

Drawing Valid Inferences From Case Studies

Alan E. Kazdin

Western Psychiatric Institute and Clinic
University of Pittsburgh School of Medicine

Although the case study is usually accorded an important heuristic role, its limitations as a basis for drawing valid conclusions about the efficacy of treatment are widely recognized. The major problem with a case study is the ambiguity related to the precise influences that are responsible for change. Threats to internal validity, normally ruled out in experimentation, make ambiguous the basis for therapeutic change. However, case studies may vary along several dimensions that dictate the extent to which specific threats to internal validity are addressed. The present article discusses the possibility of conceptualizing and conducting cases in such a way as to maximize the likelihood that internally valid conclusions can be drawn. Although case studies can never replace experimentation, the scientific yield from case reports might be improved in clinical practice where methodological alternatives are unavailable.

The case study has played a central role in clinical psychology. Indeed, understanding the individual person is occasionally considered to be a distinguishing characteristic of clinical psychology relative to other branches of the field (Korchin, 1976; Watson, 1951). The intensive study of the individual has contributed to clinical research and practice by providing a rich source of hypotheses about the bases of personality and behavior and by serving as a place to develop and apply intervention techniques (Bolgar, 1965; Garfield, 1974; Kazdin, 1980; Lazarus & Davison, 1971).

Despite its recognized heuristic value, the case study is usually considered to be inadequate as a basis for drawing valid scientific inferences. Relationships between independent and dependent variables are difficult to discern in a typical case study because of the ambiguity of the factor(s) responsible for performance. For example, treatment for a particular clinical case may be associated with therapeutic change. However, the basis for the change cannot be determined from

an uncontrolled case study. Even if treatment were responsible for change, several alternative interpretations of the case might be proposed. These alternative interpretations have been catalogued under the rubric of "threats to internal validity" (Campbell & Stanley, 1963).¹

The case study has been discounted as a potential source of scientifically validated inferences, because threats to internal validity cannot be ruled out in the manner achieved in experimentation. Even though the case study is not experimental research, under several circumstances it can lead to knowledge about treatment effects for a given client that approximates the information achieved in experimentation. The present article examines the case study and

¹ The threats to internal validity refer to classes of variables that might produce effects mistaken for the effects of treatment. The major threats include the influence of (a) history (specific events occurring in time), (b) maturation (processes within the person), (c) testing (repeated exposure to the assessment procedures), (d) instrumentation (changes in the scoring procedures or criteria over time), (e) statistical regression (reversion of scores toward the mean or toward less extreme scores), (f) selection (differential composition of subjects among the groups), (g) mortality (differential attrition among groups), and (h) selection-maturation (and other) interactions (where differential changes occur as a function of other threats with selection). For additional threats, see Cook and Campbell (1979) and Kazdin (1980).

Preparation of this manuscript was facilitated by Grant MH31047 from the National Institute of Mental Health.

Requests for reprints should be sent to Alan E. Kazdin, Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15261.

its variations as a research tool. Alternative ways in which case studies are conducted and reported have important implications for drawing scientifically validated information. The present article discusses what can be done with the clinical case to improve the scientific inferences that can be drawn.

The case study as a potential source of scientifically valid information warrants careful scrutiny for several reasons. First, the case study has had tremendous impact on psychotherapy. Individual cases (e.g., Little Hans, Anna O., Little Albert) and series of cases (e.g., Masters & Johnson, 1970; Wolpe, 1958) have exerted remarkable influence on subsequent research and practice. Second, the case study draws attention to the frequently lamented hiatus between clinical practice and research. Clinicians have access to the individual case as their most convenient and feasible investigative tool, but inadequacies of the case study as a research strategy limit the inferences that can be drawn. Researchers often rigorously investigate psychotherapy but may sacrifice clinical relevance in the populations, therapeutic conditions, and standardization of treatment that research may require. Hence, investigation of psychotherapy often obscures aspects of the phenomenon of interest (Strupp & Hadley, 1979), and even under the best conditions, treatment may only be an "analogue" of the clinical situation (Kazdin, 1978).

One suggestion to help bring research and practice of psychotherapy closer together is to encourage clinicians to utilize single case experimental designs (Hersen & Barlow, 1976). The designs permit experimental investigation of the single case. The designs have been applied successfully and often dramatically in case reports where the effects of treatment have been carefully documented with complex clinical problems seen in individual treatment. However, single-case experimental designs often impose special requirements (e.g., withdrawing or withholding treatment at different points in the designs) that are not always feasible in the clinical situation. Hence, some authors have suggested that the designs have not really been applied as widely as they should (Barlow, 1980) and perhaps often cannot be applied, because of the ethical, methodological,

and practical obstacles inherent in clinical settings (Kazdin & Wilson, 1978).

Apart from the merits of single-case experimentation, nonexperimental alternatives need to be examined carefully. Indeed, in other areas of psychology, experimentation is not always possible. In such instances, important alternatives have been provided by elaborating the requirements for quasi-experiments (Campbell & Stanley, 1963), which can achieve some if not all of the goals of experimentation. Similarly, in the context of clinical practice, experiments are to be encouraged when opportunities exist. However, it is very important to elaborate the conditions that can be invoked to achieve several goals of experiments when rigorous investigations are not possible. The uncontrolled case study is a widely available investigative tool, and its methodological limitations, advantages, and alternatives need to be elaborated.

Characteristics of the Case Study

The case study has been defined in many different ways. Traditionally, the case study has referred to intensive investigation of the individual client. Case reports often include detailed descriptions of individual clients. The descriptions rely heavily on anecdotal accounts of the therapist to draw inferences about factors that contributed to the client's plight and changes over the course of treatment. Aside from the focus on the individual, the case study has come to refer to a methodological approach in which a person or group is studied in such a fashion that inferences cannot be drawn about the factors that contribute to performance (Campbell & Stanley, 1963; Paul, 1969). Thus, even if several persons are studied, the approach may still be that of a case study. Often cases are treated on an individual basis, but the information that is reported is aggregated across cases, for example, as in reports about the efficacy of various treatments (e.g., Lazarus, 1963; Wolpe, 1958). Hence, there is some justification for not delimiting the case study merely to the report of an individual client.

In general, the case study has been defined heterogeneously to denote several different things, including the focus on the individual,

reliance on anecdotal information, and the absence of experimental controls. A central feature of the diverse definitions is that case studies differ from experimental demonstrations. Texts on methodology usually discount the case study as a preexperimental design and use it as a point of departure to show that experimentation is the alternative means for obtaining scientifically validated knowledge (Campbell & Stanley, 1963; Hersen & Barlow, 1976; Kazdin, 1980). However, the purpose of the present article is to suggest that case studies and experiments fall on a continuum that reflects the degree to which scientifically adequate inferences can be drawn. More importantly, several types of uncontrolled case studies can be identified that vary in the extent to which they permit valid conclusions.

The purpose of experimentation is to rule out threats to internal validity, which serve as alternative rival hypotheses of the results. For example, in clinical treatment research, single-case or between-groups experimental designs are required to rule out the impact of extraneous factors that might account for the findings. Case studies do not provide the arrangements that permit conclusions that are as clear as those available from experimentation. However, many of the threats to internal validity can be ruled out in case studies so that conclusions can be reached about the impact of treatment.

Dimensions for Evaluating Case Studies

Case studies have been loosely and heterogeneously defined to include a variety of uncontrolled demonstrations aimed at showing that treatment produces therapeutic change. However, case studies may vary in how they are conducted and reported. The distinctions that can be made among case studies have important implications for drawing unambiguous conclusions. Major dimensions that can distinguish case studies, insofar as they relate to internal validity, are presented below.

Type of Data

The main criterion that distinguishes case studies is the basis for claiming that a change has been achieved. At one extreme, anecdotal information may be relied on and include narrative accounts by the client and/

or therapist regarding how client functioning has improved. Anecdotal reports are subject to a variety of limitations and sources of bias that need not be elaborated here. Suffice it to say that the anecdotal reports usually are not sufficient to conclude that changes really occurred in client behavior.

Case studies can include objective information, which refers to the large category of measurement strategies in which systematic and quantitative data are obtained. The specific measures encompass the gamut of assessment modalities and techniques (e.g., self-reports, ratings by others, overt behavior). Depending on other dimensions discussed below, objective information is a basic condition of a case study that has important implications for drawing inferences about the effects of treatment. The type of data obtained in a case study, that is, anecdotal or objective information, perhaps is the most important precondition for drawing inferences from a case study. Without some systematic data collection, other dimensions that might be applied to evaluate the case become almost irrelevant. Scientific inferences are difficult if not impossible to draw from anecdotal information. Indeed, it is the anecdotal information that is the problem rather than the fact that an individual case is studied. Even a rigorously designed experiment would be completely uninterpretable if anecdotal reports rather than objective assessment procedures served as the dependent measures.²

² The absence of quantitative information may not necessarily rule out drawing causal inferences. Occasionally, intervention effects are so powerful that qualitative changes appear to be produced, and the certainty of change and the reason for this change are relatively unambiguous. These effects, occasionally referred to as "slam bang" effects (Gilbert, Light, & Mosteller, 1975), are evident throughout the history of medicine and psychology. For example, Shapiro (1963) described a case of a patient with terminal cancer who showed a dramatic remission of symptoms on separate occasions as a function of receiving inert substances (an inactive drug and water). In this case, the changes were so strong and immediate ("slam bang" effects) that the absence of careful measurement did not pose as serious a problem for drawing inferences that the administration of treatment led to change. Apart from occasional examples, in most instances, quantitative information is required to attest to the fact that changes have in fact occurred and that these changes can be assessed with procedures that are replicable.

Assessment Occasions

Other dimensions that can distinguish case studies are the number and timing of assessment occasions. The occasions in which this objective information is collected have extremely important implications for drawing inferences from the case. Major options consist of collecting information on a one- or two-shot basis (e.g., posttreatment only, pre- and posttreatment, respectively) or continuously over time (e.g., every day or a few times per week) for an extended period.

When information is collected on one or two occasions, say before or after treatment, difficulties arise in inferring that change has occurred as a result of treatment. Other interpretations of the change might be proposed (e.g., testing, instrumentation, statistical regression). With continuous assessment over time conducted before or after treatment, artifacts associated with the assessment procedures become less plausible. That is, changes as a function of the measurement instrument, if evident, normally would be detected prior to treatment and would not necessarily obscure the pattern of data relied on to infer changes associated with treatment.

Continuous assessment provides an additional advantage that can strengthen the internal validity of the case study. Data from continuous assessment prior to treatment can serve as a basis for making predictions about likely performance in the future. Extrapolations about the likely direction of performance provide implicit predictions about what performance would be like. The effects of treatment can be judged by the extent to which departures in the data are evident from the previously projected performance.

Past and Future Projections

The extent to which valid inferences can be drawn about treatment effects is influenced by past and future projections of performance. The past and future projections may derive from continuous assessment which shows that the problem is stable and has not changed for an extended period. As noted above, continuous assessment can provide information extrapolated to the future that may serve as an implicit but testable

prediction. If behavior appears stable for an extended period, changes that coincide with treatment suggest that the intervention may have led to change.

Past and future projections of performance also may be derived from understanding the course of a particular clinical problem. For some problems (e.g., obesity, social withdrawal), an extended history may be evident. The extended history is important from the standpoint of drawing inferences when change occurs. When change has occurred for a client whose problem has been evident for a long period, the plausibility that treatment caused the change is greatly increased. On the other hand, an acute clinical problem that has emerged relatively recently or is associated with a clear precipitating event may make evaluation of treatment slightly more difficult than it would be for a chronic problem. The acute or even episodic problem might be more amenable to the influence of extraneous (i.e., nontreatment) factors. Hence, it will be relatively difficult to rule out factors other than treatment that may account for the changes.

Projections of what the problems would be like in the future derived from understanding the particular clinical problem are also very relevant to drawing inferences about treatment. Research may suggest that a particular clinical problem is very likely to improve, worsen, or remain the same over a period of time. These alternative prognoses may be important when drawing inferences about treatment effects in a given case. For example, knowledge about the disorder may suggest that the problem will deteriorate over time (e.g., terminal cancer). The likely future for such a problem is highly relevant for evaluating whether treatment may be responsible for change. In the case of a patient with a terminal disease, improvements associated with a highly experimental treatment provide a strong inferential basis. Patient improvement strongly attests to the efficacy of treatment as the important intervention, because change in the disorder controverts the expected prediction.³ Of

³ This, of course, is not to say that the component the investigator believes to be important in the treatment was the one actually responsible for change, but only that the treatment and all that the treatment encompassed was the important event. The aspect(s) of treat-

course, with some clients and clinical problems, the future projections may indicate that improvements are likely even if no treatments are provided. For example, in treatment of children who have specific fears, conclusions about the short-term or long-term effects of treatment in a clinical case may be difficult to reach, because the projection for the future is for improvement (Agras, Chapin, & Oliveau, 1972). In general, inferences about the effects of treatment in a given case are more easily made to the extent that predictions can be made on the basis of extraneous information that the problem, if untreated, will follow a particular course. The plausibility that the changes are a result of treatment depend in part on the extent to which changes in client performance depart from the expected and predicted pattern of performance.

Type of Effect

The degree to which inferences can be drawn about the causal agent in the treatment of a clinical case also depends on the kinds of changes that occur. The immediacy and magnitude of changes contribute to judgments that treatment may have caused the change. Usually, the more immediate the therapeutic changes after the onset of treatment, the stronger a case can be made that treatment was responsible for the change. Of course, as any other of the dimensions discussed, showing that the conditions are met, in this case an immediate change, does not by itself mean that treatment was responsible for the change. But the more immediate the change, the less likely alternative sources of influence coincident with treatment account for the change. Alternatively, when change is gradual or delayed rather than immediate, the plausibility of associating the change with a particular event in the past (i.e., treatment) decreases. As the latency between treatment administration and behavior change increases, the number of ex-

traneous experiences that could account for the change increases as well.

Aside from the immediacy of change, the magnitude of change also contributes to the extent to which treatment can be accorded a causal role. More confidence might be placed in the causal role of treatment when relatively large changes are achieved. Of course, the magnitude and immediacy of change when combined increase the confidence that one can place in according treatment a causal role. Rapid and dramatic changes provide a strong basis for attributing the effects to treatment than more gradual and relatively small changes. A rapid and large change suggests that a particular intervention rather than randomly occurring extraneous influences accounts for the pattern of results.

Number and Heterogeneity of Subjects

Dimensions related to the subjects may influence the confidence that can be placed in conclusions about treatment effects. The number of cases included is important. Obviously, a stronger basis for inferring the effects of treatment stems from demonstrations with several cases rather than one case. The more cases that show changes associated with treatment, the more unlikely an extraneous event is responsible for change. An extraneous event that covaries with treatment and leads to therapeutic change is an unlikely rival hypothesis of the results, because the event must be common to all of the cases. The sheer number of cases obviously can contribute to the extent to which conclusions about treatment can be drawn by making implausible other explanations.

Aside from the number of cases, the heterogeneity of the cases may also contribute to drawing inferences about the cause of therapeutic change. If change is demonstrated across several clients who differ in various subject and demographic characteristics and the time that they are treated, the inferences that can be drawn are much stronger than if this diversity does not exist. Essentially, different persons have different histories and rates of maturation. As the diversity and heterogeneity of the clients and the conditions of treatment increase, it becomes increasingly implausible that the

ment that caused change in experimentation is not a question of internal validity but rather one of construct validity (Cook & Campbell, 1976, 1979). In experimental research, the particular aspect of treatment may still be debated (construct validity), even though the experiment has ruled out the impact of extraneous influences (internal validity).

Table 1
Examples of Types of Cases and the Threats to Internal Validity That They Address

Measure	Case example		
	1	2	3
Characteristic of case			
Objective data	+	+	+
Continuous assessment	-	+	+
Stability of problem	-	-	+
Immediate and marked effects	-	+	-
Multiple cases	-	-	+
Major threats to internal validity			
History	-	?	+
Maturation	-	?	+
Testing	-	+	+
Instrumentation	-	+	+
Statistical regression	-	+	+

Note. + indicates that the threat to internal validity is probably controlled, - indicates that the threat remains a problem, and ? indicates that the threat may remain uncontrolled. In preparation of the table, selected threats (mortality, selection, and others; see Footnote 1) were omitted because they arise primarily in the comparison of different groups in experiments and quasi-experiments. They are not usually a problem for a case study, which of course does not rely on group comparisons.

common experience shared by the clients (i.e., treatment) accounts for the changes.

Application of the Dimensions

The above dimensions do not necessarily exhaust all of the factors that contribute to drawing firm conclusions from case studies. Also, each of the dimensions is discussed separately. Yet, any particular case can be examined in terms of where it lies on each of the dimensions. Precisely where a particular case falls on all of the dimensions determines the extent to which particular threats to internal validity or rival alternative hypotheses can be ruled out in interpreting the results.

All of the possible combinations of the dimensions would yield a large number of types of case studies that cannot be presented here. Many of the dimensions represent continua where an indefinite number of gradations are possible so a large set of types of cases could be enumerated. However, it is important to look at selected types

of cases that vary on the dimensions mentioned earlier to show how internal validity can be addressed.

Table 1 illustrates a few types of cases that differ in their standing on the dimensions mentioned earlier. The extent to which each type of case rules out the specific threats to internal validity is also presented using a format that parallels similar analyses for true and quasi-experiments (Campbell & Stanley, 1963). For each case type, the collection of objective data was included. As noted earlier, the absence of objective or quantifiable data usually precludes drawing firm conclusions about whether change occurred. Drawing conclusions about the basis for change is premature, because the change itself has not been carefully documented. In the case types illustrated in Table 1, the assumption will be made that some form of assessment was completed, even if only one or two occasions to measure performance before and after treatment.

Case Example Type 1: With Pre- and Postassessment

A case study where a client is treated may utilize pre- and posttreatment assessment. The inferences that can be drawn from a case with such assessment are not necessarily increased by the assessment alone. Whether specific threats to internal validity are ruled out depends on characteristics of the case with respect to the other dimensions. Table 1 illustrates a case with pre- and postassessment but without other optimal features that would address and rule out threats to internal validity.

If changes occur in the case from pre- to posttreatment assessment, one cannot draw valid inferences about whether the treatment led to change. It is quite possible that events occurring in time (history), processes of change within the individual (maturation), repeated exposure to assessment (testing), changes in the scoring criteria (instrumentation), or reversion of the score to the mean (regression) rather than treatment led to change. Hence, even though the case included objective assessment, the conclusion that can be drawn about the basis for change

is not greatly improved over an anecdotal report.

Case Example Type 2: With Repeated Assessment and Marked Changes

If the case study includes assessment on several occasions before and after treatment, and the changes that occur at the time or over the course of treatment are relatively marked, then the inferences that can be drawn about treatment are vastly improved. Table 1 illustrates the characteristics of the case along with the extent to which specific threats to internal validity are addressed.

The fact that continuous assessment is included is important in ruling out the specific threats to internal validity that are related to assessment. First, the changes that coincide with treatment are not likely to result from exposure to repeated testing or changes in the instrument. When continuous assessment is utilized, changes due to testing or instrumentation could have been evident before treatment began. Similarly, regression to the mean from one data point to another, a special problem with assessment conducted only at two points in time, is eliminated. Repeated observation over time shows a pattern in the data. Extreme scores may be a problem for any particular assessment occasion in relation to the immediately prior occasion. However, these changes cannot account for the pattern of performance for an extended period.

Aside from continuous assessment, this case illustration was proposed to include relatively marked treatment effects, that is, changes that are relatively immediate and large. These types of changes produced in treatment help rule out the influence of history and maturation as plausible rival hypotheses. Maturation in particular may be relatively implausible, because maturational changes are not likely to be abrupt and large. However, a “?” was placed in the table, because maturation cannot be ruled out completely. In this case example, information on the stability of the problem in the past and future was not included. Hence, it is not known whether the clinical problem might ordinarily change on its own and whether

maturational influences are plausible. Some problems that are episodic in nature conceivably could show marked changes that have little to do with treatment. With immediate and large changes in behavior, history is also not likely to account for the results. However, a “?” was placed in the table here too. Without a knowledge of the stability of the problem over time, one cannot be too confident about the impact of extraneous events.

For this case overall, much more can be said about the impact of treatment than in the previous case. Continuous assessment and marked changes help to rule out specific rival hypotheses. In a given instance, history and maturation may be ruled out too, although these are likely to depend on other dimensions in the table that specifically were not included in this case.

Case Example Type 3: With Multiple Cases, Continuous Assessment, and Stability Information

Several cases rather than only one may be studied where each includes continuous assessment. The cases may be treated one at a time and accumulated into a final summary statement of treatment effects or treated as a single group at the same time. In this illustration as characterized, assessment information is available on repeated occasions before and after treatment as in the last type of case. Also, the stability of the problem is known in this example. Stability refers to the dimension of past–future projections and denotes that information is available from other research that the problem does not usually change over time. When the problem is known to be highly stable or follows a particular course without treatment, the clinician has an implicit prediction of the effects of no treatment. The results can be compared to this predicted level of performance.

As evident in Table 1, the threats to internal validity are addressed by a case report meeting the specified characteristics. History and maturation are not likely to interfere with drawing conclusions about the causal role of treatment, because several

different cases are included. All cases are not likely to have a single historical event or maturational process in common that could account for the results. Knowledge about the stability of the problem in the future also helps to rule out the influence of history and maturation. If the problem is known to be stable over time, this means that ordinary historical events and maturational processes do not provide a strong enough influence in their own right. Because of the use of multiple subjects and the knowledge about the stability of the problem, history and maturation are considered to be implausible explanations of therapeutic change.

The threats to internal validity related to testing are handled largely by the assessment over time. Repeated testing, changes in the instrument, and reversion of scores toward the mean may influence a comparison of performance from one occasion to another. Problems associated with testing are not likely to influence the pattern of data over a large number of occasions. Also, information about the stability of the problem helps to further make implausible changes due to testing. The fact that the problem is known to be stable means that it probably would not change merely as a function of assessment.

In general, the case study of the type illustrated in this example provides a strong basis for drawing valid inferences about the impact of treatment. The manner in which the multiple case report is designed does not constitute an experiment, as usually conceived, because each case represents an uncontrolled demonstration. However, characteristics of the type of case study can rule out specific threats to internal validity in a manner approaching that of true experiments.

General Comments

From the standpoint of experimentation, all of the above types of cases share a similar methodological status by being preexperimental and by not providing a sufficient basis for drawing scientifically valid inferences. The results that may emerge are usually rejected, because the data are from case studies. However, it is extremely important to

shift the focus from the type of demonstration (i.e., case study versus experiment) to the specific threats to internal validity that interfere with drawing valid inferences. The focus on rival alternative hypotheses that may be proposed draws attention to characteristics of the case reports that can be altered to improve the scientific yield.

The purpose of experimentation is to make as implausible as possible alternative explanations of the results. At the end of an experimental investigation, the effects of the treatment should be the most plausible and parsimonious interpretation of the results. Case studies can also rule out alternative explanations that might compete with drawing inferences about the impact of treatment.

Specific procedures that can be controlled by the clinical investigator can influence the strength of the case demonstration. First, the investigator can collect objective data in place of anecdotal report information. Clear measures are needed to attest to the fact that change has actually occurred. Second, client performance can be assessed on several occasions, perhaps before, during, and after treatment. The continuous assessment helps rule out important rival hypotheses related to testing, which a simple pre- and posttreatment assessment strategy does not accomplish.

Third, the clinical investigator can accumulate cases that are treated and assessed in a similar fashion. Large groups are not necessarily needed but only the systematic accumulation of a number of clients. As the number and heterogeneity of clients increase and receive treatment at different points in time, history and maturation become less plausible as alternative rival hypotheses. If treatment is given to several clients on different occasions, one has to propose an intricate explanation showing how different historical events or maturational processes intervened to alter performance. As in ordinary experimentation, in such cases, treatment effects become the more likely interpretation.

Some features of the case study that can help rule out threats to internal validity are out of the control of the clinical investigator. For example, knowledge about the stability

of the problem over time comes from information extraneous to a particular client. Knowledge about the course of the disorder is required. However, even though this is not controllable by the clinical investigator, he or she can bring available information to bear when interpreting results. This is already implicit in some instances where, for example, the problem in treatment is known to have a high remission rate (e.g., childhood fears). Remission of the problem, which may normally occur over time, requires special care for interpreting the long-term effects of treatment with any particular case or intervention.

The clinical investigator cannot easily control whether the changes in treatment are immediate rather than marginal. However, the data pattern that does result should be examined specifically in light of other rival hypotheses that might explain the results. Could any historical events (e.g., family processes, job experiences) or maturational processes (e.g., decreased depression as a function of the passage of time since divorce or death of a relative) be brought to bear that might explain the pattern of results? Perhaps the pattern of the data can help rule out specific rival hypotheses.

It is not merely how we conduct case studies that might warrant reconsideration but how we conceptualize them as well. Much can be done in carrying out case studies to increase the strength of the inferences about causal events. The well-known criticism of case studies as research tools has fostered a methodological learned helplessness about what can be done. In fact, much can be done to rule out specific threats to internal validity within case studies, such as the use of assessment on multiple occasions and the accumulation of several cases.

Some of the dimensions that help rule out threats to internal validity are out of the control of the clinical investigator. For example, one cannot by fiat achieve immediate and marked therapeutic changes nor be sure of the stability of the problem over the past and future. However, the clinical investigator can bring to bear the available research on the nature of the problem and evaluate the likelihood that historical events and maturational processes could achieve the sorts

of changes evident in treatment. It is not necessarily the lack of control over the clinical situation that is a problem. Within the limits of the situation, the clinical investigator might keep in mind some of the specific alternative rival hypotheses that need to be ruled out or made less plausible.

Conclusion

The case study occupies an extremely important place in clinical work both in inpatient and outpatient care. Case studies are widely recognized to serve as an important place to develop hypotheses about clinical problems and to explore innovative treatments. However, cases are usually considered to be completely inadequate as a basis for drawing scientifically validated inferences.

Case studies encompass several types of demonstrations that may differ in the extent to which inferences can be drawn. The issue is not whether a particular report is a case study. The focus on classifying reports on the basis of their lack of experimental design detracts from the more pertinent issue. Drawing inferences, whether in case studies, quasi-experiments, or experiments, is a matter of ruling out rival hypotheses that could account for the results. In case studies, by definition, the number of rival hypotheses and their plausibility are likely to present greater problems than they would in experiments. However, it is possible to include features in the case study that help decrease the plausibility of specific rival hypotheses.

The present article discusses several possibilities for assessing performance that rule out selected threats to internal validity. The purpose in adopting this approach is not to legitimize the case study as a replacement for experimental research. Experiments based on intra- and intersubject methodology can uniquely rule out threats to internal validity and can provide relatively clear information about the impact of treatment. Although the case study is not a substitute for experimentation, it has and probably will continue to contribute greatly to the information available in the field. Hence, it is important to consider the case study as a potential source of scientifically useful in-

formation and to adopt procedures, where they exist, to increase the strength of case demonstrations in clinical situations when true or quasi-experiments are not viable options.

References

- Agras, W. S., Chapin, H. H., & Oliveau, D. C. The natural history of phobia. *Archives of General Psychiatry*, 1972, 26, 315-317.
- Barlow, D. H. Behavior therapy: The next decade. *Behavior Therapy*, 1980, 11, 315-328.
- Bolgar, H. The case study method. In B. B. Wolman (Ed.), *Handbook of clinical psychology*. New York: McGraw-Hill, 1965.
- Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.
- Cook, T. D., & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976.
- Cook, T. D., & Campbell, D. T. (Eds.), *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally, 1979.
- Garfield, S. L. *Clinical psychology: The study of personality and behavior*. Chicago: Aldine, 1974.
- Gilbert, J. P., Light, R. J., & Mosteller, F. Assessing social innovations: An empirical base for policy. In C. A. Bennett & A. A. Lumsdaine (Eds.), *Evaluation and experiment*. New York: Academic Press, 1975.
- Hersen, M., & Barlow, D. H. *Single-case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press, 1976.
- Kazdin, A. E. Evaluating the generality of findings in analogue therapy research. *Journal of Consulting and Clinical Psychology*, 1978, 46, 673-686.
- Kazdin, A. E. *Research design in clinical psychology*. New York: Harper & Row, 1980.
- Kazdin, A. E., & Wilson, G. T. *Evaluation of behavior therapy: Issues, evidence, and research strategies*. Cambridge, Mass.: Ballinger, 1978.
- Korchin, S. J. *Modern clinical psychology*. New York: Basic Books, 1976.
- Lazarus, A. A. The results of behaviour therapy in 126 cases of severe neurosis. *Behaviour Research and Therapy*, 1963, 1, 69-79.
- Lazarus, A. A., & Davison, G. C. Clinical innovation in research and practice. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis*. New York: Wiley, 1971.
- Masters, W. H., & Johnson, V. E. *Human sexual inadequacy*. Boston: Little, Brown, 1970.
- Paul, G. Behavior modification research: Design and tactics. In C. M. Franks (Ed.), *Behavior therapy: Appraisal and status*. New York: McGraw-Hill, 1969.
- Shapiro, A. K. Psychological aspects of medication. In H. I. Lief, V. F. Lief, & N. R. Lief (Eds.), *The psychological basis of medical practice*. New York: Harper & Row, 1963.
- Strupp, H. H., & Hadley, S. W. Specific vs. nonspecific factors in psychotherapy. *Archives of General Psychiatry*, 1979, 36, 1125-1137.
- Watson, R. I. *The clinical method in psychology*. New York: Harper & Row, 1951.
- Wolpe, J. *Psychotherapy by reciprocal inhibition*. Stanford, Calif.: Stanford University Press, 1958.

Received February 11, 1980 ■