: 49) described it, the program instilled an optimistic attitude, "that the patient is suffering from a temporary setback, that he can improve, and that he is expected to perform in a socially expected way." Finally, a major emphasis was put on returning the patients to their natural environments as quickly as possible and continuing treatment after they returned: "An active effort (is made) to link patients to a network of

---

Convention of the American Psychological Association in August 1977. Requests for reprints should be sent to A. John McSweeny, Department of Psychology, West Virginia University, Morgantown, WV 26506.

supporting community resources . . . much time is spent helping the patient reestablish firm ties with his family, a stable residence, a job or job training, and other community groups" (Smith, 1975: 49).

THE NEED FOR EVALUATION

Community-oriented mental health programs proliferated during the late 1960s and early 1970s with the held of federal funds. However, few of the programs instituted formal evaluation procedures until Congress required them to do so in 1975 (P.L. 93-641). The study to be discussed (Smith, 1975) represents one remarkable exception to the aforementioned state of affairs. Soon after the inception of the regional center's first programs, an evaluation unit was established which began to collect systematic data about the costs of treating patients, the number of days patients remained hospitalized, and their level of functioning in the natural environment they returned to. Similar data were collected by the center's evaluation unit on a 10% sample of patients admitted to three neighboring traditional state hospital programs.[1] These data served as the basis of an evaluation report (Smith, 1975).

The results of this report were disappointing. Although the regional center was able to discharge patients more quickly, few differences were found between the regional center and the traditional state hospitals in terms of their ability to maintain patients effectively in the community. Moreover, it cost more to treat patients at the regional center than at the traditional hospital. As Smith (1975: 54) stated:

> Although the increased staff, increased expenditure and community orientation of the regional mental health center represent a clear advance in the humanitarian care of the mentally ill, no evidence was found that this new approach substantially altered the outcome of serious mental disorders or of the disability associated with them.

The purpose of this article is to re-examine critically the data from the original evaluation through a process called secondary analysis (Boruch and Wortman, 1978; Cook, 1974).

SECONDARY ANALYSIS

Secondary analysis involves a careful reappraisal of innovative social programs from a rigorous methodological standpoint (Campbell, 1969; Wortman, 1975). Its purpose is to determine the validity of the findings in accordance with accepted scientific methods (Campbell and Stanely, 1966; Cook and Campbell, 1976). Although there have been a variety of secondary analyses reported in the literature, they all usually include some issues dealing with the validity of the social program. These issues may involve the theoretical or conceptual base of the study (i.e., construct validity), the adequacy of the design (i.e., internal validity), and the subsequent analyses (i.e., statistical conclusion validity), or the replicability or generalizability of the findings (i.e., external validity). The reexamination of the program typically provides for a clearer interpretation of the outcomes through a reanalysis that is responsive to the methodological criticisms. Although all of these elements are not present in every secondary analysis, one or more are essential. In particular, the systematic reconsideration of the program's outcomes and their relevance for policy are key ingredients of secondary analysis and ones that distinguish it from social commentary and other forms of policy analysis.

There are a number of ways this data reanalysis can proceed. Initially, the original results should be replicated and confirmed. This is a useful procedure, for it provides a "feel for the data" and often illuminates problems with it as well as revealing overlooked portions of it (e.g., omitted variables) that can be usefully exploited. For example, the verification of simple population characteristics such as the number of patients served, percentage of males and females, percentage in each diagnostic category, whether outcome scores were in an acceptable range and the like provide a useful indication of data quality. Often major inconsistencies between similar variables will be found that indicate poor data quality and mitigate further analysis (Boruch, Davis, and Magidson, 1975). In some cases, where it is obvious that the original analyses are inappropriate or incorrect, this first step can be skipped. At this point new analyses are performed. These, more appropriate, statistical methods can be applied using either the original design or others considered appropriate for the secondary analysis. It is this particular step, above all others, that characterizes secondary analysis. In reexamining the services of the two mental health treatment programs, a reanalysis of a portion of the data relating the background characteristics of the patients and to outcome will be summarized.

VALIDITY ISSUES

There are a number of important concepts that have been developed to assess social experiments (Campbell and Stanley, 1966; Cook and Campbell, 1976). Although the evaluation of the regional treatment center's programs lacked the rigor found in a randomized experiment, the conceptual basis for secondary analysis is still appropriate for a number of reasons. First, these techniques are applicable to quasi-experiments, and, as will be demonstrated, the evaluation of the regional treatment center was a quasi-experiment. Second, the original investigation involved a number of analyses, some of which were published (Smith, 1975) and had an impact on mental health policy.[2] However, before the design and results can be considered, it is necessary to determine if the program at the regional treatment center was properly implemented.

CONSTRUCT VALIDITY

The evaluative process dealing with the translation or operationalism of theoretical variables has been called construct validity by Cook and Campbell (1976), among others. The focus of this part of the secondary analytic process is thus on the relation of theory to reality. This entails a determination of whether or not the various aspects of the theory have been implemented and whether appropriate measures of their impact have been employed.

It can be argued that there is no rigorous theory behind the operations of the community-oriented treatment center or indeed behind the community mental health movement in general. However, from the information provided by Smith (1975), Smith, Kaplan, and Siker (1974), and unpublished documents and discussions with personnel of the regional treatment center, several operating assumptions became apparent. These are: (a) patients will benefit most from treatment in their natural environments, (b) therefore, their stays on the incare unit should be kept as short as possible, with (c) their incare treatment aimed toward helping them gain the requisite social skills to return to their natural environments, and (d) the best way to accomplish objectives (b) and (c) is to make the incare program intensive. That is, the patients should be involved in treatment activities for most of the day. The intensive, short-term program also

entails a high level of patient-staff interaction and requires a higher expenditure for a greater number of staff members than would a traditional mental hospital program.

In most respects, it appears as though the regional treatment center's programs were implemented as planned. Initial stays in the hospital were kept relatively short. In the first six-month follow-up period, patients stayed for a median of 25.0 days in the incare unit of the regional treatment center and a median of 54.5 days in the state hospitals, after they were admitted. Using a Mann-Whitney U test, this difference was found to be significant, $p < .01$.[3] Thus, the regional treatment center appeared to be sending patients home more quickly after their first admissions.

Although data concerning staff-patient ratios or the cost of treating patients were not available, it is apparent from Smith's (1975) report that a greater amount of staff effort was expended in treating persons. Treatment costs per patient over the total 3 years of the study was $12,000 for each center patient and $9,000 for each state hospital patient. Much of this cost difference was attributed to greater personnel costs at the center and the greater frequency of outpatient contacts made by center staff. Moreover, the first-hand experience of one of the present authors, gained as an employee of the center during the period of the study, tends to confirm the relatively high staffing levels, intense staff-patient interaction levels, and frequent outpatient contacts suggested by Smith's published data.

In some respects the implementation of the community-oriented treatment approach was less than ideal. The staff of the center program were, as Smith (1975: 48) points out, "young, enthusiastic, and intelligent, but relatively untrained and inexperienced." One could argue that using such personnel to provide an innovative treatment program is an inadequate implementation of the original concept. On the other hand, there is no reason to believe that the center administrator had much choice, given the shortage of experienced mental health personnel in the 1960s. In fact, the use of "paraprofessionals" is generally consistent with a community mental health approach.

A second problem with the implementation of the regional center program concerns the extent to which patients were actually contacted by treatment personnel while they lived in the community. A quick glance at Table 1 reveals that patients in both groups received very few visits from either regional center/state hospital personnel or from indigenous community professionals during the six-month reporting periods. In fact, the

TABLE 1

Outpatient Contacts (median visits per month by source with Wilcoxon rank-sum tests for differences between regional center and state hospital)

	Regional Center			*State Hospital*				
Follow-up Period	*N*	*Median Visits per Month*	*Mean Rank*	*N*	*Median Visits per Month*	*Mean Rank*	*Z*	*p*[a]
Outpatient Visits Made by Regional Center or State Hospital Personnel								
1	36	.200	42.2	28	.001	20.0	4.96	<.001
2	37	.005	40.0	36	<.001	33.9	1.72	.043
3	36	.010	38.3	35	.003	33.7	1.62	.052
4	31	.001	35.4	36	.005	32.8	0.87	N.S.
5	31	.020	35.3	34	.005	30.9	1.63	.051
6	27	.011	32.4	32	.000	28.0	2.23	.012
Outpatient Visits Made by Community Professionals								
1	36	.186	36.3	28	.002	27.6	2.02	.022
2	37	.055	35.3	36	.008	38.8	−.77	N.S.
3	36	.648	40.8	35	.325	31.1	2.02	.022
4	31	.271	36.0	36	.002	32.3	0.81	N.S.
5	31	.024	33.5	34	.024	32.5	0.24	N.S.
6	27	.167	33.5	32	.001	27.1	1.56	.059

a. One-tailed test.

NOTE: An outpatient was defined as a person who spent less than three months hospitalized during any given six-month follow-up period.

modal number of visits per month by any professionals for all six reporting periods was zero. Clearly it would be difficult to argue that the typical patient in either group received intensive outcare treatment even if a statistically significant difference in favor of the regional center sample can be seen for many of the reporting periods.[4]

The relatively low number of outpatient visits contrasts with the subjective impression held by the first author and other regionl center employees that a great deal of outpatient activity was maintained. Analysis of the distributions for number of visits suggested a possible reason for the contrast. A relatively small group of idividuals at the regional center did, in fact, receive fairly frequent visits from professionals. One patient, for example, received a total of 78 visits from regional center and community

professionals during one of the reporting periods! No doubt the regional center employees tended to remember the exceptional patients who received many visits rather than the patients who received very few.
few.

A few other instances of "slippage" in the implementation of the center program can be found. As Smith (1975: 48) states, the major goal was "to maintain patients in the community at a maximal level of social functioning." In a few cases, however, placing the patients "in the community" consisted of helping them find a room in a hotel for transients. This is a problem that has grown in recent years with cutbacks in public funds (Koenig, 1978) and definitely does not fit the ideal conceptions of how community-based treatment should be carried out!

The question of whether appropriate outcome measures were employed should also be considered when evaluating the construct validity of an investigation. Given that the major goal of the regional center program was to keep patients in the community and to maximize their level of social functioning, two types of outcome measures are indicated: a measure of community tenure and a measure of social functioning. Smith's measure of community tenure was derived from the number of days patients were hospitalized during the preceding six months. Although technically this is a measure of hospitalization rather than community tenure, one can safely assume an inverse relationship between hospitalization and community tenure. Therefore the measure is conceptually adequate.

Smith's (p. 51) measure of social functioning consisted of a categorical variable which he called a survival indicator. A patient was considered a success if he or she (a) was not hospitalized for more than 30 days in the preceding 6 months for a mental disorder, (b) did not come in contact with law enforcement agents for criminal activity or unacceptable behavior, (c) derived at least one-half of his economic support from his own resources, and (d) "belonged to a . . . family or family surrogate group." A patient who did not meet *any one* of these four criteria was considered to be a "failure."

Smith's measure of social functioning does cover four key aspects of social adjustment, but being a categorical measure it is probably not as sensitive as such inventories of social functioning as the Katz Adjustment Scale (Katz and Lyerly, 1963), which permit the presentation of social functioning as a continuous variable. Because social functioning is such a critical index of the effectiveness of this and other mental health programs, it is unfortunate that the Katz scale or one like it was not employed in the present study.

We attempted to construct a "social functioning index" that would have properties of a continuous variable by combining the four sections of the survival indicator. A repeated measures analysis of variance revealed no significant differences between the treatment groups. Similar results were found with other composite indices we attempted to create from various combinations of available variables.

INTERNAL VALIDITY

The evaluation and comparison of the programs at the regional center and the state hospitals can be conceptualized as a nonrandomized or "quasi-experiment" (Campbell and Stanley, 1966). The basic design of the study is the "static-group comparison" (Campbell and Stanley, 1966: 12). This design, as applied to the present study, is depicted in Figure 1. It involves nonrandom assignment to the treatment and control conditions and at least one posttreatment assessment. In the case of the present study, six posttreatment assessments took place over a three-year period (at six-month intervals). The design used in the present study differs from the standard static group comparison in that the "treatment" was not completely removed before the follow-up assessments; all patients who were discharged continued to receive some form of outpatient services and a few were eventually readmitted to the incare services at the center.

The absence of random assignment of patients to various experimental and control groups in quasi-experiments such as the present study allows a variety of factors other than the treatment itself to cause the observed effects. These "plausible rival hypotheses" provide alternative explanations of any treatment effect and therefore limit the interpretability or "internal validity" of the quasi-experiment. Although Campbell and Stanley (1966) and Cook and Campbell (1976) have identified over 30 threats to the validity of quasi-experiments, 4 are of particular relevance in the present situation.

One of the most serious threats to the internal validity of the static-group comparison is selection. If patients in the two programs differ significantly in some way that could affect their level of social functioning subsequent to treatment, it is possible that these selection differences alone could account for any differences which were observed. Conversely, selection differences could mask any treatment effects if the experimental patients' level of social functioning was lower than that of the control

Program		Follow-Up Assessments (Months)					
		6	12	18	24	30	36
Regional Center	X	0	0	0	0	0	0
State Hospital		0	0	0	0	0	0

0 = Observation (follow-up assessments)
X = Treatment (community-oriented treatment program)

Figure 1: A Quasi-Experimental Design for the Evaluation of the Services of a Community-Oriented Regional Mental Health Center

patients. Thus, the absence of observed treatment effects may also be attributed to prior (i.e., selection) differences in the nonrandomized populations. Finally, selection factors may *interact* with the treatment in some way to magnify or reduce observed treatment effects. For example, it is possible that center patients differed from those in the state hospital group in ways that made them less responsive to the treatment.

Because selection factors present significant threats to the internal validity of the static group comparison, the evaluator must attempt to rule out this plausible "rival hypothesis." Smith (1975) performed chi-squares and t-tests on seven variables which he thought might present possible selection biases: age, sex, marital status, socioeconomic status (as measured by Holingshead's 1957 index), mean days hospitalized prior to study, and diagnosis. All of the original analyses indicated no significant differences between the regional center and state hospital groups. Although the original data for six of the seven variables were not available, we performed analyses on the data published in Table 1 of the Smith article (p. 50). The reanalyses generally confirmed the original analyses. Although no statistically significant differences between the two groups

were found on any of the seven background variables, a few potential problems were uncovered.

The reanalyses of the background data revealed two findings that relate to internal validity. First, the chi-square for diagnosis approached significance, $X^2(3) = 6.96$, $p = .07$. The primary differences between the two samples appear to consist of a greater proportion of patients with a diagnosis of organic syndrome in the state hospital sample (23% to 11%) and a greater proportion of psychotic patients in the regional center sample (40% to 20%). Because there is some evidence relating such diagnostic categories to treatment outcome (Fairweather and Tornatzky, 1977), this result suggests that it may be useful to consider psychiatric diagnosis in subsequent analyses of outcome.

A second finding consisted of several small discrepancies between the original analysis and the reanalysis in terms of total sample size and the mean number of days the patients were hospitalized prior to the study. The first discrepancy (sample size) was found to be due to loss of patients from the study at the first follow-up assessment. We were unable to determine the locus of the discrepancies between our analysis and Smith's analysis in terms of the mean number of days the patients were hospitalized before the study despite several efforts. Attempts to locate the original data sheets in order to help resolve the second discrepancy (in days hospitalized prior to treatment) also were not successful.

So far two possible threats to the internal validity of the present study have been discussed: selection and selection-treatment interactions. A third threat to the static group comparison design is differential mortality, that is, a differential rate of drop outs from the two samples which could introduce selection biases at follow-up. To test for this possibility, we examined the dropout rate in the two samples at each follow-up point. The dropout rate for each sample is depicted in Table 2. Using tests of significance for two proportions (Bruning and Kintz, 1968), no significant differences in dropout rate were found between the two samples at any of the six follow-up periods (All $|Zs| \leq .75$, $p > .23$). Thus, there was no evidence of significant differential attrition.

The final threat to internal validity is "diffusion" or "initiation" of treatment (Cook and Campbell, 1976). This threat often occurs when the treatment is highly publicized or part of a larger trend. In the present case, the community mental health movement caught up with the state hospitals in the second half of the study. The state hospitals began implementing community-oriented programs of their own which included improved outcare services. Some of the state hospital patients may have

TABLE 2
Number of Patients Remaining and Percentage Loss in Each Sample During Follow-Up Assessments

		Follow-Up Assessments (months)					
	Original Sample	*6*	*12*	*18*	*24*	*30*	*36*
Regional Center	47	45	44	42	38	39	37
	(0.0)	(4.2)	(6.4)	(10.6)	(19.1)	(17.0)	(21.3)
State Hospital	51	50	47	43	42	41	41
	(0.0)	(2.0)	(7.8)	(15.7)	(17.0)	(19.6)	(19.6)
Total Sample	98	95	91	85	80	80	78
	(0.0)	(3.1)	(7.1)	(13.3)	(18.4)	(18.4)	(20.4)

NOTE: Percentage loss is indicated in parentheses.

benefited from these improved outcare services although the new programs came after most of them had been inpatients. This seepage of the innovative treatment may have tended to equalize the two groups toward the end of the study.

In sum, although mortality did not appear to be a threat to the internal validity of the present study, the selection of patients differing in their distribution of initial diagnoses is of concern. In retrospect, a stratified (by diagnosis) sampling procedure would have been more appropriate. This was done by Smith and his associates (Smith, Kaplan, and Siker, 1974) in a subsequent study that found statistically significant positive effects for the regional center when compared to a state hospital. Furthermore, some diffusion of treatment occurred toward the end of the study that may have attenuated any differences between conditions. Although these potential problems should be kept in mind, overall the center and state hospital patients do appear broadly comparable.

Of course, it is still possible that the patients differed in some other ways that could have affected the results of the study. As Campbell and his colleagues (Campbell and Boruch, 1975; Campbell and Stanley, 1966) demonstrate, it is impossible to rule out group differences in nonrandomized experiments. In addition, as Smith (1975) noted, the center and state hospital groups differed in terms of their respective natural environments. The center patients came from a largely urban area and the state hospital patients came from a largely rural area. These two environments could have differed in a manner that affected the patient's level of social

functioning before and/or after treatment, or they could differ in terms of their tolerance for deviant behavior. In general, however, selection biases among demographic characteristics, although possible, are not conspicuous in the present study.

STATISTICAL CONCLUSION VALIDITY

Statistical conclusion validity concerns the amount of error present in the data and the choice of a proper statistical test. Both factors can influence conclusions about whether any observed treatment effects (or lack thereof) are really due to chance. In the present study two aspects of statistical conclusion validity were examined. First, we checked the congruence between variables that should, by definition, be congruent. Then the appropriateness of a major outcome variable was examined.

Noncongruence between similar variables would indicate the presence of systematic errors. For example, the data had four variables listed that consisted of days hospitalized: (a) total days hospitalized in a six-month period, (b) days hospitalized in the regional treatment center, (c) days hospitalized in a state hospital, and (d) days hospitalized in a private facility. According to their definition, variables (b), (c), and (d) should sum to equal variable (a). Unfortunately, that was not always the case. There were discrepancies in the median total days hospitalized for the two samples during three follow-up periods for the center sample and in one follow-up period for the state hospital sample.[5] The differences were generally not large, however.

On closer inspection of individual cases, many of the discrepancies were found to be small and probably due to rounding errors. A few discrepancies were large and appear to be recording or keypunching errors. Although a Wilcoxon test for correlated samples indicates that the discrepancies would not result in a significant change in the sample medians, the presence of errors casts doubt on the quality of the data and their potential for further analysis.[6] Moreover, several other instances of discrepancies between variables that should be congruent by definition for a few individual cases were also detected. It should be emphasized that none of these discrepancies probably would significantly change the findings of the original study. Still, the problematic data should be eliminated or "cleaned" before further inferential analyses are attempted.

The second aspect of statistical conclusion validity that was examined concerned the appropriate parametric representation of one of the major outcome variables: total days hospitalized. In the original article the mean was used as the estimate of central tendency and the t-test was employed as the major test or significance. An inspection of the distributions of this outcome variable indicated that they were quite skewed and significantly different from a normal distribution (average skewness of the distribution is 1.55, $p < .05$). Skewed distributions generally require the median as the measure of central tendency (Runyon and Haber, 1976). Indeed, a comparison of these two measures reveals that different trends arise when the median is used instead of the mean.

The results depicted in Figure 2 indicate that after the first follow-up period the typical regional center patient was hospitalized for about 20 to 40 days during each six-month period when the mean is used as the measure of central tendency. In contrast, when the median is applied, the typical patient is seen to have spent one day or less in an inpatient program during each six-month period. The reason for this apparently large discrepenacy is simple: after the first follow-up, approximately 2/3 to 3/4 of the patients were not hospitalized at all and a very small number of patients were hospitalized almost continuously!

After the first follow-up, days hospitalized essentially becomes a categorical variable with the great majority of subjects in one category. This conclusion casts further doubt on the sensitivity and utility of hospitalization days as a response variable. Erickson (1975) has noted these and other problems in the operational definition of "length of stay" that often reduce its value as an outcome measure.

EXTERNAL VALIDITY

In any study that has applied value, it is important to consider the generalizability of the results to other sites and subjects. Within the experimental context, generalizability issues are usually subsumed under the heading of external validity. In the present project, external validity presents two immediate questions: (a) Are the patients in this study like patients elsewhere, and (b) are there other studies of community-oriented treatment programs which present similar results? Smith (1975) thoroughly considered both questions in his original article and little further elaboration is necessary. He reported nonsignificant chi-squares

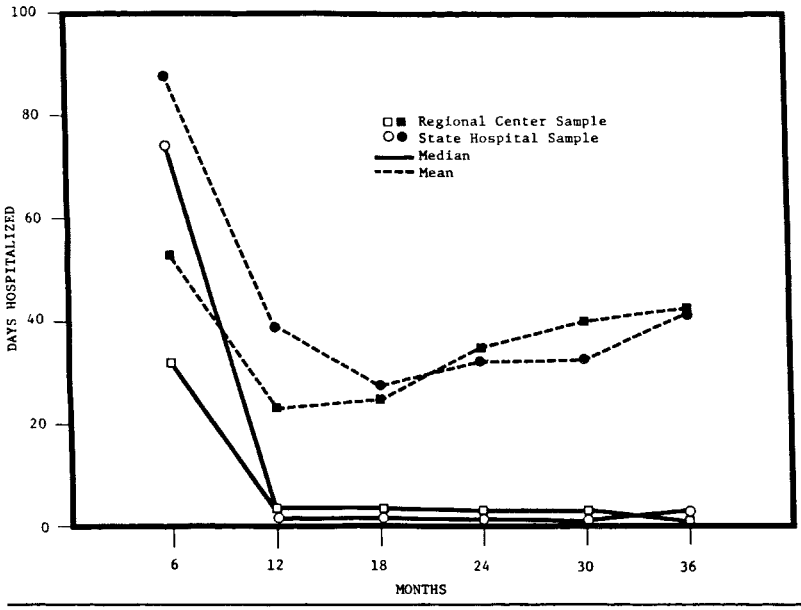

Figure 2: Mean and Median Days Hospitalized at Each Follow-Up Period

between the study sample and all admissions (excluding the study sample) to the state department of mental health in 1967 on five background variables: age, sex, marital status, socioeconomic status and diagnosis. Our reanalyses confirmed these results.

Smith (1975) also provided a brief but reasonably comprehensive review of other community treatment studies, including a very similar study by Kessel (1966). The results are generally consistent with those from the present study. In short, the external validity of the present study appears to be quite good.

One limiting factor in the present study is that it only involved one experimental site—the regional center. The external validity of the study would have been considerably improved had multiple sites been used. This would have required a statewide effort, or at least statewide cooperation, a relatively rare occurrence. Given this problem, one recent project of note is a study of psychiatric programs which involves 19 different V.A. hospitals in different areas of the country (Ellsworth, 1975a). This study, which is currently in progress, also is using a well-developed measure of social functioning: the Personal Adjustment and Roles Scale (Ellsworth,

1975b). Thus, the study should provide many valid and useful findings when it is completed.

SUMMARY AND CONCLUSIONS

A secondary analysis of an evaluation of a community-oriented treatment facility revealed several problems with the construct, internal, and statistical conclusion validities of the evaluation. The primary threat to the construct validity was the fact that few patients in the regional center program received the outpatient care a community-oriented program would dictate. The lack of proper program implementation would reduce its impact and yield nonsignificant estimates of effect. The internal validity of the study was subject to threats from selection and diffusion. Both of these would also have tended to bias the results toward the nonsignificant findings actually obtained. Moreover, the statistical conclusion validity of further analyses was cast in doubt because of numerous errors in the recording of data. The effects of these errors, although apparently small, preclude precise estimates of any treatment effects, and indicate some small bias in favor of the regional center for the first follow-up period. In addition, the utility of one of the primary outcome variables in the study, days hospitalized, appears doubtful because of the highly skewed distributions it presents. Unfortunately, the attempts to solve these problems proved unsuccessful. The external validity of the evaluation, on the other hand, was adequate.

The results of the secondary analysis are summarized in Table 3. The various threats to the validity of the study and their likely effect on the results are indicated. With the exception of the short-term bias that may have resulted from the choice of variables and statistical procedures, the general pattern indicates some systematic bias against finding an effect for the regional center. As Campbell and his associates have frequently noted (Campbell and Erlebacher, 1970; Campbell and Stanley, 1966), quasi-experimental evaluations of innovative social programs often find no, or even harmful, effects due to the bias resulting from the threats to validity that accompany these designs.

Two final comments are in order. First, the problems with the present data set underline the need for researchers to check their data for errors due to sources such as recording and keypunching (Bryant and Wortman,

TABLE 3
Summary of the Threats to Validity and Their Likely Effect on the Results

Threats to Validity	*Present (P) or Absent (A)*	*Likely Bias in Favor of:*		
		No Difference	*Regional Center*	*State Hospital*
Construct				
a. Program	P	X		
b. Staff	P	X		
Internal				
a. Selection	P	?		
b. Selection X Treatment	?	?		
c. Mortality	A			
d. Diffusion	P	X		
Statistical Conclusion[a]				
a. Congruence of Variables	P		X	
b. Outcome Measure	P	X		
External Validity				
a. Patients	A			
b. Sites	A			

a. These would only affect the first six-month follow-up.