SOME OBSERVATIONS ON CONTROLS IN PSYCHIATRIC RESEARCH

formulated by
the committee on research

Group for the Advancement of Psychiatry
The Group for the Advancement of Psychiatry gratefully 
acknowledges the support of the following foundations:

AMBROOK FOUNDATION
AMERICAN CYANAMID FUND
THE LOUIS CALDER FOUNDATION
THE COMMONWEALTH FUND
THE FIELD FOUNDATION
GENERAL SERVICE FOUNDATION
THE GRANT FOUNDATION
HARRIS FOUNDATION
NATHAN HOFHEIMER FOUNDATION
ITTIELSON FAMILY FOUNDATION
LAIDLAW WELFARE ASSOCIATION
NEW YORK FOUNDATION
NEW YORK POST FOUNDATION
THE MONA BRONTFMAN SHECKMAN FOUNDATION
SMITH KLINE & FRENCH FOUNDATION

Copyright © 1959
Group for the Advancement of Psychiatry
104 East 23rd Street, New York 10, N.Y.
SOME OBSERVATIONS ON CONTROLS IN PSYCHIATRIC RESEARCH

I. INTRODUCTION

A. Psychiatry and Science

The need for a more adequate understanding of human nature has been keenly felt and explicitly declared even in the earliest stages of our intellectual heritage. Modern developments in science and methodology are presently increasing our awareness of what must constitute such 'adequacy'.

In recent decades the field of psychiatry has given much impetus to the scientific study of human behavior, particularly through the impact of psychodynamic concepts.

Significant advances in our knowledge, in psychiatry as in any other field, have been conditioned by a critical understanding of the nature and importance of research. To a vital degree, psychodynamic principles have lent themselves to various methods that are suitable for the construction and testing of hypotheses. Yet many difficulties persist in this respect.

One such difficulty can be related to attitudes held by psychiatrists themselves. Some seem to feel that research, particularly on psychodynamic problems, is unnecessary or impossible or both. The members of this Committee, along with many other psychiatrists, believe that research in psychiatry is both necessary and possible, and have undertaken to outline some ways in which

* The Committee on Research is deeply indebted to various consultants who aided in preparation of this report, especially to its long-term consultants, Dr. Jerome Richfield and Dr. Richard Williams, who played a vital role throughout. Preparation of this report was carried to an advanced stage under the chairmanship of Dr. Alfred Stanton, and completed under the chairmanship of Dr. David Hamburg.
such research may become more adequate to its tasks. We as a profession still have much to learn about how research can be done, and the report deals mainly with this question. But there is the underlying and basic question of what needs to be done. A serious interest in research calls for some balance in awareness of the important problems that need solution and an awareness of relevant methods which offer a good chance for helping to solve a given problem.

In comparison with other fields, psychiatry does not have a strong research tradition oriented to systematic, empirical investigation of important problems. At the same time we do have a great need for immediately applicable working formulations which offer some guide to treatment. These conditions favor quick and often premature closure on many basic questions. Any information vacuum tends to be quickly filled with plausible hypotheses provided by respected authorities. In time it may easily be forgotten that they are unverified hypotheses and they come to be treated as established fact. It is, however, as true in psychiatry as in other scientific fields that authority is no substitute for evidence.

And often the best working formulations we are able to make (including some that seem quite useful clinically) are hypotheses based upon sketchy evidence not derived from systematic study. We must work with such hypotheses in order to meet our clinical responsibilities, but we should recognize what we are doing in order to maintain an inquiring open-minded, constructively critical orientation.

Much attention has been given to the special difficulties facing the psychiatric investigator. An earlier GAP report has discussed some of them in detail. While these difficulties are real and formidable, they should not be used as a rationalization for avoiding the crucial issues that require investigation. Some of the basic difficulties are: (a) the complexity of the subject matter; (b) our limited capacity to manipulate the phenomena we are interested in (i.e., control variables); (c) the easy intrusion of bias; (d) the complex transaction between investigator and subject; (e) the

2. For more extensive treatment of these issues, as well as many other problems discussed in this report, the reader may find the following volumes useful: Jahoda, Marie; Deutsch, Morton; and Cook, Sydney; RESEARCH METHODS IN SOCIAL RELATIONS, Part One: Basic Processes, Dryden Press, N.Y., 1951; Festinger, Leon; and Katz, Daniel; RESEARCH METHODS IN BEHAVIORAL SCIENCES, Dryden Press, N.Y., 1953; and Ingle, Dwight, PRINCIPLES OF RESEARCH IN BIOLOGY AND MEDICINE, Lippincott, Philadelphia, 1958.

3. The extensive controversy, historical and contemporary, as to just what it is that constitutes a scientific study of human behavior has suggested to some a need to formulate a clear and precise meaning for this notion at the outset of our report. The omission of such a formidable undertaking rests upon the view that no genuine clarification of such a basic concept can result from a formal definition. See: Black, Max, "The Definition of Scientific Method", in SCIENCE AND CIVILIZATION, edited by Robert C. Stauffer, University of Wisconsin Press, 1949; and Richfiled, Jerome, "The Scientific Status of Psychoanalysis", in THE VALIDATION OF SCIENTIFIC THEORIES, edited by Philipp Frank, Beacon Press, 1957.

B. The Meaning of Controls

The Committee on Research believes that many of the major problems of research in psychiatry can be approached most readily by a consideration of the nature, uses, and the need of controls. The relation of controls to increased rigor in concept formation, hypothesis construction, and in hypothesis-testing will be examined, as well as limitations in their usage. Both their constructive and their inhibitory effects upon various phases of the whole research enterprise will be considered.

The idea of a control is considered here in its most general sense, in which it may be seen as part of the growth, and now indispensable characteristic of any scientific activity. After establishing the basic relevance of controls to research in psychiatry, this report will endeavor to examine the nature and uses of specific, technical concepts of control in the various research settings which presently confront investigators, principally in the psychodynamic area.

The concept of control is fundamental to the procedures that characterize the sciences. It has had an especially significant role in the historical growth of medicine. Early attempts at the healing art were derived chiefly from the a priori postulates and airy speculations of the ancient Greek cosmologists. But with the distinctive
views of the physicians of the Hippocratic school, the modern conception of a positive science received a rudimentary formulation.

These physicians accepted the aim of science as the justification of belief. But they maintained a keenly critical eye upon the problem of the appropriateness of procedure. Mere cogency of argument from postulates about the world was unsatisfactory to them. To be in the service of humanity, knowledge had to be something more than reasonable. It had to give results; it had to heal the sick. A stricter concept of scientific method was thus evolved, as well as the foundations for distinctive theories of cognition, meaning, verification, and truth. Theorizing itself was approved of only if it lay its foundation in incident, and deduced its conclusion in accordance with phenomena. In a way, the primary problem of Greek medical techne was to understand man's nature and heal his disorders, in spite of the many plausible fictions which tended to hinder solutions to this problem.

Scientific activity did not come into existence until men questioned what they knew, and such questioning took the form of a critical evaluation of the sources, modes, and limits of cognition itself. In this way the concept of 'control' becomes intrinsic to the peculiar meaning of the term, 'scientific'. In its broadest use, the term 'control' refers to any operation which is designed to test or limit any of the conceivable sources of error and distortion in knowledge. These sources of error include, among many others to be considered below, the unvoiced assumptions involved in all hypothetical questions, the influence of our thought-models and attitudinal sets upon the accuracy of our observations, the correctness of inferences, the scope and flexibility of principles of classification, and the precision and clarity of our generalizing language.4

Such a broad, generic concept of control would inevitably have various inherent limits which are worthy of examination. The elimination of all bias and sources of distortion is impossible to achieve. Without inaccessible metaphysical knowledge of reality, science must formulate its goals in terms of 'warranted belief' within some system of interrelated propositions. Scientific knowledge will, then,

remain subject to continuous reconstitution. An acknowledgment of the conditions which require such revisions constitutes a control in as clear a sense as a recognition of the need for, say, a double-blind procedure.

It would be difficult to overemphasize the importance of appropriate controls in the progress of modern science. It would also be difficult to overemphasize the importance of the term “appropriate” in the previous sentence. Controls serve a purpose, and the purpose can only be sensibly specified in terms of the problem under study. A basic question which we must hold in mind while planning research is: control for what?

As indicated above, any procedure built into a research design which systematically checks on a likely source of error may usefully be viewed as a control. But what is a likely source of error in one problem may not be in another. The points at which error is likely to occur depend on the subject matter and the conditions under which it is being studied, especially the extent of well-established understanding of the problem in the field at the time of the inquiry. The appropriateness of controls, then, depends to a very large extent on the research question. What we need to check on depends on what we are trying to find out.

This point is particularly important because some workers in psychiatry and other behavioral sciences have, in their quest for rigor, adopted familiar control measures (such as matched groups) without giving careful attention to their appropriateness to the problem under study. For example, there are studies in which two groups (“experimental and control”) are matched on one or a few variables of minor significance (for the particular problem under study) while variables of major significance to the inquiry (on the basis of clinical experience) are left totally out of consideration — because they are not recognized or because they are thought to have an unwanted complexity. Such studies illustrate a kind of pseudo-control. At best, they are a step in the direction of more rigorous research; at worst, they are misleading because they appear to do much more than they actually accomplish. This kind of rigor can become rigor mortis for research, in the sense that the investigator may be willing to sacrifice the heart of his problem in order to achieve a facade of scientific respectability.

The specification and use of controls, then, presupposes con-

4. The lexical definition of 'control' as a parallel experiment, or a standard of comparison for testing the result of an experiment, is regarded by this Committee as unnecessarily restrictive. Our stipulation for a more general use of the technical term includes the entire denotation of the traditional notion.
considerable knowledge. We must know in principle, for example, what kind of information would constitute a satisfactory answer to the questions implicit in our hypotheses, and we must know what kinds and sources of error might be inherent in the variety of research in which we are engaged.

Thus, before discussing the use of controls, it is necessary to make a distinction between two general types of research, since this distinction has important implications for problems of control. These two types are: (1) Exploratory, hypothesis-developing research and (2) Confirmatory, hypothesis-testing research. These types are often not separable in practice. They frequently overlap and their classification depends upon focus or emphasis. Logically, however, they do represent different stages of the relation of theory to research, and their differences are relevant to an understanding of the nature of a genuine hypothesis.

The collection of specified types of data, the development and delineation of the possible variables in the observed data, and observations of relations between the variables constitute necessary first steps in any scientific endeavor. These first steps are characteristic of research on the level of natural history. They are not a sufficient condition for the achievement of the ultimate goal of science. The descriptive laws giving the functional relations of variables must be explained. To this end scientific hypotheses are constructed. The chief characteristic of any useful hypothesis is that it be testable, i.e., it could have verifiable implications apart from the data whose relations the hypothesis was created to explain. However, the exploratory phases of research are highly important, and must be done thoroughly — a process which may take several years and require the collaborative efforts of numerous researchers. An understanding of the proper time to construct and test hypotheses is itself a control, as broadly defined above, of fundamental importance.

II. AREAS OF PSYCHIATRIC RESEARCH AS RELATED TO THIS REPORT

Research on the basic problems of mental disorders and mental health is of necessity concerned with a wide variety of subject matters. We do not know the proportions of the variance which are attributable to physiological (including constitutional), psychological or social factors. Therefore, the attack must be made on a broad front.

For purposes of this report, the most characteristic subject matter is taken to be psychodynamics and psychopathology. This selection is made because the Committee feels that research in this field is beset by particularly crucial and difficult problems of controls. Intra and interpersonal factors loom large in the research process itself. Many of these same problems are found in research on psychosomatic relations and in psychopharmacology, and references will be made to them. Less attention will be paid to biological research, as such, not because the committee believes it is less important, but rather because the use of controls in this type of research does not present these special problems in high degrees. Distinctive problems of control are also found in relation to social psychology and sociologic factors, and come within the purview of this report.

Thus, we wish to emphasize that this report is not intended to give an exhaustive or detailed account of controls in all areas of research relevant to psychiatry. We wish to concentrate on the logic and rationale of controls which, basically, apply to all areas of research, but which present special problems when applied to psycho-dynamics, interpersonal and social relations, and therapy. Our emphasis is primarily on some of the uniquely psychological aspects of controls.
III. SOME SPECIAL CONDITIONS IN PSYCHIATRIC RESEARCH, AND THEIR IMPLICATIONS FOR CONTROLS

The crucial problems of controls in the area of primary concern in this report, as indicated in the preceding section, stem from some special conditions. An explicit statement of the Committee's views about these conditions is needed to put the major sections of the report into proper perspective.

A. The Components of the Research Situation

Four components can be identified in the structure of any research situation: (1) the observed; (2) the investigator; (3) the instruments of observation; and (4) the setting. While they are relevant for research in any field, it may be profitable to consider which, if any, of their characteristics assume special importance in psychiatric research. With this focus in mind, the last two components may be subdivided and expanded as follows:

(3) Instruments of Observation
   (a) Person or persons
   (b) Apparatus

(4) The Setting
   (a) The culture at large
   (b) The institutional setting
   (c) The research group as a setting
   (d) The therapeutic setting

These subdivisions are introduced because of special conditions which relate to either (1) the type of controls available for use in particular circumstances, or (2) the need for additional provisions to insure adequate control. They will be discussed in subsequent illustrative material and in more detailed consideration of special types of control. Here, we wish to emphasize a characteristic of the interaction between these components in the areas of research with which we are concerned, which is likely to produce sources of error.

B. Special Sources of Error

In any type of research, the emotional involvement of the human investigator is a factor for which appropriate controls may be desirable. In the psychiatric research with which we are concerned, this characteristic almost invariably belongs to some aspect, at least, of the "observed" and may additionally require specific types of control. Furthermore, the instruments of observation and the setting have this attribute in high degree. Thus, the interaction occurring between the various components of the research situation introduces very specific problems for control. It should be stressed that the requirement is control for, and not control of, this aspect of the interaction. Emotional involvement, and the interactional processes stemming from it, are an essential element of many subject matters of psychiatry, and attempts to rule out this element by controls will inevitably change the nature of the problem being studied.

These problems are not as central in other types of information gathering situations such as, say, observing a clock to determine the time it reports. If the common problems in psychiatric research were actually involved in reading a clock, it would be necessary to consider, not only the possibility of the time being mis-read but also these additional possibilities: (1) that the frequency with which the clock is consulted may modify the time it reports; (2) that the time the clock is expected to show may modify the time it actually reports; (3) if the observer dislikes the clock (let us say from an aesthetic viewpoint), it will report time differently than if he is fond of it; (4) if the observer sends someone else to consult the clock, it will report differently; (5) that the time indicated by other clocks adjacent to the one being consulted, or the position of this clock relative to other clocks might influence the time the clock in question reports.
IV. TYPES OF RESEARCH

If, as we have said, the term "control" refers to any operation which is designed to test or limit any of the conceivable sources of error or distortion in knowledge, the type of control particularly appropriate or valuable will vary with the type of research. We will therefore turn our attention to types of research, considering with this, some aspects of the psychology of the research enterprise to the extent that this is related to the problems of controls. We will not undertake, of course, to discuss the substantive issues of each research topic we consider.

When one speaks of "types" of research, one generally thinks of such things as "controlled" laboratory experimentation, clinical investigations or intensive case studies, survey research, longitudinal or cross-sectional research, uni-disciplinary, inter-disciplinary, or multi-disciplinary research and the like. Here, we are using a classification which cuts across all of these more specific types, and which refers to the developmental phases through which the research process — but not necessarily individual research projects — tend to pass. A given project may remain within one "type", or may have elements of more than one type simultaneously.

A. Development of Research — Its Sources

1. The scientific attitude:

A scientific attitude provides the underpinning for effective research. Such an attitude is perhaps better expressed in the activities of researchers than it is in words. In one way or another, we all have some familiarity with vital elements of a scientific attitude — the restless curiosity; the spirit of inquiry; the search for evidence; innovation and careful checking; the striving for understanding, prediction, and control of phenomena that are important in human experience.

The soil out of which psychiatric research grows is composed of the practice of observation, the intrapsychic freedom to see or suspect a meaningful, previously unsuspected connection between portions of what is observed, the curiosity to know the how and the why of this connection, and a need which demands rigorous research action and is not satisfied by speculation. This process of sensing connections is experienced as the feeling of discovery and is called a hunch. The conversion into action of the need to check the hunch, to know the how and the why of the what, is the formal research process and is the subject matter of this report, as it pertains to psychiatry.

We feel that this scientific attitude can be learned by the student and should be a primary goal of education in any scientific discipline. It represents an important baseline of natural, built-in controls which runs through all phases of research. It also increases the likelihood that such endeavors will more consistently result in sound, productive research.

2. The hunch:

Little is definitely known about the sources of hunches. We wish to emphasize them because of a prevailing tendency either to deprecate or to romanticize their role in research. It is our feeling that the hunch is an important element in the development and formal stages of psychiatric research. The hunch is

5. "Hunch" is neither a technical nor a precise term. It is used here to denote those beliefs, generally sudden in occurrence, the grounds for which are inadequately understood at the time. The Committee does not pretend to have a definitive psychology of the hunch, but does wish to present a tentative working viewpoint about hunches which are important to the argument of the report as a whole.

6. Robert K. Merton (Social Theory and Social Structure, Free Press of Glencoe, Illinois, 1949), under the heading "The Serendipity Pattern" discusses the manner in which the unanticipated, anomalous and strategic datum exerts a pressure for initiating theory. A detailed example of the way this actually works out (or fails to work out) is recounted by Bernard Barber and Renee C. Fox in "The Case of the Floppy-eared Rabbits: An Instance of Serendipity Gained and Serendipity Lost" (Am. J. Sociology, 64: 128-136, 1958). Two scientists, each pursuing his own problems, made the independent discovery that after the intravenous injection of papain, a reversible collapse of the rabbits' ears occurred. One went on to make a discovery about the nature of associated changes in the matrix of the ear cartilage based on this serendipitous or chance occurrence; and the other did not. Both scientists were interviewed for the light that could be shed on the process of scientific discovery in general and on the serendipity pattern in particular. An effort was made to elucidate the factors that facilitate or hamper the element of creative imagination that is necessary to complete instances of serendipity by supplying testable explanations of the unusual effect.
a sudden discovery which occurs without the individual who experienced it being immediately aware of the processes involved. Logical reasoning and evidence to support the hunch at first appear to be absent. Hunches may operate on all levels of research requiring psychodynamic judgments. They probably share at least the three following sources and can be classified for expositional purposes according to the degree to which one source predominates over the other two.

a. The pseudo-hunch (hunch of forgotten information).

b. The observational habit (observational hunch).

c. The intrapsychic system of the investigator (intrapsychic hunch).

a. Pseudo-hunch: The pseudo-hunch comes from partially remembered psychiatric learning and simply involves the completion of an old forgotten formula. For example, the psychiatrist had once known that \( A + B = C \) and that the dynamic factor, \( B \), is usually unconscious or withheld. This relation is not clearly remembered; in fact, it may not be part of the psychiatrist's consciousness at all. Still, it was once learned and thus, whether in the pre-conscious or in consciousness, when \( A \) and \( C \) are identified, \( B \) emerges by association. This emergence is experienced as a hunch, but due examination of the process afterwards causes the investigator to recall the previously learned, and then forgotten, psychodynamic knowledge. This is the daily workhorse type hunch, as opposed to the creative one which enables us to suspect or detect new connections and meanings. The investigator sometimes fails to recognize his "hunch" as the echo of previously learned and forgotten knowledge. Our traditional method of reviewing pertinent literature before major research commitments in time and money are made, is a control against the pseudo-hunch. It guards the researcher against the unknowing repetition of work previously done by another investigator.

b. Observational hunch: There is a special quality to some hunch-breeding observations which is difficult to describe and to understand. It is the ability to take a fresh look at a familiar object, to see beyond that which ordinarily traps attention, and to record its relationships with other things in time and space.

An observational hunch probably stems from repeated accurate observations of an offbeat quality — combined with the habit of seeking connections. This hunch comes from different observational experiences, such as repeated observations of patient behavior, which finally allow one to see that several elements are almost always arranged in a certain order. We grow preconsciously to expect that \( C \) will follow \( A \) and \( B \). The realization of this expectation is experienced as a hunch. \( A \), \( B \), and \( C \) may belong to the observer only to the degree that they have become part of his experience with the observed patient. They belong to the patient and need not have previously existed in the investigator's life or personality organization. This hunch arises primarily from astute, continued observation. The investigator's psychic apparatus is involved, not through intrapsychic sharing of patterns with the patient, but rather because it supplied the motive force which developed and sustained the afore-mentioned attitude of inquiry.

c. Intrapsychic hunch: The intrapsychic hunch depends heavily upon the inner perception of the investigator and less so upon his outer perception. It is probably arrived at through:

1. Investigator composition. This occurs when an investigator has a hunch concerning a meaningful relationship between two variables (\( A \) and \( C \)) previously regarded as uninvolved with each other. The dynamic link (\( B \)) is not actually in the material but is supplied by the investigator's personality organization, range of emotional experience or the realization of the relevance of knowledge from other areas.

2. Flexible empathy is the capacity correctly and feelingly to comprehend many different ego states in another human being. It is particularly characterized by the empathic one's ability to be simultaneously inside and outside of the empathic state. This enables him to recreate within himself the emotional impact of \( A \) or to react backwards from the feeling of \( C \). Self-analysis of his own associations or impulses at that critical moment will supply a \( B \) — correct or incorrect.

3. Self-analysis of the investigator's personal reaction to the patient. This is not based on empathy but upon the reactivity of one human being to the meaning of another's behavior. This is the automatic personal response in the interaction between two people.

d. Mixed hunch: This garden variety, commonplace hunch incorporates degrees of the last two types. The observational intrapsychic hunches should be viewed as abstractions in their pure
form. They merge together in the mixed hunch. This is the most important and fruitful combination since our ideal is a high degree of training in a man possessing the characterological attributes conducive to a well developed attitude of inquiry (observational hunch) plus a free system of intrapsychic communication (intrapsychic hunch).

3. The use of controls in relation to hunches:

The observational and intrapsychic hunches are more characteristic of research involving human interaction and the making of psychodynamic judgments, since such research is dependent upon the usage of the complicated intrapsychic apparatus of the investigator as an instrument for observing and understanding the meaning of the patient-subject's productions. There is one consequence stemming from this discussion of sources of hunches which should be understood in the process of working out research designs:

In the case of the observational, or of the flexible empathy hunch, the investigator's emotional investment in his brainchild sometimes makes it difficult for him to face evidence to the contrary. The needs of the researcher are more satisfied by positive findings than by negative ones. Psychiatric research based on psychodynamic judgments of another human being lends itself to unconscious distortions by the investigator in the service of defending a shaky brainchild. This process can be facilitated by ignoring adequate control measures during the early planning phases of research design. Lack of planned controls allows judgment leaks to occur if they ever become essential to the preservation of the hunch's intactness.

Hunches are the germinating points for research leading to discovery, and there is no need for formal controls at this level. We should feel, instead, a freedom to seek new connections. The scientific attitude with its attendant self-disciplines is an important, built-in, natural control system. This section of our report touched upon possible ways of understanding the nature and sources of hunches, so that we can recognize the necessity to control for bias in the original hunch-getter, when the formal phase of research is begun. We trust that our effort will not be misinterpreted as a warning about hunch formation. Here we have been speaking in the context of discovery; later comes the context of justification. Both are absolutely essential in the scientific process.

Before turning our attention explicitly to the use of controls, it is useful to develop the distinction mentioned earlier, between two overlapping phases or aspects of the research process: (1) Exploratory, hypothesis-developing research, and (2) confirmatory, hypothesis-testing research. This distinction has important implications for problems of controls.

B. Exploratory Research

In this instance, the investigator is faced with a problem in which he has little reliable information as to what variables are highly relevant to the phenomenon under study; or perhaps he has some reliable data on one or two variables but feels there are probably others of equal or greater importance in relation to his problem. He then must make an estimate of the situation to determine the feasible and most suitable ways of proceeding, among several alternatives. He may, for example, review the problem empirically, probably surveying a fairly large number of variables in a microscopic way in order to get a more adequate basis for deciding which ones are relevant to this problem. Once he has identified them, he will probably want to carry out a more detailed, systematic study of their nature and relationships — but this comes later, and somewhat different controls will be appropriate than those used in the initial work.

The exploratory phase of research may be likened to the familiar first step in viewing a microscopic section. That is, one first looks at the slide with the naked eye or with the eyepiece reversed in order to gain over-all perspective — e.g., to identify the tissue from which the section has been taken. In the process of doing this, one begins to form a tentative answer as to the nature of the section — or at least a series of likely alternatives begin to emerge from the total (enormous) range of possibilities. In order to decide which of these alternatives is most likely correct, the viewer now turns to greater magnification, checking thoroughly and systematically with detailed observation of those variables earlier identified as relevant to the diagnosis. This latter procedure is similar to the (later) confirmatory phase of formal research.
In practice, the exploratory phase frequently includes movement from an earlier impressionistic stage to a 'tightened-up' exploration that is transitional in character to the definitive confirmatory research. These two overlapping stages of exploratory research may be illustrated by two common patterns of psychiatric research which often form a sequence of research projects.

1. **Incidental exploration.** The investigator develops a relatively speculative hypothesis on the basis of his usual clinical experience, or on a basis of theoretical elaboration. The information on which the hypothesis is based comes mainly from his work with patients, which usually means that it is gathered *incidentally* in the course of treatment. The investigator has not set out primarily to collect this information. He has not been able to set up optimal conditions to minimize bias and other sources of error. He simply noticed something interesting in the course of his work, thought about it afterward, and developed a new idea. Some of the most fruitful ideas in the history of psychiatry have been developed in this way.

2. **Systematic, goal-directed exploration.** If, let us say, a psychiatrist has developed a new idea about the precipitating dynamics of neurotic depressive reactions in the course of his clinical work, and feels that this idea is still relatively speculative, he may then want to put his hypothesis on a basis of greater documentation. This shift toward a relatively well-documented hypothesis will probably necessitate a shift to a more systematic, goal-directed method of investigation, in which additional controls may be quite useful. An example of this type of research might be as follows: the investigator has previously treated a few depressed patients, along with many other patients, and has formed certain clinical impressions; now he decides to put these impressions on a sounder basis by making special arrangements so that he will be able to study (and perhaps treat) a larger number of depressed patients with the help of a colleague who is asked to check his observations and interpretations in each case; they formulate certain specific questions which they hope to answer in each case, and give special consideration to points at which their own personalities and/or theoretical preferences might lead them to unconscious selectivity in observation or interpretation. On the whole, such a procedure would probably favor the development of a more adequately documented hypothesis than was originally formulated, and might well lead to changes in the formulation.

C. **Confirmatory Research**

In this phase of research, the hypotheses are well developed, with adequately defined concepts, and put to tests which usually involve experimental manipulation of relevant variables and/or prediction of consequences derived from the hypotheses. At this point in the research process, the investigator is no longer dealing with impressions of several possible answers to his question, but instead has one definite answer to test. Usually, this answer is a plausible one which could explain the known facts bearing on the research question. It may have grown out of the investigator's own previous (exploratory) work, or it may be an hypothesis prevalent in the field. In spite of its promise, the investigator is aware of the considerable chance that it may be in error — for example, because it was evolved under conditions of observation that allow much room for the operation of bias; or perhaps his own exploratory observations lead him to question an hypothesis which is uncritically accepted by many workers in the field. In either case, he now wishes to set up conditions which will put the hypothesis to a test, giving it a critical, investigative challenge that will be very likely to reveal its errors if they exist.

Basically, this testing involves a prediction. The hypothesis in effect predicts that a certain relationship will be found between the variables under study. 'If so-and-so, then such and such.' Usually, though not always, these predictions are made and checked in an experimental setting. The reasons for this have been stated concisely as follows:7

1. The experimenter makes the event happen at a certain time and place and so is fully prepared to make an accurate observation.

2. Controlled conditions being known conditions, the experimenter can set up his experiment a second time and repeat the observation; and what is very important in view of the

---

social nature of scientific investigation, he can report his conditions so that another experimenter can duplicate them and check the data.

3. The experimenter can systematically vary the conditions and note the concomitant variation in the results."

Although the control of variables in the study of human behavior is generally more problematic than in most of the physical and biological sciences, there have been indications in recent years that experiments may be quite fruitful in the behavioral sciences. Yet it should be borne in mind that the laboratory experiment is not the only setting in which predictions can be made and checked. For example, the experiment-in-nature has a significant place in research. If specified conditions essential to checking a prediction cannot be produced in the laboratory, the investigator can wait for them to occur naturally or seek them out where they are most likely to be found. Thus, the accuracy of his prediction can be determined even though he does not produce the essential testing conditions; the essential point is that he can specify the conditions necessary for a test and these can be independently checked.

The conclusion reached in an exploratory study claims less than the conclusion reached in a confirmatory study. The exploratory conclusion says in effect, "There is some evidence for this finding; it is worth considering; it needs much more investigation." The confirmatory study claims more. It says, in effect, "This finding is very likely correct; it has been checked with exceptionally great care; your confidence in it is justified, and you have a full warrant for your belief." However, we wish again to emphasize that all phases in the development of research are essential, hunches and exploratory research as well as confirmatory research. And, again, an understanding of when it is appropriate to move from one phase to another is in itself an important control.

D. The Relevance of the Distinction to Types of Controls

The distinction between exploratory and confirmatory research does not mean that the former is completely uncontrolled. The research attitude prevails at every step of the way. The investigator's basic orientation in any research should include a consistent critical scrutiny of his work, unceasingly scanning for sources of error. He attempts to formulate the problem as sharply as he can, defines his terms, specifies his operations as far as he can and is alert to the possibility of bias. Therefore, in a basic sense, there is an element of control in any well conducted investigation, however tentative and exploratory it may be.

Thus, exploratory and confirmatory research may best be viewed as two ends of a continuum, with a gradual transition from one to the other. At the extremes a useful distinction can be made between the two, as has been outlined above, but around the center of the continuum are many studies which are in some ways exploratory and in other ways confirmatory. The investigator should provide as much control as is reasonably possible at any stage of the investigation. He is, however, limited by at least three factors in his decision as to what controls are indicated at a given point:

1. His knowledge of the problem. As the investigator's knowledge of the problem increases through his exploratory research (or through the exploratory research of others), it becomes possible for him to devise more and more specific controls based on his knowledge of the problem. They may be added to, or perhaps may even replace, the earlier controls.

2. Economy and feasibility in setting up control measures. In an early stage of research, the investigator may be tempted to set up elaborate controls to check on some possibility, even though he has no strong reason to believe that it would be important. The expense in time and money of undertaking such a procedure is likely to be prohibitive, and in any case it is uneconomical to follow such a plan before one has a grasp of the problem and some good leads as to the variables which are probably most important.

3. Necessity for a wide field of observation in developing adequate hypotheses. In the early stage of any research it is quite useful to scan the problem in an extensive way, getting some overall perspective on it before centering attention on any one variable. If the researcher immediately decides on a priori grounds to concentrate on one variable from the start, there is great danger that he may overlook others which are at least equally important to the question he is trying to answer.

As research proceeds, the situation changes. When the researcher comes up with some tentative answers based on exploratory work and wishes to check these more thoroughly, he will prob-
ably have to focus on certain crucial variables in order to study them exhaustively, since it will usually not be possible for him to study all the variables in a sufficiently exhaustive way to provide rigorous checking.

From time to time, studies appear in the literature which may well be misleading in their use of controls. One difficulty is that, in order to find variables which he can fully control, the investigator may pick out some trivial aspect of the problem to study. That is, he avoids more complex but probably crucial variables which he is less secure about controlling, and then completely controls one or two variables which are very likely insignificant. Another difficulty which occasionally arises is a sort of pseudo-control in which the investigator appears to believe that he has fully controlled the variable by making a gesture in this direction, while leaving out of consideration other crucial aspects of the same variable. Many of these studies appear to be almost exact replicas of experiments carried out in older, well-established physical and biological sciences. They appear to represent a bodily transfer of procedures from other sciences, made rather uncritically without sufficient consideration of the special problems involved in the study of human behavior.

Very little psychiatric research has so far reached the advanced stage of confirmatory work. At the present time, the greatest need is for 'tightening up' exploratory studies, on the road to eventual rigorous testing. The development of suitable methods for carrying out genuine confirmatory studies is, however, one of the great challenges facing psychiatry in the years to come. But we must also guard against the danger of strangling the clan of early exploratory work with a superstructure of controls. Perhaps the greatest danger does not lie so much in pushing prematurely toward true confirmatory research as in failing to recognize exploratory research for what it is and does, and imposing inappropriate controls on it.

E. Illustrative Material: Development of Research in a Problem Area

In order to make this general description of research types more meaningful, and to give a more concrete notion of how they relate to each other, let us consider some specific examples. This will have the advantage of giving a sense of the progressive development of research on a given topic. The topic we have chosen is the effect of early mother-child separation on the child's development. Through the examples of research on this topic, we hope to show the necessity for each type of research (hypothesis-developing and hypothesis-testing) and the way they fit together. We hope to make clear that progress requires both types; neither one without the other takes us very far, but a continuing interaction between the two makes for substantial progress in a field. In clinical fields, sensitive observation and creative imagination of the clinician in the course of his usual work often leads to promising new ideas. This in turn leads to planned, systematic, carefully designed efforts to check out these ideas, to determine where they are accurate, where they require modification, and where they are wrong. Thus, there is a logical movement from clinical impressions to sharpening of formulations and tightening of research designs. Yet, even as the hypotheses are tested new questions emerge, sometimes as a result of the central focus of the research and sometimes through quite incidental observations. Thus, in these examples we hope to give the flavor of the continuing, developing, cumulative process which is research.9

The topic we have chosen for our examples is one which has interested many workers and we must of necessity limit ourselves to a few examples. We regret that there is no space to include a number of fine studies which have been done in this area, both in the study of human behavior and of animal behavior. Bowlby has prepared for the World Health Organization an extensive review of the literature in this field up to 1952. We will draw upon this report at several points, and the reader with special interests may turn to it for detailed review of the evidence up to that time.9

Bowlby states, "During the late 1930's, at least six independent workers were struck by the frequency with which children who committed numerous delinquencies, who seemed to have no feelings for anyone and were very difficult to treat, were found to have

8. In selecting this research area for illustrative purposes, we are not suggesting that it represents an ideal field of inquiry or that the important issues are all settled; we do, however, believe that it is an area in which an unusual amount of systematic research has been done on important questions.
had grossly disturbed relationships with their mothers in their early years. . . . In the same year as Levy's paper (1937) and in the years following, papers were published in the USA by Powlemaker et al. (1937), Lowrey (1940), Bender (1941, 1946 and 1947), and Goldfarb (9 papers 1943-1949), and in Britain by Bowlby (1940 and 1941). With monotonous regularity each put his finger on the child's inability to make relationships as being the central feature from which all the other disturbances sprang, and on the history of institutionalization or, as in the case quoted, of the child's being shifted about from one foster-mother to another as being its cause."

The early papers were largely based on observations made incidentally in the course of clinical work. While careful observation and thoughtful consideration went into each of them, there wasn't the opportunity for extensive checking to determine whether such factors as the observer's bias or the particular sample studied or the particular method of study might lead to erroneous conclusions. Therefore these reports generally have a frankly tentative quality. The need for more systematic, thorough checking on likely sources of error is recognized by these pioneer workers. Toward this end, Bowlby and Bender each undertook rather large-scale, systematic reviews of available case material. Bowlby continues, "Bender's conclusions are based on the 5% to 10% of the 5,000 children whom she had under her care in Bellevue Hospital from 1935-1944 and who showed the characteristics already described. She gives a full clinical description of the syndrome, which she terms 'psychopathic behavior disorder of childhood'."

". . . Bowlby, besides giving fairly full case-histories, in some of which the child's response to the traumatic experience can be traced, lays especial emphasis on the tendency of these children to steal. Dividing all the cases he had seen at a child-guidance clinic into those who had been reported as stealing and those who had not, he compares a group of 44 thieves with a control group, similar in number, age and sex, who although emotionally disturbed did not steal. The thieves were distinguished from the controls in two main ways. First, there were among them 14 children whom Bowlby describes as 'affectionless characters', while there were none in the control group. Secondly, 17 of the thieves had suffered complete and prolonged separation (six months or more) from their mothers or established foster-mothers during their first five years of life; only 2 of the controls had suffered similar separations."

". . . Both Bender's and Bowlby's studies are retrospective in the sense that, as clinicians, they were called upon to examine and treat children showing neurotic symptoms and disturbances of behavior and, by working back into the children's histories, unearthed the common factor of deprivation of maternal care—caused either by their being in institutions, or being posted, like parcels, from one mother-figure to another. The objection to these retrospective studies is, of course, that they are concerned only with children who have developed adversely, and fail to take account of those who may have had the same experience but have developed normally."

Retrospective studies are also vulnerable to distortions of memory. It is possible, for example, that individuals who are experiencing serious difficulty in adult life may wish to provide a plausible, socially acceptable reason for their difficulties. Thus, by describing overt hardships in childhood, they might evoke sympathy from others and minimize their own responsibility. Such a process, perhaps operating quite unconsciously on the part of the patients, might well lead to a deceptive association of adult psychopathology with childhood deprivation. Therefore, it is important to have direct observation of the impact of mother-child separation at the time when it occurs as one step toward determination whether it produces serious disruption in the child. We would now like to present a few, necessarily selective examples of direct studies on this problem.

". . . Spitz, with Wolf, . . . has more recently made a systematic study of the adverse effects which occur during the first year if the child is kept throughout in an institutional environment. They studied altogether four groups of children, in three of which the babies were with their mothers and one where they were not. Though the absolute levels of development, not unexpectedly, differed according to the social group the babies came from, there was no change of quotient during the year in the case of babies, 103 in all, who lived with their mothers. The group of 61 brought up in an hygienic institution, on the other hand, showed a cata-
trophic drop of developmental quotient (DQ) between the ages of 4 and 12 months."

The developmental quotient of the children separated from their mothers dropped markedly during the first year so that by 12 months it was clearly the lowest of the four groups. By the end of the second year it had made an even deeper fall of such an extent as to indicate a serious retardation.

"In confirmation of earlier work, Spitz & Wolf's results show that most of the drop in DQ had taken place during the first six months of life. It is true that these infants were living in conditions especially bad from the psychological point of view, as not only was there but one nurse to some seven children, but, for reasons of hygiene, the children were kept restricted to cots and cubicles in what amounted to solitary confinement. However, studies such as those of Rheingold and Levy make it plain that retardation may occur in conditions which are far from being as adverse as these. Rheingold studied 29 children aged from 6 months to 24 years (mostly between 9 and 15 months) all of whom were awaiting adoption. All had been cared for by foster-mothers; 15 with no other young children, the remainder with up to three others in the same foster-home. Those receiving all the foster-mother's attention were on the average accelerated in development while those who had to share it with other babies were retarded to a statistically significant degree."

The student of human behavior is often prevented by ethical considerations from setting up the exact experimental conditions to provide a test for his hypotheses. To some extent, and sometimes quite impressively, he can get around this limitation by wisely capitalizing on natural occurrences which provide a rough approximation of the conditions that would be necessary for such a test. These experiments-in-nature are quite relevant in the area of research on mother-child separation. We have already seen how Spitz and Wolf have taken advantage of naturally occurring circumstances to make their observations. Another striking example occurred during World War II in England, when Burlingham and Freud set up a residential nursery in the country to care for children who were evacuated from industrial centers that were under attack. Here they had the opportunity to make extensive observations on the impact of mother-child separation. For example, they point out that, in spite of painstaking efforts to ease the transition from home to nursery, rather severe reactions usually occurred in young children. In general, children who underwent the wartime stresses with their mothers fared better than those separated from their mothers.

These authors, like several others working in this area, observed the immediate, short-term after effects of relatively prolonged separation. They were interested in determining whether the disruption of the child's behavior noted during separation would carry over into the period following reunion with his parents. Quoting Bowlby again, "Those most commonly observed are (a) a hostile reaction to the mother on her return, which sometimes takes the form of a refusal to recognize her, (b) an excessive demandingness towards the mother or substitute mother, in which intense possessiveness is combined with intolerance of frustration, acute jealousy, and violent temper tantrums, (c) a cheerful but shallow attachment to any adult within the child's orbit, and (d) an apathetic withdrawal from all emotional entanglements combined with monotonous rocking of the body and sometimes head banging. These reactions have been observed by many clinicians but are nowhere more vividly described than in the two publications of Burlingham & Freud."

Quite recently, Heinicke has taken the observations of Burlingham and Freud, as well as those of Bowlby and associates, as a point of departure for a more systematic study of these same issues, introducing a quantitative element and a number of additional controls. His study, "Some Effects of Separating Two-Year-Old Children from Their Parents: A Comparative Study," is one of the most carefully designed in this area of research and therefore merits special attention. Heinicke points out, "... there was already available a considerable body of data and hypotheses in regard to the problems to be studied. However, none of the data was in quantitative form and the hypotheses had been advanced very tentatively. . . . It has long been recognized that a fruitful approach to the further clarification of the nature of a variable is to think of it as running along a dimension and then to proceed to

explore two or more points along this dimension. We were interested, therefore, in studying the effect of differing degrees of separation. That is, we wanted to be able to say that these effects, and especially any differences between the groups, are probably not simply due to general child development or to the fact that we observed and tested a certain group of children. . . . We chose for the study two different kinds of separation environment—residential nurseries and day nurseries—and studied groups of children in each setting. . . . In terms of time a Residential Nursery child experiences more separation than a Day Nursery child; thus, the primary set of comparisons will be between the two groups. . . . It is important to note that the emphasis of this study is on group comparisons.

"The variables of concern can be grouped in terms of certain general areas. We were interested in how the separated two-year-old child relates to his parents (this, both when the parent is absent and when the parent visits). We also wanted to study how the separated child relates to the human beings in the nursery setting. Then we were interested in how certain separation experiences affect the development of such things as sphincter control."

Heinicke continues, "In order to make comparisons possible we attempted to standardize the criteria for selecting the children to be included in each of the two groups. . . . It was hoped that by thus limiting the nature of the population a certain degree of homogeneity would be achieved. Although cases not falling into these limits are of interest, it was clear that to include all the variety of separation experiences would be too ambitious. . . ."

"The five criteria chosen for the selection of the children were: 1. that the child had had no previous separations of more than three days; 2. that he fell within the age limits of 15-30 months; 3. that he did not enter the Nursery with a sibling; 4. that he was living with both his mother and father at the time; and 5. that there were no obvious indications that he was being rejected by the parents by being placed in the Nursery. Although many other variables are of importance in relation to the phenomena here studied, it was hoped that these others would tend to randomize and thus not be systematically different for the two groups under comparison."

Heinicke goes on to provide a detailed description of the sample, a description of the setting of the research, and an extensive description of the methods used. In connection with the methods, he describes a number of categories which he determined in advance would be relevant to the problem under study. He and a co-worker, who also made regular observations, set out to record observations in each of these categories for each child, thus striving for systematic, even coverage in every case. In this way, they would later be able to make careful comparisons of the various children. He goes on to describe their effort to determine the reliability of the observation categories. He points out, "We wanted to ensure so far as possible that the subjective bias of any one Observer did not influence the results to be analyzed. One method of dealing with this problem is to see that every child is observed by two different Observers, and so far as possible both Observers did observe each child each day. . . . At the beginning of the project and before any case was formally studied, the two Observers observed some children together and then discussed the general nature of the categories and the method of observation very thoroughly. This was a period of definition and training. Once an adequate level of confidence had been attained, a more systematic check on the reliability of the categories was started. This consisted of a series of nine sessions during which the two Observers categorized the behaviour of a child simultaneously."

A detailed analysis was made of the extent to which the two Observers independently agreed on the categorization of behavior. A similar, though less extensive, attempt was made to check on the reliability of other kinds of behavior observations as well. Contact was established with each child in his home prior to separation. Upon admission to the nursery, he was observed every day for about an hour and usually by both observers. In addition, each child was tested with doll play fantasy methods during two major periods: one in the first ten days following admission and the other during the fourth week in the nursery. Thus the observers attempted to get a picture from several vantage points of the changes in behavior over a considerable period of time.

The central finding of the study is briefly summarized as follows: "The responses of the Residential children tend to be more intense and extreme—their seeking of the parents is accompanied by crying, their seeking of relations with staff is of a very
demanding sort, their hostile expressions are more severe, they indulge in more autoerotic activities, and they show a greater degree of breakdown in such areas as sphincter control and illness.” He adds, “Where comparisons are possible our findings parallel those reported previously.”

The development of research in this area may now be seen in some perspective. We began with the clinical impression that psychiatric patients often have experienced significant separation from their mothers in early childhood. We saw that these impressions held up when clinical records were surveyed systematically. We then saw that direct observation of infants experiencing separation showed considerable disturbance, that the disturbance was generally proportional to the extent of separation, and that disturbances persisted during the period immediately following return to the mother. We are still left with the question whether these separations do in fact lead to long-term after effects. Is the adult personality significantly influenced by such separations? Is it not possible that the child has ample opportunity to work out his reactions to a separation during subsequent years of childhood and therefore show no damaging after-effect in adult life? The clinical observations alone do not settle the question, since it may be that the psychiatrists are seeing only those who are damaged and that a much larger proportion of children who undergo early, prolonged separation experiences may emerge unscathed and therefore never come to the attention of psychiatrists. In a major program of research, Goldfarb tackled the problem of long-term after effects. He undertook detailed study of a relatively large group of children institutionalized in infancy and attempted systematically to determine the course of their development over many years. In a recent review, Goldfarb has summarized the main features of his research.

“... One does not normally set about depriving a child of the accepted mode of child rearing in our culture. However, there are infant institutions in which the psychologic climate offered the babies is the polar opposite of that in families. Usually these institutions are physically hygienic living units in which a single adult nurse cares for a large group of babies. The adult-child ratio is small. Further, during each day up to three or four adults may be actively responsible for the child’s care. Over a period of months there are frequent changes in the staff caring for the child. Thus, the child does not have a good opportunity for continuous, intimate or intimate contact with specific adults. Relationships are dim and fleeting. The opportunity for warmth and affection from the same adult person is minimal. Usually the sensory environment is less stimulating than in the family. Finally, the child has less possibility for actively regulating his own life and activities. Group requirements and program needs dominate the individual child’s own wishes and inclinations. There is less opportunity for assertive expression of what is unique in each child.

“... These were controlled studies in which we made a systematic effort to shed light on the problem of psychologic deprivation in infancy. Children reared until about three years of age in an institution and then placed in foster homes (institution group) were compared with children who had been reared continuously in families (foster home group). Both groups had been placed in substitute care in the early months of infancy. The institution group had been reared in an institution outstanding for its standards of physical hygiene. To prevent epidemic infection, babies below nine months of age were kept singly in separate cubicles. They had brief, hurried contacts with adults when they were cleaned and fed by the nurses. During the first year of life, therefore, each child lived in almost complete isolation. During the next two years, the experience was only slightly less impoverished. In these studies, care was taken to demonstrate that the foster homes of both institution and control family groups were equivalent in regard to children’s facilities, economic status, cultural status, sociability, occupational status and educational status. In addition, the foster homes of both groups were equivalent in regard to such subjective aspects of family care as the over-all satisfactoriness of the families for each child and the degree to which the foster families had assimilated the child. Further, the true mothers of the institution children were superior to those of the foster home children in occupational, educational and mental status.
This tended to justify the conclusion, therefore, that differences in intelligence, emotionality and total behavior were a consequence of the differences in early life experiences of both groups.

"The most thoroughgoing investigation was that of fifteen equated pairs of institution and foster home children at a mean age of about twelve years and three months and ranging between ten and fourteen. The institution group had entered the institution at about five months of age and had transferred to the foster home at a mean age of three years and eleven months. They had remained in the foster homes until studied. The two groups of children were individually studied by a long series of tests, experiments and ratings.

"... In brief, the institution children were more retarded intellectually. Of great importance also was the finding that they were distinctly impaired in conceptual ability. ... Along with the cognitive disability there were distinct emotional trends; chiefly, the absence of a normal capacity for inhibition. The institution group showed extremely difficult behavior with symptoms of hyperactivity, restlessness, inability to concentrate and unmanageability. Further, although indiscriminately and insatiably demanding of affection, they had no genuine attachments. They were incapable of reciprocating tender feeling, and their meager love potential was associated with the absence of normal tension in situations which would ordinarily arouse such tension. Similarly, there was an absence of normal anxiety following aggressive or cruel behavior. Finally, the institution children showed specific impairment in social maturity."

An important factor contributing to the development of research in any area is a spirit of free inquiry and criticism in the scientific community. Hopefully, such criticism will be constructive, oriented to issues rather than persons, striving for clarification and adequate evidence. In the field of our present illustration, a good deal of criticism has emerged, some of it polemical, some constructive.

O'Connor, for instance, published a rather extensive critique of the position taken by Bowlby in his World Health Organization review. He cited instances in which he felt Bowlby gave preferential treatment to evidence fitting his own hypothesis and tended to neglect evidence that might contradict his hypothesis—a familiar problem in all fields of scientific inquiry. O'Connor vigorously challenged Bowlby's proposition that permanent damaging effects are very likely to follow maternal deprivation.

That such controversial issues need not degenerate into either polemics is amply demonstrated by Bowlby's most recent study of this problem in which he modifies his earlier hypothesis in light of additional evidence. Bowlby and associates carried out a careful study of children hospitalized at an early age for tuberculosis. While they confirmed that damaging effects do occur in some children, they observed also that psychopathic or affectionless characters are not commonly produced by such experiences. Thus, we move toward a more differentiated concept of mother-child separation and its effects. This naturally raises questions as to the factors that may be relevant to a damaging outcome. In a thoughtful recent review, Glaser and Eisenberg attempt to specify such factors. "The predominant symptoms and their intensity can be expected to vary with the age at separation, the child's experience prior to separation (i.e., the intensity and adequacy of his earlier relationships), the length of the separation, the relationship to adults during the period of separation, the age at which substitution or replacement appears, if it occurs at all, the adequacy of the substitution, and finally, factors as yet unknown, such as the intrinsic endowment of the child."

Future studies will probably be addressed to a further differentiation of relevant factors and their significance in determining long-term outcome. Such differentiation reflects a more adequate analysis of the research problem, but does not remove the importance of the general tendency toward damaging effects of early prolonged mother-child separation. As Eisenberg notes, "The fact... that some children may be able to survive the consequences of maternal deprivation hardly argues its unimportance any more

than the fact that not all children get acute nephritis minimizes
the significance of streptococcal infections."

Thus, the unfolding of research in this problem area, with in-
creasing rigor, has established a general tendency. Children sub-
jected to prolonged separation from their mothers in infancy clearly
tend to show disturbed behavior during and shortly after such
separation; further, some of these children show signs of impaired
mental development many years later. Yet the very establishing
of this general tendency raises new questions. Why are not all such
children so affected? What are the protective factors which permit
recovery and adequate development in some of these cases? Do
any of these children emerge strengthened by their difficult ex-
periences? We may expect new hypotheses to emerge on these ques-
tions, partly from clinical practice, partly from systematic ex-
ploratory research. Perhaps a new cycle will then form as investi-
gators move toward rigorous testing of these emergent hypotheses.

V. A CLASSIFICATION OF TECHNIQUES OF CONTROL

In the preceding sections, we have discussed various scientific
principles which control the research process to increase its rigor
and discipline, and to yield a greater warrant for belief in its
conclusions. We are now ready to examine some of the more
specific techniques of control which are used for this same purpose.

These techniques can be classified in a variety of ways accord-
ing to the types of variability which must be brought within known
limits to provide reasonable answers to research questions. Each
of the components of the research situation discussed in Section
III, above, contain sources of variability which, if left uncontrolled,
may produce bias or error. For present purposes, and following
the general argument in Section III, we are classifying the tech-
niques of control into two major categories according to whether
or not they are directly concerned with or emerge from psychoso-
cial involvements in the research process. These involvements
may pertain to any of the components of the research situation:
observed, investigator, instrument or setting. If such involvement
is a significant aspect of the technique of control itself, we shall
refer to it as a "psychosocial control." If it is not, we shall refer
to it as a "formal control." Formal controls can be and are used
in relation to psychosocial aspects of the situation, but they do
not themselves, intrinsically, contain psychosocial involvements.
Psychosocial controls are also used to control for such factors,
but their distinguishing feature is that they also make use of these
involvements, in the control process itself, and thereby turn a
potential source of error into a constructive instrument.

The remainder of the report will be devoted to each of these
major types of control techniques in turn.
VI. THE USE OF PSYCHOSOCIAL CONTROLS

A. Background of the Problem

Earlier in this report, the special conditions of psychiatric research were pointed out. Attention was drawn to such important considerations as the following: 1) the personality of the investigator may well influence his observations; 2) the responses of the subject may be shaped to a significant but undetermined extent by the impact of the investigator; 3) the emotionally charged significance of the characteristic data of psychiatry may well bring prejudice to the research activities. In other words, it is quite possible that certain factors inherent in human psychology and in the functioning of human groups are likely to introduce error into psychiatric research. Furthermore, these factors may operate in such a way as to minimize our awareness of the probable sources of error. Thus, it becomes important to have ways of maintaining alertness to these risks and of estimating the nature and extent of their influences on research.

Before asking what controls might be appropriate in dealing with these psychological factors, let us attempt to clarify their nature. The psychiatrist shares with other human beings a necessity to maintain his own self-respect as well as the respect of the other people who are important to him. He must protect his essential values, avoid anxiety and depression, and preserve his sense of belonging in whatever group he feels is crucial to his own existence.

Furthermore, we value a predictable environment. We strive for a sense of understanding in relation to whatever challenges us. Being able to offer a plausible explanation for some problem which is important in human life is likely to be a powerful source of self-esteem and of status in society. To have a sense of understanding, even if it is illusory, implies a comforting measure of control and mastery over what otherwise might be difficult, threatening elements of experience. Physicians and others active in the field of psychiatry are assigned an expert role in our society. It is widely anticipated that they will be able to provide solid answers for troublesome questions, and thereby make the environment more predictable. Thus, there is implicit pressure on the person who fills such an expert role to provide answers whether he has them or not. This pressure may well be felt in psychiatric research.

Let us consider a few examples of the ways in which personal problems and group membership may have an adverse influence on the psychiatric investigator. For one thing, he lives in a professional community. He will very likely have some reference group within the community which is of particular importance to him. His self-esteem may well rest on his feeling that he is fully accepted and respected within his reference group. He may feel a strong need to please those of higher status and view deviation as a threat to his membership in the group. He may see research as a means of self-enhancement in the group, particularly if it pleases those in key authority positions. Or, at the other extreme, he may have a dislike or even fear of conformity, and hence he biased in favor of deviance. Thus, his observations may be unconsciously tailored to meet the requirements of his relation to the group.

Similarly, the investigator may use his research in an attempt to work out a variety of personal problems and may be largely unaware that this is happening. For example, if he is a person who has grown quite bitter out of unfortunate experiences in his earlier life, he may displace much hostility through his research and express it in a socially acceptable way. He may attempt an intellectual resolution of his destructive tendencies and use research as a means of retaliation against those who have, at least symbolically, injured him in the past. Thus, his observations may consistently be weighted in a direction which could feed the fires of his anger, all within a well-justified context.

Again, the investigator may use research as self-reassurance. While some elements of this may be present in all human activity, it is more likely to occur in rampant form in those who suffer from

\[16\] We recognize that validity is logically independent of motivation. However, the adverse influences of personal problems may affect observations in ways difficult to check, and prejudice the research at any stage of development.
intense personal insecurity. The investigator may feel a great need to confirm his own belief or those of his group, come what may. He may, therefore, build into his observations a series of self-fulfilling prophecies. In other words, he has a strong tendency to observe only what he had hoped to find in the first place.

These, then, are some examples of the types of psychosocial factors which may introduce serious distortions into the observations of a psychiatric investigator. If they are indeed serious threats to the accuracy of his observation, it then becomes a matter of great concern to find ways of controlling for them. In addition, we should give careful attention to ways in which the psychosocial aspects of the research process can be put to constructive use and indeed, can serve as appropriate controls.

We can be more effective in thinking of appropriate control if we have a clear idea of some of the common ways in which these factors appear concretely in current psychiatric research. Let us, therefore, now briefly consider how they may relate to our chief data-gathering technique, the therapeutic interview. The therapeutic interview serves as a central example here because of its great importance in clinical investigation in psychiatry. Much of our data is obtained in therapeutic interviews, and even diagnostic studies which require only brief contact with each patient are frequently set in a therapeutic context. Therefore, any likely source of error in the data obtained through therapeutic interviews is one which would have wide ramifications throughout the field of psychiatry. 17

In order to maintain a perspective on the therapeutic interview, we must consider its advantages as a data-gathering instrument as well as the likely sources of error which it entails. Several valuable potentialities may be briefly stated as follows:

1) Since the patient is seeking help and often help of a rather profound sort, he is likely to be highly motivated to reveal himself more fully than under most other circumstances, including unconventional aspects and those which would ordinarily escape detection or even awareness.

2) Conditions are provided which facilitate full revelation of personal experience; the intimacy, privacy, and trustworthiness of the therapeutic situation may well permit the patient to examine and reveal aspects of himself which otherwise could not be brought to light, however much he might want to do so.

3) The long-time course of many therapeutic relationships makes possible a richness of data that would otherwise be difficult to obtain, both because it provides very extensive information and because it gives access to fluctuations over time.

4) The therapeutic relationship, particularly if it is a long-term one, permits the analysis of distortions both on the part of the patient and of the therapist. There is considerable opportunity to recognize discrepancy between early and later evidence, to reflect on the material as it accumulates over time, to clarify the relationship of patient and therapist. Thus errors that might otherwise have escaped attention may be detected.

These are potent advantages. They have played and will undoubtedly continue to play an important role in the development of psychiatric research. We should not, however, lose sight of the significant difficulties which are also involved. Therapeutic interviews, especially if frequently repeated over a long period of time, give us a great mass of complex data. In every session and between sessions, a wide variety of factors may be operating to influence the thought, feeling and action of the patient, as well as his relationship with the therapist. This mass of information lends itself to highly selective perception, interpretation, and influence. Selective perception and interpretation is an essential part of all scientific activity. However, it is important to distinguish between explicit selectivity which is relevant to the research questions, and covert selectivity which is relevant to the bias of the investigator. By bias we mean the tendency to collect evidence in such a way that one possible answer to a research question is consistently favored over all others.

The therapist may tend systematically to notice certain things while not noticing others. It is very difficult for him to maintain an even alertness to all the significant events that transpire. In terms of his training, experience, and personality, he is likely to be differentially sensitive to different aspects of the patient's behavior, and to do so outside of his own awareness. This is what we mean by selective perception.

17. For a similar analysis of the advantages and limitations of psychoanalysis as a research method, see: Janis, Irving: Psychological Stress, Chapter 9, "Values and Limitations of Psychoanalytic Research." Wiley, New York, 1958.
But if selective perception does not occur, there are still important possibilities for selective interpretation. He may consistently overlook the possible significance of some parts of his record while tendentiously viewing other aspects as being crucial. With the passage of time, it is almost inevitable that the latter must stand out more clearly in the therapist's mind and will play an increasing role in his therapeutic activity. Those elements which he interprets as being comparatively unimportant may not receive sufficient attention and exploration when they occur in other contexts later in therapy. The therapist will almost surely have sufficient data about the patient to be able to construct a plausible explanation for some of his important difficulties—an explanation which is also bound to have at least some documentation. Alternative explanations, perhaps equally plausible, are likely to suffer progressively as the therapist commits himself to one interpretative line. This may or may not be helpful in therapy, depending on the nature of the case, but it will always be a hindrance to accurate observation and discovery.

Furthermore the therapist responds differentially to the patient. He necessarily takes a greater interest in some aspects of the patient’s behavior than others. He makes more comments about some aspects than others. He implies certain expectations about the patient and communicates, perhaps unwittingly, certain of his own beliefs and values to the patient. In view of the great importance which the therapist often assumes for the patient, these differential responses may have a considerable influence in shaping the long-term course of the patient’s behavior. This, too, may be useful therapeutically; but the hazard from a research viewpoint is that the therapist may be influencing the patient to give the desired responses, without realizing that he is doing so. Finally, the therapist’s retrospective description of interviews may be highly selective in terms of his own bias—so selective recall must be considered as well.

These cautions regarding the possible research risks inherent in the selective perception, interpretation, influence, and recall of the therapeutic situation are reminiscent of a humorous but pertinent remark made by the psychologist Harry Harlow in another context. He said, “In the course of human events, many psychologists have children, and these children always behave in accord with the theoretical position of their parents.” Perhaps the same may be said of our patients. At any rate, in moving toward increasing refinement of our research, we need to consider the possibility that there is some tendency for the patient to behave in accordance with the theoretical preferences and/or personal problems of his therapist, and for the psychiatric investigator to perceive and interpret the behavior of patients along these preferred lines, whether the patients actually do so or not. (The analogy of the clock, in Section III, is pertinent to this argument.)

Illustrative Material: Relation of Research Question, Likely Source of Error, and Control Procedures

Research Question (1)

What are the factors in the patient, in the treatment, and in the life situation that determine the changes that occur during the psychotherapy?

Likely Source of Error (1)

Dynamic theory of personality and psychopathology is sufficiently comprehensive, supple, and (still) imprecise, so that when change is evaluated upon termination of psychotherapy, the changes discerned can usually be retrospectively satisfactorily “explained,” almost regardless of the exact nature and extent and direction of the changes.

Control Procedure (1)

Predictions are made in advance of the psychotherapy about anticipated changes in the psychotherapy. Each prediction is specifically stated, together with the contingencies to which it relates, the assumptions (in data or theory) on which it is based, the inferential process by which it is derived, and the evidence which will later be necessary to confirm or refute it. That is, the evidence which will later be sought in order to test the prediction is pre-determined.

The psychiatrist and patient do not operate in a vacuum. The context in which a study is carried out may have considerable hear-
ing on the outcome and may call for control in its own right. There are many aspects of the research context which could be singled out for discussion in this connection, but we will first briefly summarize four general categories: 1) the research group; 2) the institution in which the research is done; 3) the psychiatric subculture; 4) the general culture.

By the research group we mean those who are directly engaged in gathering the data and trying to make sense of it. There is a variety of ways in which the composition and functioning of the research group may also tend toward selective perception, interpretation, and influence, in much the same way as this occurs in the case of the individual therapist. For example, if the research group has a dominant member who wields great power over the other participants in other areas of their lives, there is a real danger that they may unconsciously fit their observations and interpretations to meet his wishes as they perceive them.

Similarly, the institution in which the research is carried out may influence the results of the research. If, for example, staff members of the institution widely share a strongly felt ideology which assumes certain facts to be virtually incontrovertible, then it will be extremely difficult for any research group, however able, to come up with findings which differ basically from the postulates maintained in the institution. Also, characteristics may be attributed to patients as an enduring part of their personality organization, which in fact are manifestations of the institutional impact on them. Numerous descriptions of schizophrenic patients as being utterly hopeless are to be found in the literature of the not-too-distant past. The patients in the particular institutions which were used for these studies did in fact make little if any progress. In light of subsequent developments, however, it seems very probable that their lack of progress was in large part attributable to the deadening influence of the institutions themselves. Put in a different type of institution, many of these patients might well have made considerable progress. Therefore, the unrecognized impact of the institution led to a seriously distorted formulation of the patients' difficulties.

The psychiatric subculture is important in this connection also. By this term we mean the professional audience toward whom the investigator's communication will be aimed through his presenta-

tion and publications. While it may be assumed that these communications will in general be aimed at what we might call the psychiatric community, there will be individual differences. A given psychiatric investigator may be concerned primarily with a certain segment of the psychiatric community. Whatever his particular reference group may be, the investigator may well consider whether his observations are likely to be unwittingly molded by the unseen audience in ways which lead to bias.

Finally, the culture at large deserves consideration. The same type of issue just set forth in relation to the professional subculture applies to the general culture in which the investigator lives. There is, for example, the familiar tendency of investigators to label whatever is common or prominent in their own culture as an inherent part of human nature. Similarly, conventional behavior that is clearly acceptable socially tends to be regarded as appropriate behavior and so its determinants may avoid scrutiny. In general, there is probably some subtle but pervasive pressure on investigators to come up with findings which will be at least readily acceptable to the culture in which he lives, and perhaps even pressure to come up with findings which appeal so strongly to the values of the culture as to foster excitement, popularity and widespread recognition.

Thus, the observations made by a psychiatric investigator may well be influenced by a variety of social factors including the research group in which he works, the institution in which the research is carried out, the professional audience which is important to him, and the general culture in which he lives.

Now let us turn to a consideration of the ways in which we may minimize or take into account these interfering factors so as to increase the accuracy and reproducibility of our research. We have deliberately sketched the interfering factors in a rather bald way so as to stimulate concern and thinking about them. We have not done this because we regard them as insuperable obstacles to effective research. On the contrary, it is our belief that careful consideration of these difficulties can lead to substantial progress in the foreseeable future. In the following section, we would like to suggest some guidelines for thinking about these problems. It should be understood at the outset that these suggestions are offered in a tentative, exploratory, heuristic spirit, since we do not yet have
sufficient research on methods per se to give hard-and-fast answers to the questions we raise. But the questions need raising, and careful attention to them by many psychiatric investigators will lead to increasingly effective ways of controlling for interfering psychosocial factors, and to more adequate ways of utilizing the facilitating psychosocial factors which are of great significance in research on human behavior.

B. Specific Control Measures

1. Controls for error-producing factors in observation: group and interpersonal methods.

If one is dealing with a complex field of observation in which numerous variables are necessarily interacting, the analysis of the problem is very difficult; and yet it may be possible for a sensitive, knowledgeable observer to detect processes and relations of importance. In view of the various complicating factors mentioned above, two such observers can often provide a more adequate description and understanding of the field of study than either one could by himself. Under optimal conditions, the two observers may have a kind of complementary relationship in which each one provides what the other cannot. Thus, the total range of observation is extended, its accuracy is strengthened, and sufficient perspective is provided to minimize idiosyncratic distortions.

A major value of group methods is to tighten up exploratory research, in the sense of providing consistent though perhaps rough checks on potentially serious sources of error. They may also create secondary gains; talking things over may be a process of socialization which consolidates the research group and sustains it through difficult periods of tedious or disappointing effort. A particularly important advantage of such groups, at their best, is a tendency to move toward the highest common denominator. That is, standards of observational accuracy, honesty, thoroughness, and scientific scrutiny may be set for the entire group by a single member, thus gradually increasing the effectiveness of all participants.

In many psychiatric problems, clinical practice provides ideas which might serve as a guide to the solution. Yet the documentation is often skimpy, the evidence collected in a haphazard way, and there are several competing ideas of about equal plausibility, each of which would account for the phenomenon in question. What, then, should be the next step in research on this problem?

The clinical investigator often cannot expect to be able to control variables in the traditional sense of holding them constant, or eliminating them from consideration, but he does hope to achieve a modest though significant measure of control by taking relevant variables into account. He sees a number of factors, such as his own theoretical preferences and personal problems, which might influence his results and give him a spurious picture of the patient's dynamics. At the same time, however, he is not prepared to invest the time, money, and effort that would be required to make a minute check on every conceivable source of error. Therefore, he decides to move one step forward, a step which seems reasonable and likely to sharpen his observations sufficiently to provide for a more adequate hypothesis than his earlier clinical impressions. He does this, in our previous example, by asking another psychiatrist to check his observations and interpretations in each case; they formulate certain specific questions which they hope to answer in each case, and give special consideration to points at which their own personalities and/or theoretical preferences might lead them to unconscious selectivity in observation or interpretation. This, then, might reasonably be a first significant step from an incidental exploration of ordinary clinical experience to the systematic exploration of a partially controlled study. It seems to us that this is a step which is particularly appropriate and important at the present stage of inquiry into many psychiatric problems.

It should not be imagined, however, that this step can readily be accomplished, nor that it is sure to be an improvement over earlier procedures. There is a fundamental danger in these group methods of control; a kind of pseudo-consensus may arise in which one participant dominates the recording and evaluation of evidence.\textsuperscript{19} Thus, there is in effect only one observer, and yet there is an illusion of others contributing when they are not. In addi-

\textsuperscript{19} An important literature has grown up in the field of social psychology regarding group influences on individual judgments and attitudes. Much careful observation and ingenious experimentation with small groups has brought out clearly that group pressures may drastically affect even simple judgments. See, for example, papers by Asch and Sherif in Swanson, Guy; Newcomb, Theodore; and Harlcy, Eugene: \textit{Readings in Social Psychology}, 2nd Edition, Henry Holt, New York, 1962.
tion to the danger of pseudo-consensus, it must be recognized that these methods tend to be fairly costly in time, money, and effort. While it seems that, in many respects, two heads are truly better than one, it is also often true to say that bringing in additional people may well bring in additional complications.

From what we have said so far about the advantage of additional observers and/or interpreters, it might be inferred that we are arguing that the accuracy of observation increases in direct proportion to the number of observers in a complex situation. While there may be some truth to this within fairly narrow limits, a point may easily be reached beyond which this relationship fails to hold. As groups become large, they tend to become cumbersome and confusing to the participants. Tensions almost invariably arise within the group and in time the strains and complications may well outweigh the advantages. Indeed, as with any means of control in research, a kind of cost accounting is required, in which one balances the gains of the procedure against its limitations.

Group methods of control have considerable usefulness in psychiatric research. They have not been very much exploited in related fields. Indeed, psychiatric research may be able to make a distinctive methodological contribution here. As we now consider a few examples of group methods, we will attempt to delineate both the advantages and limitations which we see as most characteristic of each.

a. Consultation. In the simplest case, a consultant who is familiar with the aims and materials of the research comes in per-

20. Though this entire discussion is written within the framework of the advantages that can accrue from group methods of psychiatric research, we are aware of the many scientific and personal problems that come to the fore in such enterprises; these are even greater when the research is cross-disciplinary, but also exist in a homogenous group sharing a common conceptual framework and a common background of training and experience. There are useful discussions of these issues in (1) Redlich, Fredrick C.; and Brody, Eugene B.: “Emotional Problems of Interdisciplinary Research in Psychiatry.” *Psychiatry.* 18:233-239, 1955. (2) Bush, Vannevar: “Professional Collaboration.” *Science.* Jan. 11, 1957. (3) Simmons, Ozzie C.; and Davis, James A.: “Interdisciplinary Collaboration in Mental Illness Research.” *Am. J. Sociology.* 63:297-303, 1957. (4) Mitchell, Howard E.; and Mudd, Emily H.: “Anxieties Associated with the Conduct of Research in a Clinical Setting.” *Am. J. Ortho.* 27:210-330, 1957. Attention is being called to these articles on the basis that to be forewarned about these inevitable complexities of group research is to be in large part forearmed.

periodically to review the data and to check with the investigator for possible sources of error. His function is to check the investigator constructively, pointing out discrepancy, confusion, bias, and oversight. He usually works from some sort of summary of primary data, which means that a considerable selectivity element has already been introduced by the time he comes in contact with the material. He may attempt to supplement the experience of the observer by providing interpretive guide posts. He may focus on certain possible sources of error which he and the investigator have perceived as a particular point of danger. Ordinarily, he takes no direct responsibility for the conduct of the research; his advice may be taken or not, as the investigators wish.

b. Supervision. If such a consultant has considerably greater experience than the investigator, and if he participates on a frequent, regular schedule, the relationship may become one of supervision. There is opportunity for the checking processes to be more thoroughgoing than in the occasional consultation. The supervisor often has a direct responsibility for the conduct of the research. There is less chance that his recommendations will not be followed. Sometimes his interaction with the person being supervised is included in the data to be studied.

Experience, somewhat paradoxically, presents a problem in consultation and supervision as a means of control. On the whole, experience is valued highly by physicians, largely because of its value in meeting clinical demands. This attitude has become an important part of medical tradition and is evident throughout psychiatry. The experienced clinician is highly respected and for good reasons. However, the general background of clinical experience does not in itself necessarily provide any significant control. Two additional conditions should be met in order for the clinical experience of the consultant or the supervisor to serve a significant function in controlling for the bias or the observational limitations of the investigator.

1. The experience of the consultant or the supervisor should

21. We are speaking here of the research function of the consultant. He may also have other functions in relation to therapy and to training.

22. A thoughtful discussion of the important limitations of supervision as ordinarily practiced, along with suggestions for methodological research on supervision, may be found in: Kuble, Lawrence: “Research into the Process of Supervision in Psychoanalysis,” *Psychoanalytic Quarterly* 27: 229-236, 1958.
be relevant to the problem being investigated. For example, if the investigation is primarily concerned with schizophrenia, there would be severe limitations on the control functions which could be served by an experienced clinician, however distinguished, who had never worked with schizophrenic patients. This is not to say that such contacts might not be quite stimulating to the investigator and quite valuable in the long-run research, but it is doubtful whether they could serve much of a control function in any systematic sense.

2. Experience does not mean closure. If the consultant or supervisor has been able to assimilate his clinical experience within the framework of a genuinely inquiring viewpoint, then he would be much more likely to serve a useful control function than if his experience had led him to crystallize fixed doctrines, in the sense that his views were not modifiable by evidence and he could permit only one possible answer to the research question.

Careful thought is needed as to the reasons for bringing in a consultant. The type of consultant utilized should depend on the reasons why he is needed. However rough a control function he may serve, the basic control question still applies: What are we trying to control for? Thus, a consultant may be needed chiefly because the investigators have had limited clinical experience with the problem at hand, and so they seek out a consultant who has extensive, relevant clinical experience, as we have just indicated. Or, the investigators may operate in a relatively homogeneous institution, so they rarely get different views about the research question; they may therefore need a consultant who has examined the problem carefully from a quite different angle. Or, the investigators may have had ample, relevant clinical experience but know little of research method; hence, they call in a research consultant who has extensive experience in planning effective investigations, even if he knows little of the clinical problem. In the latter case, it is particularly important that the consultant be brought in early in the formulative stage of the work.

There are of course many personal problems which may arise through the introduction of supervisory control. A great deal depends on the competence and judgment of the supervisor as well as his capacity for constructive interpersonal relationship. If the supervisor is a staff member on a higher level in the same institution as the observer, complications may arise because of a confusion of roles. The supervisor may be a person who can influence the observer's assignment, status, and advancement within the institution. This may detract from the freedom of thought and expression which is optimal in research. On the other hand, a person from within the institution is likely to be more familiar with the observer than an outsider would be. This has the advantage that it may facilitate his knowledge of probable difficulties. Furthermore, his familiarity with the social context of the research, and the people conducting it, may well contribute to the salience of his questions and interpretations. Yet, his closeness to the situation may deprive him of the valuable perspective of a newcomer. There are potential advantages associated with each of these procedures—consultant-or-supervisor from within the research institution or from outside it. Experience with these procedures is still insufficient to determine the conditions under which each is optimal. Choice of procedure here must depend largely on the personal qualities of a specific consultant, and on important idiosyncratic features of a particular research problem and setting.

The use of a supervisory technique for the purpose of observing the investigator alters the social system in which the research is proceeding and introduces an additional variable. This is true even if the supervisor does not come in direct contact with the patients, since they may be affected indirectly through his effect on the investigator and consequent alterations in the latter's relationship with the patients. For example, the investigator may become increasingly insecure and may decrease his participation in the research. Or, he may become hostile to the intruder who is looking over his shoulder and may then displace this hostility to the patient. Anxiety, hostility, or competitive feelings aroused in the investigator through his contact with the supervisor may be erroneously attributed to the patient.

Since a variety of emotional problems may arise in connection with supervisory techniques, it is important that the supervisor be alert to the possible meanings which he may have for the investigator, and that their relationship be taken into account in their discussion as a variable in the research. Thus, in the gain-and-cost analysis of supervisory techniques, the advantages are very likely to outweigh the complications. Of course, an infinite regress would be possible in which an additional supervisor would be needed to control for the effects of the relationship of the first supervisor and the investi-
gator, and in turn a third supervisor would be needed to control for the effects of the second supervisor and so on. In practice, this should surely not be necessary if the supervisor is well chosen, and if both the investigator and supervisor are alert to the issues described here.

Illustrative Material: Relation of Research Question, Likely Source of Error, and Control Procedures

Research Question

Do certain drugs (e.g., veriloid) effect a persistent drop in blood pressure of hypertensive patients?

Likely Source of Error

Blood pressure of hypertensive patients may be lowered by other factors which in the usual clinical situation are operative at the same time as the drug. Thus, it becomes the problem to determine whether any observed hypotensive effect is due to a drug or to these other factors. They include:

1. hospitalization
2. intercurrent life situations
3. doctor-patient relationship

Control Procedures

The effect of hospitalization per se could be separated from effect of a drug by observing the patient for a considerable period of time after admission without administering a drug.

Intercurrent situations of major emotional significance could not be eliminated from the patient's life, but could be carefully searched for so that they could be taken into account when they did occur.

Effects of the doctor-patient relationship (e.g., doctor's optimism) could be separated from drug effect by using "double-blind" design—administration of active agents and placebos of identical appearance was alternated in a sequence unknown both to patient and doctor.

In another phase of this research, the clinical investigator was administering known drugs (new hypotensive agents) to hypertensive patients. A special observer was utilized to evaluate the doctor-patient relationship. During a period when this observer classified the investigator's attitude as enthusiastic, the drugs had a striking hypotensive effect. During a later period, following disturbing events in the life of the investigator, this special observer classified the investigator's attitude as unenthusiastic or negative; and the drug effects during this period turned out to be much less pronounced.

c. Multiple observers and multiple data analysis. There are research groups in which two or more people work intensively and systematically on the primary data. In the case of multiple observers, two or more people directly observe and record the entire process which is of interest in relation to the question being asked in the research. In the case of multiple data analysts, two or more people are given the same detailed record, and they attempt to classify the material contained in it. Of course, the two systems may be combined in various ways. The essential point is that a thorough-going effort is made to bring in comprehensive checking procedures through the use of the diverse capacities of more than one person working extensively on the data.

The basic rationale in these methods is simply an extension of what has already been described in connection with consultation and supervision, and with group methods in general. It is the "two heads are better than one" rationale to which we might add "particularly if both examine the primary data quite thoroughly." If the situation being studied is especially complex, this rationale may be extended, as noted earlier, so as to maintain that several heads will be still better than two. Hopefully, the complementary capacities of several individuals may greatly increase the range of observation and the sensitivity to bias.

This happy outcome is by no means assured, however. There are difficulties which may arise in the group setting as we have indicated earlier. The essential point of the group in research is to make available the independent, fully functioning capacities of additional people. If these capacities are constrained by the presence of certain key members, group norms, or other factors in group dynamics, effectiveness is to that extent lost. At the worst, such a group may lead to the danger of a spurious legitimacy. In other

words, the presence of a group may suggest that cross checking is occurring on all major items and therefore that a considerable degree of control of relevant variables is being accomplished. In fact such a group may become a kind of rubber stamp for a dominant member, or for its particular group norms.

There appear to be certain basic interpersonal conditions within a research group which are probably essential for effective function of these methods. It is important that each observer have sufficient independence to defend his judgment when he has evidence to support it; yet, each observer must have the capacity to admit error. If he is hopelessly committed to his own judgment once it is made, there is very little possibility of correcting his error; or, it may be corrected at such a price in terms of intragroup tensions that the process may become disruptive to the research effort. These interpersonal considerations are of great importance in all of the group methods, ranging from occasional consultation through regular supervision and including the more thorough, systematic multiple observer and multiple analyst systems.

Another basic issue in the group methods is the way in which the estimates of various observers are combined into a final estimate. Ordinarily, the investigators hope to emerge with a consensus of the various observers which is more accurate than would be the estimates of any one observer alone. One way to do this is through continuous interaction of the observers. They may discuss their observations at every point, even perhaps from minute to minute, as the events are occurring. This may have the advantage of achieving a maximum sharing of their observational capacities. On the other hand, it has a serious drawback. It also maximizes opportunity for one observer to influence the other. Through continuous interaction, a particularly powerful or facile person can come to mold the judgments of the others in a far-reaching way, or non-rational group norms can exert a controlling influence. These group factors can operate in quite a subtle fashion, and may be largely unrecognized by the participants. If this interdependence method is used, close attention is required to assure that the observers have equal opportunity of correcting each other as the work proceeds.

Another way of proceeding, and one which has certain fundamental advantages, is to obtain an independent estimate from each observer on each relevant variable before attempting to reach any consensus. In other words, each observer is required to commit himself in writing on certain specific issues which have been decided upon in advance as crucial to the question under study. The primary advantage of this procedure is that it does much to eliminate the problem of covert dominance by a single observer. Significant problems remain in arriving at the ultimate consensus, but at the very least this procedure forces a greater independence on the part of each observer and thereby decreases the opportunity for his domination by others. Since each observer is committing himself in writing to a decision on certain crucial issues, he will make every effort to arrive at a well thought-out and soundly documented decision, one which he can readily defend. It will therefore take explicit evidence to change his decision later. Furthermore, he will be free to follow out a line of observation and cannot be distracted from it by the influence of others while the observational period is under way. In other words, the full individuality of each observer is most free to operate under this system, and therefore differences between observers may point up particularly significant phenomena. In addition, some problems will present an opportunity to make systematic comparisons of the estimates made by each observer in relation to some independently measured variable. In psychosomatic research, for example, several observers can make independent estimates of a psychological variable. When a sufficiently large series has been accumulated, the estimates of each observer may be independently related to the physiological process under study. In this way, it may be discovered that one observer has a particularly great predictive capacity in relation to a given problem. It may further be possible to determine the specific set of criteria he uses for his predictions, and thus to arrive at a clarification of the problem through comparative study of independent observers. This possibility is not present in the previously described approach, interdependent estimates.

This leads us to the question of how independent observations may be ultimately combined. If each observer puts his final estimate into a numerical form according to a convention that has been agreed upon by the research group, then these numbers can be manipulated mathematically. This has the powerful advantage that the final pooling of judgment is a mechanical one and therefore no one observer’s rating can count more than any other. Therefore, a research group gets the benefit of two or more fully independent
judgments, and at the same time gets a pooling of these resources into a final consensus. There is, however, an element of arbitrariness in this procedure. If, instead of averaging, the observers are required to discuss their differences in order to arrive at a consensus, each observer must produce the evidence upon which his rating is based and the final rating goes to the more solid evidence. This has the additional advantage of building into the procedure a lever to pry evidence out of the intuitive realm. By trying to provide the evidence upon which these ratings are based, and in the course of discussing the consensus rating, the observers tend to become increasingly aware of the cues which serve as criteria for each rating.

This discussion has so far assumed that each observer had access to the same data — that is, all observers had exactly the same exposure to the primary phenomena under study. This might happen, for instance, in a situation where the observers were interviewing the patients together, or were observing interviews through a one-way vision screen, or were seeing the interviews on sound motion pictures, or listening to them on tape recordings. All of these techniques are fairly difficult to arrange, however, and may be quite costly. There are situations in which it is more feasible to have the patient interviewed by a single person in private. This person then writes a full descriptive account of the interview which is made available to the other members of the research group. They scrutinize it carefully and then discuss it with the interviewer. Here again, the question arises as to the way in which the total resources of the group will be combined into a final judgment. Many combinations are possible and we will not attempt to discuss them all. One procedure is to continue the discussion until something like full agreement is reached spontaneously. This is likely to prove inextricable, particularly if the group is rather large—let us say, five people or more. Another way is to allow perhaps several hours for full discussion and then resolve differences by a vote.

Still another procedure is to give the interviewer maximum opportunity to hear the views of the members of the research group, but give him final authority to make the ultimate decision. In other words, the group proposes but the individual interviewer disposes.

He thus may be able to take into account the observations and interpretations of his colleagues, but he is able to utilize them in the context of his actual experience with the patient. Since he is the only one who had this first-hand experience, and therefore is the only full observer in the situation, there is considerable justification for this procedure. At any rate, having actually experienced sustained interaction with the patient, he is probably less likely to make gross errors in terms of theoretical preconceptions than are his colleagues who are less bound by the actual experience with the patient. Therefore, this procedure essentially widens the horizon of the interviewer while recognizing the primary value of first-hand observation. It also gives an efficient mechanism for arriving at a decision and avoiding interminable discussions. Therefore, it has much to recommend it, both from the viewpoint of observational accuracy and group efficiency.

Illustrative Material: Relation of Research Question, Likely Source of Error, and Control Procedures

Research Question

Are factors in the behavior of infants related to their gastric functions? E.g., does gastric acidity vary with affective fluctuations?

Likely Source of Error (1)

Because there was hardly any previous data on these relations, the investigators would need to observe a wide range of behavioral phenomena. Furthermore, affective processes are often transitory; and the infant cannot report verbally on his experiences.

Control Procedure (1)

The infant was observed repeatedly and for long time intervals. A few situations of primary research interest were repeated time and again. For each period of observation, two or more detailed behavioral records were made independently and subsequently analyzed into discrete categories according to agreed-upon, explicit criteria. Systematic check on the reliability of observation was made

24. The last two techniques mentioned have an important additional advantage with respect to reproducibility. However, they have formidable practical problems of time, effort, and expense.

25. This example is drawn from Engel, George; Reichman, Franz; and Segal, Harry: "A Study of an Infant with a Gastric Fistula, I. Behavior and the Rate of Total Hydrochloric Acid Secretion," Psychosomatic Medicine 18:374-398, 1956.
by determining whether two observers could usually agree independently on categorization of the infant's behavior during a given period. Films were made periodically to permit further checking of observations.

**Likely Source of Error (2)**

Affective fluctuations observed (reliably) in the ways already noted might be caused by the infant's interaction with any of a variety of elements in her current environment as well as by internal processes. One important possible source of affective fluctuation lies in the behavior of the experimenter himself (i.e., the person who consistently comes in direct contact with her, e.g., to remove gastric juice from her gastrostomy). The experimenter is observing her behavior, but may be limited in his awareness of his own behavior and its effect on her because of his personal involvement.

**Control Procedure (2)**

Most of the studies took place in a laboratory with the observers behind a one-way vision screen. Thus, the stimuli impinging on the infant were greatly decreased in comparison to the usual pediatric ward situation. Further, the investigators formulated explicitly the view that “the experimenter is part of the experiment . . . to be observed carefully and recorded by one or more observers behind a one-way vision screen.” Thus, the interaction of experimenter and subject could be taken into account as a variable in the research.

2. **Controls for Error-Producing Factors in the Social Environment of the Research.**

a. **Awareness of problem.** Even in the absence of more systematic measures, many studies can be improved by the investigator's awareness that the findings of his research may well be influenced by the setting in which it is carried out. If, for example, the research is done in a hospital, the investigator may be able to take this into account, provided he has considerable familiarity with the hospital, its personnel, their attitude toward the research, their beliefs about the topic under study, and so on. Unfortunately, many investigators do not indicate any recognition of the possibility that the findings may be significantly related to the setting.

b. **Special observer to focus on transactions between the investigator (or research group) and the setting.** In some studies, the investigator's awareness of this particular problem is not in itself sufficient to cope with it, because complex processes are involved which require study in their own right. Under these conditions, it seems reasonable for the research group to delegate a special observer to focus on transactions between the investigator (or research group) and the setting. In this way, relevant information (such as attitudes of hospital personnel toward the patients being studied) can be systematically collected, rather than left to haphazard impressions.

c. **The same observer(s) in different social fields; i.e., the research group investigates the same problem in different settings.** This is difficult but might be very productive in some circumstances. It is analogous to the study of a given aspect of human behavior (e.g., mother-child relationship) in different cultures by anthropologists. The influence of the setting would be highlighted by differences in obtained results (provided other factors are taken into account—e.g., differences in patient population). An example of current interest is the problem of evaluating the drugs that affect the central nervous system (reserpine, et al). There is good reason to suspect that the effect of a given drug on the behavior of patients is strongly influenced by the setting in which the drug is administered. Thus, studies of the psychodynamic effect of such drugs, or of their value as therapeutic agents, should take into account not only the drug itself and the patient, but also the setting. For example, the results with reserpine and chlorpromazine seem poorer in general hospitals than in state hospitals. It would probably be highly instructive for a research group to set up a study of responses to these drugs in which the same observers using the same methods would work in two such different settings.

3. **Control for Bias: Concealment of Information.**

Much of what has been said in connection with group methods of control of psychosocial factors has touched on the problem of bias. By bias we mean simply the collection of evidence in such a way as to favor consistently one alternative answer to a research question over all others. Bias in psychiatric research has two main sources: the theoretical preferences of the investigator (or his research group), and the personal problems of the investigator. In
either case, one answer, or one set of answers, to the research question may be strongly favored even though the investigator or the research group are unaware of it. In view of the great emotional significance of the characteristic data of psychiatry, the problem of bias is extremely difficult and pervasive.

The group methods make some inroads on this problem. At their best, they offer a modest but significant degree of control for bias. In addition, they do not restrict the observer’s field of observation and do not force him to premature closure. However, when sharply formulated hypotheses have emerged, and they have received considerable documentation in exploratory studies, there is an increasing indication to put these hypotheses to a more rigorous test. The investigator wishes to check and cross-check his impressions, putting in every feasible safeguard against inadvertent error. In this context, surely one of the major sources of error would be the investigator’s bias. Therefore, the phase of hypothesis-testing will usually call for rather extensive, difficult, taxing measures to control for bias. The concealment of information is very important in this connection.

Before proceeding, let us attempt to clarify what we have just said by using a familiar clinical illustration. Sometimes arguments arise as to the relative efficacy of various types of treatment. Let us suppose that we would like to study this question systematically. Assume that we are able to draw patients at random from a common pool so that we can be reasonably sure that any difference in treatment results will not be attributable to differences in the type of patient. The various treatments will be applied to essentially the same kind of patient. Now, each therapist uses the type of treatment that he considers most effective for this type of patient. When the treatment is completed, its results must be evaluated. This after all is a pivotal step in the entire research. If the people who have done the therapy are to make their own evaluation of treatment results without check, there is a real danger that a higher percentage of excellent results may be reported for each group than would occur with independent checks. Since we know that the therapists are personally and professionally committed to the type of therapy they are using, it is difficult to see how we could legitimately place high confidence in the results they report unless we assume an extraordinary degree of objectivity, comprehension, and integrity.

In all likelihood, obtaining their own reports would be better than having no reports at all, particularly if they are required to provide considerable specific documentation for their claims and provided that we can assume that they are honest people. Even so, we would be left with a nagging doubt of serious proportions, since modern psychiatry has made us keenly aware of the remarkable capacity for self-deception in emotionally-charged areas of living, even in the face of conscious honesty. We would have good reason to suspect that, in general, each group of therapists was reporting somewhat better results than they had actually obtained. But how much better? It would sooner or later become apparent that we would gain a great deal by having an assessment of treatment results that would be entirely independent of the therapist’s. Indeed, if it were possible, we should like to have the results evaluated according to explicit criteria applied by people who had no idea of the kind of therapy each patient received and no personal or professional commitment to any particular kind of therapy. While this is clearly desirable in principle, since it would eliminate the factor of bias in evaluation of treatment results, we can at once see that very formidable practical difficulties would arise if we were to undertake such a completely independent evaluation. It may even be argued that, in this particular example, such total independence would be impossible, though perhaps a more limited, yet significant degree of independence could be achieved. The practical difficulties here should not lead to an abandonment of the principle. These difficulties pose a great challenge for psychiatric research; they tax the ingenuity of present and future psychiatric investigators. Yet there are already some areas of research in which such controls are feasible and have been fruitfully applied.

For example, in psychosomatic research the aim of the investigator is usually to determine the relation of a psychological and a physiological process. Increasingly, with the accumulation of experience in this field, the tendency of psychosomatic investigators has been to conceal all information about the physiological variable from the psychiatrist until he has fully committed himself in writing as to his estimate of the psychological variables. The psychiatric data are analyzed and recorded prior to any examination of the physiological data. The reason for this lies in one of the difficult problems of psychosomatic research. The psychological data are
often highly complex and contain elements which may plausibly be interpreted in various ways. If an investigator has before him a set of physiological findings, showing that one out of a series of measurements was abnormal, he is likely to be influenced by this in evaluating the psychological material of the corresponding occasion. In this way, spurious correlations may appear between the two sets of data. There is hardly a day in the life of any person in which he does not undergo some experience that might be regarded as potentially stressful on a symbolic, unconscious level. If the investigator is highly sensitized to such possibilities because he knows in advance that an unusual physiological event has occurred, the chance of his finding something that appears correlative may be greatly increased. Thus, the independence of the two sets of data until each has been analyzed in its own right is a useful safeguard against inadvertent distortions.

However, we wish to emphasize that such a safeguard should not be taken so rigidly as to promote a constriction of observation and creative imagination necessary for the development of hypotheses. Concealment of information is usually quite inappropriate in exploratory research.

**Illustrative Material: Concealment of Information**

In order to check their hypothesis, to validate their hunches, the authors used multiple blind predictions. The hypothesis to which the authors, after a number of observations, arrived is that a spontaneous change in depth of hypnosis occurs when either infantile wishes or hostile impulses are threatening to break through and the usual mechanisms of defense are not adequate; i.e., fluctuations in depth of hypnosis are seen as an indication of endangering the impulse-defense balance of the patient.

The method consisted in presenting two psychoanalysts with pairs of verbatim sections of material gained from the hypnosis of a patient. In one section the patient had reported a change in depth; in the other no change was reported. Care was taken that the neutral section was chosen at random.

In the typed copy of the verbatim record presented to the pre-

26. This example is drawn from Brenman, Margaret; Gill, Morto; and Knight, Robert P.: “Spontaneous Fluctuations in Depth of Hypnosis and their Implications for Ego-Function.” *Internat. J. Psychoan.* 33: 22-33, 1952.
happen that the active drug regularly produces subjective experiences which the inert placebo does not; therefore, patients and/or staff members may come to identify the patients who are getting the drug and the patients who are getting placebo, even if they do not verbalize it. They may then respond selectively, perhaps taking a greater interest in the patients who are detected from subtle cues as being on the drug. This may call for a more complex design in which a sort of "triple blind" arrangement is undertaken, so that there is included in the study a group of patients who are receiving a placebo that has some active ingredient which causes a subjective experience roughly comparable to that of the experimental drug. There are other approaches to this question, and much methodological research is needed here as elsewhere to determine the most valuable approaches to these difficult problems.

The point we wish to make is that anything like a solid, decisive evaluation of drug effects, or of psychophysiological relations, requires efforts of this sort to control for the potent and pervasive element of bias which has so far undermined much of the work on these problems. While the desirability of blind controls in the hypothesis-testing phase of research is impressive, we must keep in mind several problems which arise in connection with the withholding of information. First, as we have briefly noted, there is always a serious question as to whether such secrecy can actually be achieved. For example, if the investigator deals with edited transcripts, extremely elaborate censorship precautions must be set up and rigidly observed. If, on the other hand, the investigator talks directly with a reporter who is to conceal specific aspects of the total information, there is a considerable risk of non-verbal communication in which the investigator may get the desired information in subliminal ways, even though no explicit reference is made to it. There are also serious human problems which arise for investigators, subjects and the morale of the entire research setting, when concealment of information is used over an extended period of time. Perhaps the most serious difficulty is in the use of blind controls at the wrong stage of concept development in research, as we have indicated. These techniques are appropriately used in the direct testing of discrete, clearly formulated hypotheses. If used prematurely, the concealing of information may seriously limit the field of observation, and may result in overlooking many important phenomena which would be pertinent and productive in the exploratory or naturalistic phase of research. Therefore, if the concealment methods are to be useful, they must be applied quite judiciously, and when applied must be used with great thoroughness and determination. A particularly important role in long-term blind studies is that of coordinator. He not only guards against leakage of information ("contamination") but also deals with ethical and morale problems, and provides feedback to the participants at the end.

Illustrative Material: Relation of Research Questions, Likely Sources of Error and Control Procedures.  

Research Question (1)  

Does psychological stress produce increase in circulating concentration of an adrenocortical hormone (hydrocortisone)?

Likely Source of Error (1)  

Factors other than psychological stress are known to influence circulating concentration of hydrocortisone.

Control Procedure (1)  

Patients in whom one of these factors is present are eliminated at the outset. Thus, patients who had any of the following conditions were excluded: (1) active, severe infection, (2) malignancy, (3) endocrine disorder, (4) menstruation, (5) central nervous system drug intake. They were excluded because any of these conditions might influence hydrocortisone level and thereby obscure possible effects of psychological stress on hydrocortisone level.

Research Question (2)  

If hydrocortisone elevation occurs under psychologically stressful conditions, what is the psychological variable(s) with which they are most closely associated?

27. This example is drawn from Price, Douglas; Thaler, Margaret; and Mason, John: "Pre-operative Emotional States and Adrenal Cortical Activity: Studies on Cardiac and Pulmonary Surgery Patients." A.M.A. Archives of Neurology and Psychiatry 77:646-656, 1957.
Likely Source of Error (2)

At the time of this study, there was not much solid evidence available on behavior-hormone relationships. Hormonal fluctuations might be related to any of a variety of psychological variables. Therefore, observation had to cover a wide range of complex phenomena.

Control Procedure (2)

The behavior was studied in detail by two investigators using different methods, in order to determine whether different people using different kinds of data could independently arrive at the same conclusion. To facilitate comparison of their different sets of data, a common rating scale was employed by each investigator in estimating key variables.

Likely Source of Error (3)

In view of the complexity and ambiguity of much of the behavioral data, the investigators might be influenced by prior knowledge of the endocrine data in such a way as to get spurious correlations.

Control Procedure (3)

The three investigators (psychiatrist, psychologist, and endocrinologist) kept all their material separately until all final estimates had been made. Only then were behavioral data and hormone findings related.

VII. THE USE OF FORMAL CONTROLS

The measures just described are those which have been found useful as controls which minimize the bias arising either from the observer or other aspects of the research undertaking. They are particularly pertinent to the fact that psychiatric research is characteristically carried out by people upon people. Psychiatric observers have developed many of these measures directly in the clinical situation and have applied them to the research areas because of their usefulness in the treatment situation. They are, then, more or less characteristic of the dynamic psychiatrist's method of approach.

Historically the dynamic psychiatrist has tended to avoid many other types of research controls which have become increasingly prominent in other fields. Experiment and statistical analysis have been the object of much intensive and fruitful study in recent years. There have been many advances in these techniques and their range of applicability which will undoubtedly prove to be applicable to the psychiatric field. Indeed, a fair proportion of the advances in this area have arisen from efforts to apply formal and mathematical analysis in social sciences.28 Much controversy has occurred concurrent with the development of these newer methods of analysis. Some writers have spoken as if measurement were synonymous with science, and on the other hand, others have spoken as if anything measurable would be unimportant. Much of this difficulty resulted from reasonable but premature application of measurements or special methods of analysis without concern for their limitations or restrictions. To some extent problems have been tailored to meet the measurement tool rather than the more ideal

---

28. In this section of the report, the Committee wishes to give only a brief introduction to complex, technical issues which have been thoroughly covered elsewhere. It is beyond the scope of this report to pursue these matters in detail. We have tried instead to give particularly pertinent references on each topic so that the interested reader may pursue the subject in depth.
use of a measuring tool suited entirely for the problem. Often this is justified, but only if the reason for the change in problem is recognized. In any case the Committee recognizes that many extremists of the recent past have modified their positions, and assumes that the task confronting psychiatric investigators at present is not inherently so controversial—the use of methods of data collection and analysis which are as precise as the problem permits, the avoidance of pseudo-rigor achieved by number magic of any sort, and the constant alert search for more rigorous ways of handling their research problems.

Rigor may be achieved in many ways: by greater clarity in definition, achieved by greater specification of the limits of a concept; by the discovery and elimination of ambiguities in key concepts; by divorcing the data obtained and their treatment from the investigator as a person, so that the data becomes more objective; by handling observations and data by different people to estimate the degree of reproducibility, and in other ways. In many cases, half this battle for rigor is won when there is serious acceptance of its desirability in advance of data collection, careful planning and some industry in refining a research hunch into a research plan. Planning and eliminating ambiguities in advance perhaps can be pressed too early, but should not be avoided entirely.

Controls of this sort used in a research plan are applicable to all types of research, from natural history and survey types of research to the most careful hypothesis-testing experiment. We will consider here various controls available, under the headings of: (A), Conceptualization, (B), Specification of Techniques of Data Collection, (C), Sampling the Objects of Study, and (D), Control Groups in Experimental and Ex Post Facto Studies.

A. Conceptualization

A higher degree of conceptualization of a research hunch will often permit others, and the investigator himself, to proceed about its study in a more systematic way and with greater confidence that the same procedure can be carried out a second time. Uncertainty or actual discomfort accompanies this exercise of secondary process upon a treasured idea, but with real gain. It is often a process which takes considerable time. It has, of course, many aspects but there will be mentioned here—(1), Categorization, (2), The Formal Structure of the Problem, and (3), Measurement.

1. Categorization or classification has become somewhat un-fashionable among psychiatrists in recent years, in the wake of the reaction against the preoccupation with classification which dominated psychiatry following the turn of the century. It is impossible, however, to avoid categorization, and the effort to do so merely leads to bad categorization; classification is implicit in language itself, and particularly in scientific language.

The first step in planning a study will often be simply the awareness of the key categories which are planned. What is the unit—a person, a clinical entity, some perceptual act, a particular personality characteristic before and after some event? Some units are very easily defined, others only with difficulty, some only arbitrarily (How long after childbirth is a psychosis ‘postpartum’?). Yet the nature of the limits set on the definition are important for both investigator and audience. It will often be both practicable and important to estimate the reliability of the definition of the unit by seeing how frequently independent observers differ in their identification of the unit.

It may be necessary to devise a whole set of categories which will exhaustively cover all possible occurrences, and to test out such a set of categories in a pilot study to find if in fact it is exhaustive. In many studies, it is particularly important to be certain, if possible, that the categories chosen are mutually exclusive—that is, that they are on the same logical level.

(Ryle recently spoke of this logical error in many body-mind formulations, using the analogy of a tourist who, having seen every College at Oxford, then asked where the University was.)

We have spoken of ‘devising’ a set of categories when it may be possible to use categories already given wide acceptance. The desirability of using familiar technical terms is obvious but, since many technical terms are used differently by different schools of thought, it is often not safe for critical concepts. Checking with independent observers, and at the very least, full specification of which usage is intended will usually be wise.29

29. The ‘categories’ or ‘units’ we have mentioned are often referred to as ‘variables’; it should be noted that ‘isolation of variables’ is then one of the early steps in research. Such ‘variables’, however, are not necessarily numerical or even measurable.
2. The Formal Structure of the Problem is the corollary of careful categorization of the unit of study. This is a somewhat awkward way of pointing to the nature of the research question. A research hunch, or an observation which the investigator wants to pursue to see if it is a possible general observation, or the testing of an hypothesis, are all different ways of stating the research question. We speak of the formal structure of the problem because it calls attention to the fact that the problem often is not a simple question when it is brought to its most precise formulation. Once the categories are defined (or as they are defined) it will be wise to pay as close attention as possible to the way they are thought to be related to each other.

For a statement to carry any information, or give guidance to research, it must be capable of being proven wrong. Its contradiction, or denial, must be meaningful. Unfortunately, many statements made in the psychiatric literature do not meet this criterion. The investigator should try to conceive of evidence which might refute the proposition in which he is interested, as well as evidence which might support it. (Negative findings may be as valuable as positive ones, but all too often they are left unpublished.)

Failure to achieve this type of precise formulation is usual in psychiatric studies, and should not block an investigator from gathering data; but, again, he should have made efforts enough to establish that he is working at the best level of logical rigor which he can reach. Here also, the explicit effort and serious awareness of this possibility, perhaps with the use of consultants who are not as close to his problem as he is, will often mean that he can be much more rigorous than seemed likely at first glance. This will mean that his work will be more productive for the same amount of labor—or even for less.

3. Measurement is an historically most important refinement of conceptualization in science, and recent years have brought a very considerable increase in the range of measurement techniques applicable to psychological and psychiatric problems. Measurement is essentially a further step in classification. Implicit measurements are widely used in clinical work—rough estimates of severity of disorder, intensity of affect, and the like. Aside from the ability to manipulate figures which makes measurement so central a technique in science, the work involved in finding appropriate measures—without destroying the problem under study—may often be a very effective aid in the clarification and categorization mentioned earlier. To say this is not to suggest that all the most important psychiatric problems now needing study can be treated with refined quantitative techniques. However, attitudes of ignoring any non-quantitative data or analyses, on the one hand, or, on the other, of lightly dismissing the possibilities or values of measurement are both unjustified.

Attention should be drawn, however, to the several different types of measurement which have proved useful, since statistically untrained persons have a tendency to think of measurement only in terms of a highly developed type of scale—like the foot rule. The foot rule is an example of an ‘interval scale’; this means that the scale has equal units of measurement, thus permitting specification of distance between points. The ‘ratio scale’ is still more complex. These scales are highly developed and, where applicable, permit a wide range of mathematical analysis. However, other types of scaling which are less highly refined also have many possible uses, and are often applicable where the interval or ratio scales are not. There are scale types where the simple facts of ‘greater than’ and ‘less than’ can be used; for instance, the power hierarchy in an army will fall clearly into a scale of this type with no use of numbers at all. This is not the place to treat these and other scales at any greater length, but simply to call attention to their existence and value, and to the fact that they can be analyzed mathematically in ways in which non-scalar statements cannot be analyzed.30

Whatever sort of measurement may be decided upon, it will usually need to have its ‘reliability’ estimated if possible upon

materials like those of the study itself. A reliability check determines whether essentially the same results would be obtained if the observations were made again under the same conditions. Various techniques available for this estimation all include in some way the determination of the ability to repeat the findings, by different observers, on different subjects or the like, depending upon which is appropriate. Overlooking studies of reliability has led in the past to launching unreliable tests into the literature where they may last for some time doing the damage of a boat which has slipped her mooring in a harbor.31

The other critical characteristic of a measuring device is its 'validity.' (This is an unfortunate term because the same word is used in logic in a different sense; it has nevertheless become so widely used in methodological literature that it seems likely to remain.) The validity of a measure is the degree to which it either (1) measures what it purports to measure, or (2) is effective in prediction. (These two meanings are thought by operationalists to be synonymous.) It is clear that the questions regarding the validity of a measure are much more complex and often controversial than questions of reliability, and there are no simple or uniform techniques which permit non-controversial estimation of validity in all fields. Many critical problems in this area comprise the heart of difficulties regarding measurement and further study of the area is needed. While much of measurement is subject to criticism regarding its validity, it is also clear that difficulties regarding validity are not such as to permit a hunting license on all measurement techniques indiscriminately. In any event, there is general agreement that one major building block of validation is the testing of predictions from theory.32

Before turning from problems of measurement, it is important to reiterate that inability to measure should by no means imply that one abandons the field of research automatically as unworkable. The choice of a research problem is, of course, a matter which scholars should and will continue to exercise and which the Committee can make few statements about. But we believe that many problems which now offer great difficulties in rigorous treatment are being unwisely neglected, precisely because of this reason.

B. Specification of Technique of Data Collection

The considerations just discussed apply also to the whole area of planning, analyzing, and describing the techniques used in collecting the data. This aspect of the research process has often been neglected, particularly the important feature of the impact of the study itself upon the persons studied, a problem which has been treated in another section of this report. However, a special point needs to be emphasized here: there are now certain standard procedures—psychological testing techniques, designated types of psychotherapy, and the like—and it is the responsibility of the investigator, if he indicates that his material is obtained through the use of one of these techniques, to use the technical terms and procedures of the techniques in their standard form, or else to specify in detail the modifications he has employed. In the case of "psychoanalysis" (speaking of the term as referring to a method of data collection),


32. Validity leads us into fundamental questions of evidence and verification in science. Several recent contributions may be helpful to those who wish to pursue these questions. A clear, concise introduction from a modern, broadened operational viewpoint may be found in Rapoport, Aniel; OPERATIONAL PHILOSOPHY, Harper and Bros., N.Y. 1953, Ch. 3, "The Problem of Verification: Is X True?" For a different viewpoint, see Frank, Philipp: VALIDATION OF SCIENTIFIC THEORIES, Beacon Press, Boston, 1957. A comprehensive discussion by a distinguished philosopher of science may be found in Braithwaite, R.: "SCIENTIFIC EXPLANATION, A STUDY OF THE FUNCTION OF THEORY, PROBABILITY AND LAW IN SCIENCE," Cambridge University Press, England, 1963. Two thoughtful, complex discussions of special aspects by behavioral scientists are as follows: Peak, Hellen: "Problems of Objective Observation," Ch. 6 in RESEARCH METHODS IN BEHAVIORAL SCIENCES, ed. by Festinger, Leon; and Katz, Daniel: Dryden Press, N.Y. 1958; also Cronbach, Lee; and Meehl, Paul: "Construct validity in psychological tests," Psychological Bulletin, 52: 281-302, 1955.
and to some extent in other areas, this requirement has been spoken of as unscientific and authoritarian; but this probably results from a confusion of research requirements with other issues. So far as research is concerned, this is simply a question of accurate specification. An investigator may wisely use any data-gathering technique he has reason to believe is appropriate to his problem, including variations of standard techniques. The solution of some problems may best be pursued by more comprehensive or rigorous use of existing, well-standardized methods; the solution of other problems may require the development of new methods. Our point here is that these methods will be of little value unless they are fully and accurately described.33

While there has, we believe, been a lack of clear statements of the operations involved in many research efforts, the same considerations apply to rigor here as in other areas of this report. The decision that precise formulation is not possible should not be reached with responsibility and with reluctance, but it should not mean the automatic abandonment of the area of study. Whether or not fully operational statements will ever be reached, scientific advance need not await that time.

C. The Objects of Study — Sampling

1. The Nature of Sampling. Medical research started with studying the patients who came to the physician, and much clinical research is still based upon these individuals. Much work carries with it the unspoken assumption that anything found in one case will be found in all. This assumption is often justified, and the tradition of research upon self-selected patients has obviously enough had great success.

In every case, however, what the worker has done under these circumstances is to use a particular type of sample. An awareness of this fact, and of technical advances in allied fields of study, can add significantly to the effectiveness of the research worker. For it is often wise to rely for a time on a self-selected sample in pursuing a line of inquiry, but the times when other types of sample selection are indicated can now be determined with a certain precision. Sampling, again, is a relatively new technology which fulfills reasonable expectations which sampling theory can make quite explicit. One scholar recently remarked that a sample study of the country’s population might well prove more accurate than a complete census, since the unwieldy administrative problems of the latter would be largely avoided.

Scientific sampling has two distinct functions of special importance for this report—(1) the effort to achieve a ‘representative’ sample (more technically to discover the composition of a large population from a sample, with a measurable probability of error) and (2) the effort to achieve a homogeneous sample (or samples) to test a statistical hypothesis in an experiment. We will deal here only with the first, treating the latter as part of the design of an experiment.34

Probably the most important matter for the investigator is for him to be aware of sampling practices. It is impossible not to draw a sample; one has only the threefold choice of using (1) an unspecified sample or a sample of generally unknown characteristics as when one uses self-selected patients; (2) of systematically selecting a sample incorrectly, in such a way that avoidable error is built into the procedure; or (3) of drawing a sample systematically and in such a way that the composition of the sample is independent of avoidable errors and the amount of unavoidable error can be estimated rigorously. In particular, the effort to select ‘typical’ cases is prone to error, in that the cases regarded by the selector as typical are likely to be also those which vindicate his own view. The ‘typical’ cases may be so clearly defined (according to criteria other than those to be tested by the study) that the patients can be selected by another observer. If this is so, the cases are not selected because of being ‘typical’ but because of their specified characteristics.

The question of any scientifically selected sample is what relation it bears to the composition of the population from which it is drawn. If this cannot be precisely estimated, systematic sampling procedure is either valueless or worse—it might be strongly biased and therefore actually misleading. In any event, the decision to

33. An informative discussion of the role of new methods in scientific discovery may be found in Benjamin, John: "Directions and Problems in Psychiatric Research," Psychosomatic Medicine, 14: 1-9, 1952.
34. An exceptionally clear, well-illustrated introduction to sampling and to the field of statistics may be found in Wallis, W. Allen; and Roberts, Harry: STATISTICS, A NEW APPROACH, The Free Press, Glencoe, Illinois, 1956.
draw a particular sample should be governed by the population to which one wishes to generalize as a result of the research.

2. **Sampling in a Sample Survey.** When it becomes important to determine how frequently some item occurs in a population, what the character of an illness typically is, or some similar problem, it is usually wise to proceed by a specified sampling procedure—often called a 'sample survey.'

Sampling for purposes of a survey can be carried out instead of a census in order to reduce cost, to do the study speedily, for greater accuracy, or to give greater depth to the study, since more highly trained workers can do the data collection than in a full-scale census. Along with the type of sample selected, other decisions must be made—the definition of the population to be sampled and determination of data to be collected and of organization of the field work.

Four procedures are now in common use: simple random sampling, stratified random sampling, systematic sampling, and cluster sampling:

**Simple random sampling** is a method of selecting sample units out of the parent population such that every sample unit has an equal chance of being chosen. The samples are drawn unit by unit, often by a table of random numbers, or by drawing from a bowl, or by some other procedure equally free of bias or observer subjectivity. The method is rigorous, well worked out, but tedious and sometimes very time-consuming and expensive. It is often not as sensitive as stratified sampling methods. By less sensitive we mean that a sample of the same size would give less precision in estimation.

**Stratified random sampling** is a modification whereby the accuracy of a sample can be increased without increasing its size. The parent population is divided into subpopulations ('strata'). These subpopulations must be exhaustive—that is, they must include the whole of the original population, even if a 'miscellaneous' stratum must be included. A simple random sample is taken from each stratum, an estimate made for each stratum, and then they are combined to give an estimate for the whole population. The criteria for dividing the population into strata need not be directly related to the question being studied, although if this is practicable, the gain in sensitivity is likely to be greater. Data can be obtained regarding each stratum through this procedure, data which would not be available from simple random sampling; on the other hand, more data is needed to begin with, since something must be known in order to divide the population into subpopulations.

**Systematic sampling** is carried out by some arbitrary device being used to determine which units will be drawn—for instance using every 25th case, or selecting a card every inch in the card pack. It is much more rapid than either of the preceding, and may also be more precise. It is frequently used. It fails if the population being sampled is periodic in any way, and if the systematic sampling procedure happens to coincide with the period; when this happens the samples give no evidence in themselves of the existence of a periodic bias. It may not be possible to compute sampling error. For these reasons, there is disagreement among authorities regarding the wisdom of systematic sampling.

**Cluster sampling** is a procedure derived from census-type studies of population whereby small areas—or groups—of the whole population are used as the units from which to draw samples. It was originally called 'area sampling'. It has quite important uses, but is less likely to be pertinent to psychiatric investigation than the other types, and consequently is simply mentioned here.

3. **General considerations on sampling.** While sampling appears as a short cut or a convenience to the sociologist, the reverse is often true for the medical worker. The additional work involved in selecting a technically drawn sample rather than some easier sample such as a spontaneous one, is rewarded by the greater confidence in general inferences which may be made from the study when it is completed. Sampling, then, like anything else, is best done as a considered part of the research procedure. There are times when it is not worth the effort, times when it is indispensable, and many times when it is a matter of judgment and preference. If it is to be done, it should be done with the care an important scientific tool merits. Modern sampling theory includes ways in which the cost of sampling and the determination of sample size necessary for a certain precision of the sample can be estimated in advance, and

---

it will often be wise for the investigator to get expert advice while planning his work, rather than after it has been started.

D. Experimental Testing of an Hypothesis

The degree and precision of knowledge in certain areas may permit the investigator to formulate some question which is capable of critical testing in either a specially-arranged, planned experiment, or by a retrospective critical treatment of some event which has occurred—the ex post facto design. In either case, he is likely to find his testing best done by the study of the effects of some intervention upon a group of persons. The amount of this particular effect can be estimated (as opposed to the amount of change, which may stem from any source) by contrasting the amount of such effect with the amount of change which has occurred in a comparable group who did not have the treatment or intervention under consideration. This latter group is a control group.

When differences are found between two groups (experimental and control), how are they to be interpreted? When are the differences to be taken seriously and when not? Let us consider this question by quoting a particularly clear statement given by Jahoda, Deutsch, and Cook.36

"When we are contrasting samples or studying the differences between experimental and control groups, we usually wish to test some hypothesis about the nature of the true difference between the larger populations represented by the samples. Most commonly, the statistical analysis is directed toward testing the null hypothesis, which states that the obtained differences between the groups being studied could have occurred as the result of chance alone."37 Suppose the data from our two samples showed that Englishmen attend an average of three motion pictures a month and that Frenchmen attend an average of two and one half pictures. The null hypothesis here would be that this difference might have occurred by chance through sampling error even if there were actually no difference


in the average frequency of movie-going on the part of the English and French populations as a whole.

"If our statistical analysis (called the statistical test of significance) leads us to the conclusion that the observed differences in our samples could have arisen by chance only a small percentage of times in the absence of true differences between the two populations, we reject the null hypothesis; that is, we conclude that there probably is a genuine difference between the two populations. In the social sciences, it is more or less conventional to reject the null hypothesis when the statistical analysis indicates that the observed difference would not occur more than 5 times out of 100 by chance alone.

"If the statistical analysis indicates that the difference between the two samples might have appeared by chance more than 5 times out of 100, the null hypothesis is not rejected. However . . . the failure to reject a hypothesis does not establish its truth. Failure to reject the null hypothesis simply means that the hypothesis that there is no difference between the populations from which the samples have been drawn is now tenable; it does not mean that it is the only hypothesis which can be regarded as being tenable."

1. Composition of Control Groups. Control groups can be selected according to principles reminiscent of those of sampling which have already been discussed. The experimental and control groups may be selected from a parent population by (a) Randomization, (b) Frequency distribution, (c) Precision controls, (d) Self-matching.

a. Randomization is a procedure whereby the units in the total population to be studied are divided into control and experimental groups in a way that gives each individual an equal chance of being assigned to any given group.

b. Frequency distribution controls are used by selecting groups which are matched to each other in terms of the overall distribution of certain highly relevant characteristics. It is important to recognize that these characteristics must be known in advance, since it is best to match on the basis of all the most important factors. Where it is possible to do this, the frequency distribution controls and experimental groups are likely to be more sensitive, but the sacrifice of subjects is likely to be high in the process of matching.
c. Precision controls are those where each individual in control and experimental groups is matched with a partner who shares certain highly relevant characteristics. As the term suggests, this method of determining the composition of control groups is the most precise. However, unless the number of relevant characteristics is small, or the population is homogeneous in terms of these characteristics, the sacrifice of subjects may be so high as to make this method unfeasible. It is generally worthwhile to make at least a trial effort toward precision controls. One can often learn some significant facts about the population under study, even if the trial suggests the method cannot be used.

Either the precision control or the frequency distribution control may be applied in ex post facto matching—such a situation may be useful for instance where an experiment-in-nature has occurred, and two groups are selected, after the event which is being studied, are matched, and are then compared for differences. Ideally, the two groups are quite similar on all relevant variables except one. For example, after a hurricane, two groups of children are selected for comparative study. They are very much alike as to age, sex and socioeconomic background, but differ in this respect: children in one group were with their parents throughout the hurricane; while children in the other group were separated from their parents during the hurricane; the prevalent psychological reactions in the two groups are studied and compared. The contrasting groups could not be picked out in advance of the hurricane, but could be matched afterward. Differences between the two groups might well clarify the functions of the child-parent relationship in protecting against stressful conditions.

The size of control groups necessary for a given precision can, again, be predicted roughly if there is some way of reasonably estimating the degree of difference that may be found.

d. Self-matching. This is one way in which the individual may be used as his own control. There are various ways in which an individual may be compared with himself at different times. For example, a particular kind of behavior may be observed before and after a specific intervention, in order to find out whether this type of intervention has any effect on it; or, a given stimulus may be applied at different times, with one highly relevant condition having changed in the interval. Such comparisons may become cumulative so that group comparisons are possible, but these groups are not composed of entirely different individuals; rather, each individual is represented in the experimental group and in the control group. For example, the experimental occasions may be those in which each person was anxious and the control occasions those in which each person was not anxious.

The basic assumption in these self-matching procedures is that an individual on two occasions (reasonably close together in time) is more like himself than like any other individual. Many intrapersonal variables may be assumed to remain fairly stable, at least over a short period of time, while one or a few crucial ones are allowed to vary and their presumed effects are observed. Under most circumstances, it is unlikely that any two groups of different individuals could be matched as well as each individual can be matched with himself. Self-matching procedures may turn out to have special value in psychiatric research.  

2. Treatment of Control and Experimental Groups. Several types of design have been used. The simplest is the observation of one (experimental) group before and after an event. While the difference may be observed, and confirmed as significant, this provides no evidence permitting one to attribute any change to the event. For this reason a control group, chosen in one of the ways indicated above, is treated identically except for the particular treatment whose effect is being studied. This permits inferences regarding the effect of the specific treatment under study. It suffers, however, if

---


---

the testing itself (part of the experiment) may cause effects and, for this reason, it has been proposed that a third group be chosen which will receive the experimental treatment and the testing procedure which follows the experimental treatment, but which did not receive the preliminary testing; such a control would measure the effect, if any, of the testing procedure.40

The strategy of using control groups cannot be prescribed in general, and is an index of the ingenuity of the investigator in dealing with an individual problem. Control groups are used to measure the effects of possible confounding factors, and to discriminate different elements which may contribute to an observed and measured effect. There have been many different types of experimental design developed in recent years which cannot be described in this report, but the general principle of controls as indicated above are followed.

In psychiatric research, the problem of the “normal” control group is an especially thorny one. By “normal” controls we usually mean individuals who happen not to be in treatment or not manifestly symptomatic. This dimension is often a chance one, and is often not relevant to the variables under scrutiny in the particular research endeavor. The personality variables more relevant to the research question (and therefore more in need of “control”) usually do not have this beguiling simplicity. “Volunteer” controls introduce an additional systematic (but often unknown) distortion in terms of whatever bias or personality predilection has led these individuals to select themselves for participation in the particular experiment.41

Control groups of this type, used for testing a null hypothesis, should not be confused with comparative groups which are often used productively in exploratory research. The simultaneous study of members of two ethnic groups, or of two cultures will tend to


VII. SUMMARY

We have reported our deliberations on the use of controls with the intent to give a perspective of their potential fruitfulness for psychiatric research. We are convinced that this perspective should be a broad one. Any operation designed to test or limit the various sources of error and distortion in knowledge is a control. The broad principles which govern the scientific process, as well as many of the more specific scientific techniques and procedures, are controls in this sense. We are equally convinced of the importance of understanding both their constructive and their inhibitory effects. The appropriateness of controls depends on the questions which the research is attempting to answer. Controls are useless unless they are appropriate to the problem under study.

Particular attention has been devoted to those aspects of psychiatric research in which intra- and interpersonal factors loom large in the research process itself. Emphasis has been placed on some of the uniquely psychological aspects of controls. The emotional involvement of the human investigator is especially important in this context. It can and, in much of the psychiatric research, it does affect all components of the research situation: the observed, or subjects, the investigator, the instruments of observation and the research setting. Also, the psychosocial interaction between these components introduces special problems of control.

The research questions being asked are variable with types or phases of the research process. Therefore, it is important to be able to distinguish these phases, and to recognize both the values and the limitations of each. A common scientific attitude permeates all phases of research, which includes frequent scanning for sources of error. Although we know little about the source of the hunch in science, its constructive use should be encouraged. Hunches should not be stifled by an elaborate system of controls, but they should subsequently be subjected to controlled scrutiny and testing. Investigators should also feel free to engage in exploratory research, both incidentally and in more systematic ways. Significant hypotheses may be developed in this way, which may later be subjected to the more exacting procedures of confirmatory research. However, this process may take a long time, and no individual project or investigator should feel constrained to move toward the use of controls necessary for confirmation, lest he be led to premature closure. There is no sharp line between these types of research. The transition between them is appropriately gradual, and moves in the general direction of narrowing the field of observation and the addition of controls not formerly feasible to use.

The techniques of control have been classified, for purpose of this report, according to the presence or absence of psychosocial involvements in the research process itself. "Psychosocial controls" make constructive use of these involvements, as a significant aspect of the control technique itself. "Formal controls" do not. Both types of controls have their proper uses in psychiatric research.

Psychological factors in the research process contain many sources of bias and error, such as selective perception and selective interpretation. They also have many potential advantages, as, for example, in the therapeutic interview, which yields types and levels of data which cannot be had in any other way. The social dimensions of these factors, in groups, institutional and cultural settings, also have both dangers and advantages for the research process. The solution to the problem, so that advantages outweigh disadvantages, consists not in ruling these factors out by controls, but in putting them to constructive use. Consultation and supervision, multiple observers and other group methods, special observations of the social environment, and concealment of information can all, potentially, contribute to bias-free findings in which belief is warranted. But each technique contains its own new sources of bias. They are appropriate in some situations and in relation to some research questions, and not others. This report has attempted to indicate the uses and misuses of each one.

The more formal controls, common to all fields of science, are also potentially useful in psychiatric research. Three general categories of controls of this type are inherent in the scientific process in all of its phases, and must be explicitly recognized: conceptualization, specification of data collection, and sampling. The specific
type of control selected within these categories should vary with
the research question and the stage of the research. But it should
always be specified, and not left on an implicit, or unrecognized,
level. A fourth general category, the use of control groups in expe-
rimental designs, is appropriate only in confirmatory aspects of
research, and should be avoided in the exploratory. The report has
indicated the nature of several specific techniques of control within
each of these categories, and has emphasized ways in which to recog-
nize when a particular technique is appropriate and when it is not.

Controls are an integral part of the process of scientific research.
If this report has done no more than provide sharper focus and a
balanced perspective concerning their use in psychiatric research,
it will have served its purpose.
Group for the Advancement of Psychiatry

The Group for the Advancement of Psychiatry has a membership of approximately 183 psychiatrists, organized in the form of a number of working committees which direct their efforts toward the study of various aspects of psychiatry and toward the application of this knowledge to the fields of mental health and human relations. GAP is an independent group and its Reports represent the composite findings and opinions of its members only, guided by its many consultants.

Collaboration with specialists in other disciplines has been and is one of GAP’s working principles. Since the formation of GAP in 1946 its members have worked closely with such other specialists as anthropologists, biologists, economists, statisticians, educators, lawyers, nurses, psychologists, sociologists, social workers, and experts in mass communication, philosophy, and semantics. GAP envisages a continuing program of work according to the following aims:

(1) To collect and appraise significant data in the field of psychiatry, mental health and human relations;

(2) To re-evaluate old concepts and to develop and test new ones;

(3) To apply the knowledge thus obtained for the promotion of mental health in good human relations.

Committee on Adolescence

Committee on Aging

Committee on Child Psychiatry

Committee on College Student
Beauan R. Snyder, Boston, Mass.; Alfred Fleissig, Chicago; Alan Frank, Los Angeles; Daniel H. Fumonstein, Boston; Charles Fuchsherr, Washington, D.C.; Robert K. Zeiter, Madison; Bryant M. Wedge, New Haven; J. B. Wheaterlgh, San Francisco

Committee on Cooperation with Governmental (Federal) Agencies
John M. Caldwell, Jr., Miami, Fla.; Benjamin H. Rabin, New York; Norman C. Bell, Los Angeles; Elmer L. Green, Birmingham; Calvin S. Draper, Philadelphia; Robert H. Felix, Bethesda; Albert J. Glass, Washington, D.C.; Edward O. Harper, Cleveland; Edward J. Koller, Los Angeles; John R. Halden, Bethesda; Donald B. Peterson, Anoka, Minn.; Raymond Wagoner, Ann Arbor

Committee on the Family
John E. Spiegel, Cambridge, Mass.; Sidney Berman, Washington, D.C.; Murray Bowen, Chevy Chase; David Mendel, Houston

Committee on Hospitals
Alfred P. Bay, Topaz, Calif.; Walker E. Barton, Boston; Robert E. Basnett, Princeton; Daniel Blain, Sauk Centre; Robert S. Garber, Princeton; Paul Haus, Morristown, Pa.; George W. Jackson, Topaz; Frances J. O’Neill, Central Islip, N.Y.; Humphrey F. Osmond, Sandwich; Lucy B. Graile, Kansas City; Lee O. Sewell, Perry Point, Md.; Harvey J. Tompkins, New York; Cecil L. Wilso, Omaha

Committee on International Relations
Mohammed Torqu, New York, N.Y.; Francis F. Barnes, Chevy Chase
Additional copies of this GAP Report No. 42 are available at the following prices:

- 1 - 9 copies: $1.00 each
- 10 - 99 copies: $0.80 each
- 100 or more copies: $0.60 each

Please send your order including remittance to:

Publications Office
Group for the Advancement of Psychiatry
104 E. 25th St., New York 10, N. Y.