

Prologue

It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong.

—Richard Feynman, Nobel Laureate

In all affairs, it's a healthy thing now and then to bang a question mark on the things you have long taken for granted.

—Bertrand Russell

I may be wrong, but I doubt it.

—Charles Barkley, former NBA star, and many, many others

I realize that it defies publishing convention a little by actually introducing a prologue, but because this article was so central to everything else in this book, I thought I'd give it a shot.

“Is Casework Effective?”, published in 1973, sets the more or less empirical tone for this book by presenting a review of the early outcome research in social work. This review has two dire conclusions. The first is that there appeared to be no studies with adequate methodology that could lead to a conclusion that anything social workers did led to demonstrably positive changes in clients.

The second conclusion was that many of the clients of professional social workers actually did worse than similar clients in control groups or in groups treated by nonprofessionals. I called that phenomenon “the deterioration effect,” in line with what some researchers had found in the outcome studies of psychotherapy by non-social work professional personnel (Bergin, 1971).

This review of research was very stimulating to me, and to others in the field. I am using the term “stimulating” in all its possible meanings. I won't discuss that result any more here, since that impact is the entire topic of the epilogue. Suffice it to say that that article got that part of my professional career off to a bang.

Reference

- Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In Bergin, A. E., & Garfield, S. L. (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis*. New York: John Wiley and Sons.

Is Casework Effective?

A Review

The core of professional practice is a commitment to competence—a commitment that most directly refers to a concern with the effective carrying out of professional services. Unfortunately, social casework, the largest segment of the social work profession, has been criticized consistently and most dramatically for its failure to demonstrate clearly effectiveness in helping clients.¹ Much of the criticism leveled at casework, however, has been based either on ideological grounds, with little apparent concern for research data to support such criticism, or on an inadequate review of research, for example, using only one study, from which the critic attempts to draw conclusions for the entire profession. One can hardly be confident in conclusions derived from such methods of evaluation.

Although there was a flurry of interest in the question of casework effectiveness raised by the publication of *Girls at Vocational High*, the issues raised at that time are far from settled.² In fact, they never have been thoroughly discussed. It seems as if, by some tacit arrangement, major contenders in the issue of effectiveness had agreed to let the matter drop.

The thesis of this paper is that the issue of effectiveness of practice always must be of paramount concern to the profession and cannot be brushed aside. A convergence between the professional values of commitment to the scientific method and the desire to promote capably the well-being of our clients demands such a stance.³ It is surprising then that although the issue of effectiveness frequently is a topic of discussion, and there have been some attempts to examine aspects of the research on this subject, no comprehensive review of all the available

Fischer, J. (1973). *Social Work*, 18, 5–20. Reprinted with permission of National Association of Social Workers.

¹ See, for example, Scott Briar, "The Current Crisis in Social Casework," *Social Work Practice*, 1967 (New York: Columbia University Press, 1967), pp. 19–33.

² Henry J. Meyer, Edgar Borgatta, and Wyatt Jones, *Girls at Vocational High: An Experiment in Social Work Intervention* (New York: Russell Sage Foundation, 1965). For examples of the issues raised at that time, see Earl Ubell, "Social Casework Fails the Test," *New York Herald Tribune* (October 4, 1964); and Briar, op. cit.

³ Scott Briar and Henry Miller, *Problems and Issues in Social Casework* (New York: Columbia University Press, 1971).

major evaluative research on casework effectiveness is available in the social work literature.⁴

This article is an attempt to provide such a review. Its aim is to examine casework effectiveness in such a way as to generate reliable conclusions that can be scrutinized and tested through independent investigation. Utilizing analytic criteria of demonstrated validity, this review will present the findings of major extant evaluative research and will extrapolate from these studies conclusions as to whether professional casework practice has indeed been found to be effective.

What Is Social Casework?

To draw conclusions about how effective casework is, it is first necessary to consider just *what* casework is, that is, what is to be examined. Hartman poses this well:

Because people who define themselves as caseworkers define the practice so differently, and because no one has been elected to determine the definition, I assume that we can all carve out our area, practice it, teach it and write articles about it as long as the community, clients, universities and editors will support us.⁵

She also reviews a number of definitions of social casework that reflect the major streams of casework since its earliest days.

However, for research purposes, the definitions reviewed by Hartman neglect a most crucial variable—exactly what it is that caseworkers do. Complicating this problem is the increasing recognition that caseworkers do many things in many ways, all of which legitimately can be called casework.⁶ This confusion in specification of casework methodology, to paraphrase Raimy's definition of psychotherapy, points to a view of casework as a set of undefined techniques, applied to unspecified problems, with unpredictable outcome. For this approach, rigorous training is recommended.⁷

In a most general sense, then, casework could be defined—at least for the purpose of reviewing studies that evaluate casework—as the services of professional caseworkers. Specification of the details of these services generally has been held to be less important than agreement that the services should be provided by persons whose educational qualifications have met the standards of the profession.

⁴ Attempts to examine aspects of this research have been made by Scott Briar, "Family Services," in Henry S. Maas, ed., *Five Fields of Social Service: Reviews of Research* (New York: National Association of Social Workers, 1966); Briar, "Family Services and Casework," in Maas, ed., *Research in the Social Services: A Five-Year Review* (New York: National Association of Social Workers, 1970); and Ludwig L. Geismar, "Implications of a Family Life Improvement Project," *Social Casework*, 52 (July 1971), pp. 455–465.

⁵ Ann Hartman, "But What Is Social Casework?" *Social Casework*, 52 (July 1971), p. 419.

⁶ Robert W. Roberts and Robert H. Nee, eds., *Theories of Social Casework* (Chicago: University of Chicago Press, 1970).

⁷ Victor Raimy, ed., *Training in Clinical Psychology* (Englewood Cliffs, N.J.: Prentice-Hall, 1950), p. 93.

And these qualifications traditionally have consisted of a master's degree from an accredited graduate school of social work (MSW).

The implication is that educational criteria relate to a presumed basic minimum competence in the practice of casework for all those who have been educated as caseworkers, but that it is not necessary to specify the exact nature or kind of casework. Thus any conclusion about the general success or failure of casework reached from reviewing the research can be made only if two conditions are met: (1) the services evaluated are performed by professional caseworkers and can be shown to have some central core of relevance to casework practice and (2) success or failure is the *rule* in the studies evaluated, cutting across a variety of clients, approaches, and situations. Although the issue of specification of practice methodology is important, lack of specification does not preclude drawing conclusions on a broader level—the level that examines the effectiveness of services offered by professional caseworkers, no matter which techniques and methods have been used in these services.

Almost as difficult as defining casework, however, is the problem of specifying just what is meant by “effectiveness” (or “success” or “improvement”). Obviously, the effects of intervention can show up in a number of ways, from subtle psychological changes to objective, observable changes in school grades, delinquency rates, and other performance dimensions. There might be some validity in drawing general conclusions about the effectiveness of casework from changes in only a few measures of outcome, since those few measures might really be the only appropriate indicators of the kinds of changes casework services are capable of producing. However, the scope of potential changes resulting from casework intervention would suggest that one would have more confidence about conclusions when positive changes can be demonstrated using varying types of criterion measures in one study and across several studies.

Actually, the selection of outcome indicators is a task that must be determined in advance in each study.⁸ Effectiveness would then mean that differences in scores significantly favor one group over another in achieving a goal specified in advance by the researcher. Thus this review is constrained by the fact that results can only be reported in relation to the measures included in the primary investigations, even though there may have been other unknown, potentially important effects of the services.

Selection of Studies

The purpose of a study of casework effectiveness is to examine whether the services were successful in helping clients.⁹ A minimum requirement for establishing that whatever changes in clients could be found were actually a result of the

⁸ Elizabeth Herzog, *Some Guide Lines for Evaluative Research* (Washington, D.C.: U.S. Department of Health, Education & Welfare, 1959).

⁹ This type of study should be distinguished from other types of outcome research that might examine the *effects* of services (e.g., along such dimensions as continuance-discontinuance) which may be unrelated to the question of whether the services were *effective*.

specific services provided is the use of a control procedure. So evidence of change in clients is not necessarily evidence that the changes came about because of the casework services, and evidence of no change cannot be taken as a demonstration that the services had no effect (e.g., that intervention might have prevented deterioration). In either situation the researcher cannot draw definite conclusions unless some form of control has been introduced to minimize alternative explanations. As Nagel points out succinctly:

... data must be analyzed so as to make possible comparisons on the basis of some *control* group, if they are to constitute cogent evidence for a causal inference. The introduction of such controls is the minimum requirement for the reliable interpretation and use of empirical data.¹⁰

Therefore, a minimum requirement for selection of studies for this review was that some form of control group of clients was utilized in the study.

Beginning with recent reviews, major social work journals, dissertation abstracts, and unpublished agency reports were surveyed from the 1930s to the present. Over seventy studies were located that purported to examine the effectiveness of casework services. However, although these studies contained much valuable information, most neglected to include a control group in their design. Because of the difficulty in drawing a valid conclusion regarding cause and effect without a control group for comparison, the bulk of these studies had to be excluded from this review.

Two major types of control were utilized in the studies eventually selected: (1) untreated control—a group that purportedly received no treatment at all and (2) a specific form of “other-treated control.” In the second type of study the experimental group received the services of professional MSW caseworkers and the control group received services from nonprofessionals (e.g., non-MSW public assistance or probation workers). Despite obvious differences in the two categories of studies, certain assumptions basic to professional education and practice are utilized in this review.

Essentially, these assumptions are as follows: given client groups with similar problems appropriate for social work intervention (1) caseworkers with professional degrees should achieve more successful results than nonprofessional workers and (2) a program of professional intervention should achieve more successful outcome with clients than either no treatment at all or nonspecific or haphazardly selected treatment. Considerable research points to the fact that there are few pure control groups. Even when nominally in a control group, people often seek help from a variety of sources, such as family, friends, the clergy, and so forth. In such cases it is assumed that a program of professional intervention should, on the whole, achieve more efficacious results.¹¹

¹⁰ Ernest Nagel, “Methodological Issues in Psychoanalytic Theory,” in Sidney Hook, ed., *Psychoanalysis, Scientific Method and Philosophy* (New York: New York University Press, 1959), p. 53.

¹¹ See Allen E. Bergin, “The Evaluation of Therapeutic Outcomes,” in Bergin and Sol L. Garfield, eds., *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis* (New York: John Wiley & Sons, 1971).

Thus in line with the definition of casework as the services offered by professional caseworkers, this review will attempt to ascertain whether such services have been found to be more effective than no treatment or other nonspecific or nonprofessional services with which they have been compared.

Several other types of studies were excluded from this review in the hope that their omission would permit greater precision in drawing conclusions by minimizing potential biasing and the confounding effects which could have occurred if they had been included. Studies examining casework services outside the United States proper were not included.¹² Since the effectiveness of MSW caseworkers was the object of attention, several well-known studies examining only the services of nonprofessionals also were not reviewed.¹³ Those studies that examined variations in types of professional casework without utilizing an untreated or nonprofessionally treated control group were excluded as well.¹⁴ Further, those studies in which it appeared that caseworkers were only a small minority of the treatment team providing services to clients in the experimental group were omitted.¹⁵ However, when there was lack of clarity in the text of the report as to certain characteristics of the study (e.g., number or proportion of caseworkers involved or the exact nature of their training), such studies were included. This was done because it was thought that the chance rejection of an appropriate study could detract more from the generality of conclusions than the chance inclusion of an inappropriate study.

Eleven studies were located that met the minimum criteria for inclusion in this review: (1) services were provided by professional caseworkers for the experimental group and (2) an "untreated" or nonprofessionally treated control group was used. The criteria used to analyze these studies were derived from available texts on the evaluation of research.¹⁶ In general, the studies were analyzed along

¹² See, for example, E. Matilda Goldberg, *Helping the Aged* (London, England: Allen & Unwin, 1970); Rosa A. Marin, *A Comprehensive Program for Multi-Problem Families: Report on a Four-Year Controlled Experiment* (Rio Piedras: Institute of Caribbean Studies, University of Puerto Rico, 1969); and P. M. Kuhl, *The Family Center Project and Action Research on Socially Deprived Families* (Copenhagen: Danish National Institute of Social Research, 1969).

¹³ See, for example, John H. Behling, "An Experimental Study to Measure the Effectiveness of Casework Service" (Columbus, Ohio: Franklin County Welfare Department, 1961) (mimeographed); Geismar, op. cit.; Alvin Rudoff and Irving Piliavin, "An Aid to Needy Children Program: A Study of Types and Responses to Casework Services," *Community Mental Health Journal*, 5 (January 1969), pp. 20–28; and Edward E. Schwartz and William C. Sample, "First Findings from Midway," *Social Service Review*, 41 (June 1967), pp. 113–151.

¹⁴ See, for example, Margaret Blenkner, Julius Jahn, and Edna Wasser, *Serving the Aging: An Experiment in Social Work and Public Health Nursing* (New York: Community Service Society, 1964); and William J. Reid and Ann W. Shyne, *Brief and Extended Casework* (New York: Columbia University Press, 1969).

¹⁵ See, for example, Donald G. Langsley, Frank S. Pittman III, and Kalman Flomenhaft, "Family Crisis Therapy—Results and Implications," *Family Process*, 7 (September 1968), pp. 145–158.

¹⁶ See Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally Co., 1963); Herzog, op. cit.; Julian Meltzoff and Melvin Kornreich, *Research in Psychotherapy* (New York: Atherton Press, 1970); and Tony Tripodi, Philip Fellin, and Henry S. Meyer, *Assessment of Social Research* (Itasca, Ill.: F. E. Peacock Publishers, 1969). For a summary of the criteria used in these texts see Joel Fischer, "Framework for the Analysis of Outcome Research" (Honolulu: School of Social Work, University of Hawaii, 1971), (Mimeographed.)

the following dimensions: (1) formulation of the problem, (2) research design and method of data collection, (3) methods of data analysis, and (4) the authors' conclusions. Because of space limitations, detailed analysis of each study is not included here, except when problems in design either obscured potential findings or produced incomplete conclusions.

Except for a few situations in which methods traditionally defined as group work or community organization were used, the studies reviewed here “. . . addressed the practice of social casework . . . for the most part practiced ‘classically.’”¹⁷ Thus it could be assumed, and the studies themselves demonstrate, that each examines the practice of professional caseworkers, that there is indeed in all of the studies a central core of relevance to casework practice.

Since many readers may be unfamiliar with the results of these studies, the following sections present brief summaries, detailing the types of clients included, the nature and length of service, crucial aspects of the research method, and, of course, the findings. These summaries are so presented because such a review of the content of the studies is a necessary substantive basis for forming conclusions regarding the state of casework practice. For clarity of exposition, the studies are grouped into two categories according to whether they used one or the other of the two types of control groups already described.

Untreated Controls

Berleman and Steiner This study attempted to measure the impact of a service program on the prevention of juvenile delinquency.¹⁸ The researchers studied 167 black seventh-grade boys to determine past evidence of acting out and to predict future acting-out behavior. Four “high-risk” categories were formed from this group, and the boys were randomly assigned from these categories to experimental and control groups. Owing to attrition and other factors, the experimental group eventually consisted of twenty-one boys and the control group of twenty-six. Three trained social workers provided intensive individual and group services to the experimental group for five months. The dependent variable of juvenile delinquency was operationally defined as acting-out behavior and measured by school and police disciplinary records. Outcome was assessed between the preservice and service periods and at two postservice periods. No significant differences were found between the groups on the criterion measures of acting-out behavior at any of the service or postservice periods.

Craig and Furst This study was also designed to influence delinquency rates.¹⁹

¹⁷ Carol Meyer, “Implications of Evaluative Research Findings for Curriculum Concerned with Intervention on the Micro-System Level.” Paper prepared for the Symposium on the Effectiveness of Social Work Intervention, Fordham University, New York, New York, January 14–15, 1971.

¹⁸ William C. Berleman and Thomas W. Steiner, “The Execution and Evaluation of a Delinquency Prevention Program,” *Social Problems*, 14 (Spring 1967), pp. 413–423.

¹⁹ Maude M. Craig and Philip W. Furst, “What Happens After Treatment? A Study of Potentially Delinquent Boys,” *Social Service Review*, 39 (June 1965), pp. 165–171.

It included boys who rated high in predictions of probable delinquency (according to the Glueck Social Prediction Table, designed to predict future delinquency) as well as a small group of referrals from teachers. On the basis of matching, twenty-nine first-grade boys were assigned to an experimental group and twenty-nine to a control group. The boys in the treatment group were given intensive child guidance therapy by psychiatric social workers and other clinic professionals. The median length of clinic contact was fifty months. Delinquency records (presumably police and court records) were inspected over a ten-year interval and revealed the same number of delinquents in the experimental and control groups. In addition, school behavior reports, based on teacher evaluations, for nondelinquent boys were compared. These reports also revealed that the groups were not significantly different.²⁰

McCabe This study attempted to use social work intervention to diminish the deleterious effects of a “pathological environment” on intellectually superior children.²¹ From a larger group of predominantly black and Puerto Rican children in the second to fourth grades, who had demonstrated superior ability on IQ tests, sixty-seven children were matched and randomly assigned to treatment and control groups. Forty-two children were placed in treatment groups and twenty-five in control groups. Social workers conducted a program of intervention grounded in principles of ego psychology. They concentrated most of their efforts on small-group services to both the children and their parents.

Outcome was operationalized in terms of the children’s intellectual functioning, the parents’ functioning, and the family’s overall functioning and measured fifty-eight indicators of change. These measures included items from intelligence and school achievement tests, behavior rating scales, and scales of parental and family functioning. The researchers compiled an overall index of outcome that showed no significant differences between the experimental and control groups. In addition, of fifty-eight measures, only one statistically significant difference—reading achievement—favored the experimental group. The overall impact of this intensive service program, even if the one significant difference was not just a statistical artifact, was negligible.

Meyer, Borgatta, and Jones The purpose of this large-scale study was to examine “the extent to which social casework is effective in prevention” with potentially problematic subjects.²² The study subjects were four cohorts of high school girls, selected from the entire population of one school and identified on the basis of school records as “potential problem cases.” Eventually, by random assignment, 189 were referred to the experimental group and 192 to the untreated control group.

Services were provided by trained social workers from an agency specializing in the problems of adolescent girls. Both individual and group services were provided, although after the first year of the three-year program, group treatment was

²⁰ The statistics used were not specified in the report.

²¹ Alice McCabe, *The Pursuit of Promise* (New York: Community Service Society, 1967).

²² Meyer, Borgatta, and Jones, *op. cit.*, p. 3.

the primary mode of service. Three of the cohorts were included in analyses of all the data, while the last cohort, which had been exposed to treatment for two instead of the normal three years, was included only on selected measures.

Measures of outcome included a variety of subjective and objective criteria: school achievement and behavior ratings, personality and sociometric data, and client and worker ratings. Of the dozens of criteria by which experimental and control groups were compared, there were significant differences between the groups on only one of twelve factors of the Junior Personality Quiz. Although several other criteria tended to favor the experimental groups, no other between-group differences were statistically significant. To quote the authors: “. . . the conclusion must be stated in the negative when it is asked whether social work intervention with potential problem high school girls was . . . effective.”²³

Miller The goal of this study was to prevent adolescent delinquency—operationally defined as the amount of law-violating behavior—in a lower-class urban district.²⁴ As part of a large-scale “total community delinquency control project,” an experimental group of 205 gang members was matched with a control group of 172 gang members. Over a period of three years, the experimental group received both individualized and group services, with emphasis on group services. Although data on several outcome indicators were reported, the only clear comparison between experimental and control groups was on the number of court appearances. On this measure, there was no discernible difference between the groups. Reviewing the overall impact of the project, the author asked rhetorically: “Was there a significant measurable inhibition of law violating . . . behavior? The answer . . . is ‘No.’”²⁵

Powers and Witmer This was the first controlled study to examine the effects of casework intervention.²⁶ A well-designed delinquency prevention project, it matched and then randomly assigned 325 “predelinquent” boys to an experimental group and 325 to a control group. Direct individualized services were provided predominantly by caseworker-counselors. The mean length of contact per boy was four years and ten months.

Outcome was measured by court and police records, ratings of social adjustment, and psychological inventories. No significant difference was found between experimental and control groups on all major methods of evaluation. As frequently happens in the evaluation of services, the workers involved believed they had substantially helped a greater proportion of their clients than the more objective outcome measures revealed. This is an important indicator of the need for control groups and objective criterion measures.²⁷

²³ *Ibid.*, p. 180.

²⁴ Walter B. Miller, “The Impact of a Total Community Delinquency Control Project,” *Social Problems*, 9 (Fall 1962), pp. 168–191.

²⁵ *Ibid.*, p. 187.

²⁶ Edwin Powers and Helen Witmer, *An Experiment in the Prevention of Delinquency—The Cambridge—Somerville Youth Study* (New York: Columbia University Press, 1951).

²⁷ See, for example, Harvey Gochros, “The Caseworker—Adoptive Parent Relationship in Post-Placement Services,” *Child Welfare*, 46 (June 1967), pp. 317–326.

Of the six studies utilizing untreated control groups reviewed so far, all dealt primarily with children and adolescents, most in preventive rather than remedial terms. However, although most of the studies were conceptualized as prevention efforts, outcome indicators (e.g., personality measures, school achievement) are mainly the same as would be used in evaluating the effectiveness of remedial efforts. The overall outcome was clear: none of the studies revealed that their program had any significant effect on the clients when outcome measures for experimental and control groups were compared.

Other-treated Controls

Blenkner, Bloom, and Nielsen This study evaluated the effects of a program of services for the aged. A group of 164 aged persons were referred to community agencies for protective services because they had difficulty in caring for themselves. From this group 76 were randomly assigned to an experimental group and 88 to a control group. For one year the experimental group received intensive individualized services from experienced caseworkers; the goal was to do “whatever is necessary to meet the needs of the situation.”²⁸ The control group received ordinary community services from a variety of agencies. Outcome was operationalized in terms of four major aspects of the clients’ lives and situations: competence, environmental protection, affect, and effect on others.

Data were collected through structured interviews and ratings by observers. There were no significant differences between the experimental and control groups on most measures. Measures of “physical environment” and “concrete assistance” (both in the area of protection and not further delineated) and relief of stress on collaterals significantly favored the experimental group. However, most of the apparent gains in relation to these variables were explainable by a higher rate of institutionalization for experimental group subjects. In fact, overall findings from the initial part of the study led the project staff to consider the hypothesis that intensive service actually accelerates decline and to further examine follow-up data.

When data were examined at a five-year follow-up, there were significant differences between the experimental and control groups. That is, the experimental group members were found to have significantly higher rates of institutionalization and death than the control group members. Thus with survival being the ultimate outcome criterion, the effects of this intervention program favored the control, rather than the experimental group.

Brown Brown reported the findings of a program intended to evaluate the effectiveness of intervention with low-income multiproblem families.²⁹ Fifty multiproblem families receiving Aid to Families with Dependent Children (AFDC) were randomly assigned to an experimental group and fifty to a control group. The

²⁸ Margaret Blenkner, Martin Bloom, and Margaret Nielsen, “A Research and Demonstration Project of Protective Services,” *Social Casework*, 52 (October 1971), p. 489.

²⁹ Gordon E. Brown, ed., *The Multi-Problem Dilemma* (Metuchen, N.J.: Scarecrow Press, 1968).

experimental group received intensive family-centered services from professional caseworkers with reduced caseloads, while the control group received the usual services of the public assistance agency. The program lasted thirty-one months, and the dependent variable of family functioning was operationalized as movement on the Geismar Scale of Family Functioning and the Hunt-Kogan Movement Scale. There were no significant differences between the groups, which led the researchers to conclude as follows: “Whatever was done by these workers for these clients cannot be demonstrated to have had a beneficial effect. . . .”³⁰

Geismar and Krisberg This was another study dealing with the effect of reaching-out family-centered casework on low-income multiproblem families.³¹ The treatment group consisted of thirty of the most “seriously disorganized” families in one housing project. The control group was composed of fifty-one families from another housing project, all of whom were receiving AFDC and associated services. The control group differed from the treatment group on several variables. That is, it contained a far higher percentage of black families and families with absent fathers and demonstrated higher levels of family functioning at the pretest on the main criterion measure, the Geismar Scale of Family Functioning. In addition, the control and experimental groups lived in different geographic areas.

Services to the treatment group utilized various methods, primarily intensive direct services and use of environmental resources. Outcome was assessed on the Geismar scale twice for the control group and three times for the experimental group over the eighteen-month experimental period. At the conclusion of the project, the experimental group showed a gain of just under seven steps in mean “total family functioning,” while the control group gained less than one scale step. The authors concluded that this demonstrated a significant effect of treatment.

Unfortunately, the data do not support this conclusion. The initial differences previously noted between the experimental and control groups—several possibly crucial variables for which the two groups were not comparable—makes any conclusion of effectiveness or noneffectiveness potentially misleading. With neither matching nor the more preferable randomization of assignments to the experimental and treatment groups, and such obvious noncomparability, any gain for the experimental group can be explained as a “selection-maturation” artifact.³²

The treatment workers supplied information on the families’ social functioning for the experimental cases, and a different group of trained researchers supplied this information for the control group, which introduced an obvious and critical source of bias. And since the scores at pretest were more extreme in a negative direction for the experimental than for the control group, any positive change from pre- to post-test may be a product of statistical regression, independent of

³⁰ *Ibid.*, p. 127.

³¹ Ludwig Geismar and Jane Krisberg, *The Forgotten Neighborhood* (Metuchen, N.J.: Scarecrow Press, 1967).

³² Campbell and Stanley, *op. cit.* A selection-maturation artifact refers to an interaction that occurs when the selection of subjects for experimental and control groups results in groups with different potentials for rates of change.

the effects of the experimental variable.³³ In fact, the mean total family functioning score for the experimental families at the conclusion of treatment was still more than three steps below the pretest scores of the control group.³⁴ Thus the only conclusion that can be drawn from this study is that no definite conclusion about the effectiveness of the intervention program is possible.

Mullen, Chazin, and Feldstein This study utilized more satisfactory design procedures.³⁵ Eighty-eight new public assistance families were randomly assigned to an experimental group and sixty-eight to a control group. The experimental families received intensive professional casework services aimed at decreasing rates of family disorganization and enhancing family functioning. Control families received standard public assistance services. Eleven areas of family functioning, based on ratings of structured interviews, were used as criterion measures. At the conclusion of up to two years of service, no significant differences in family functioning were found between the experimental and control groups.

Webb and Riley The last study to be reviewed here was an attempt to affect the "life adjustment" of female probationers, aged 18 to 25.³⁶ Using random assignment, twenty-six recent probationers were assigned to an experimental group and thirty-two to a control group. The experimental group received intensive individualized services from family agency caseworkers for one year, and the control group received the usual probation services.

The dependent variable of life adjustment was operationalized as several dimensions of the Minnesota Multiphasic Inventory and a form of semantic differential. Subjects were also rated on sixteen "behavior correlates" by probation officers. The authors reported that the project was successful because six of twelve psychological measures showed significant improvement in the experimental group and only one of twelve showed significant improvement in the control group. In addition, five of sixteen behavior correlates "reflected markedly improved ratings of the experimental group as compared to the control group."³⁷

These conclusions cannot be sustained, however, because Webb and Riley, at least on the psychological dimensions, did not include between-group statistical measures. They only reported that the experimental group improved significantly on selected measures and that the control group did not. However, if the authors

³³ Statistical regression refers to a general tendency for those groups selected for treatment because of extremely negative scores to show evidence of improvement at a later point in time, irrespective of the treatment.

³⁴ Although a design using analysis of covariance techniques for equating experimental and control groups on pretest measures might have reduced uncertainty, such a design is hampered when the covariate is not perfectly reliable and when the samples are drawn from such obviously disparate populations. See F. M. Lord, "Large-Sample Covariance Analysis when the Control Variable is Fallible," *Journal of the American Statistical Association*, 55 (1960), pp. 437-451.

³⁵ Edward Mullen, Robert Chazin, and David Feldstein, *Preventing Chronic Dependency* (New York: Community Service Society, 1970).

³⁶ Allen P. Webb and Patrick Riley, "Effectiveness of Casework with Young Female Probationers," *Social Casework*, 51 (November 1970), pp. 113-115.

³⁷ *Ibid.*, p. 569.

had utilized a more appropriate statistical test—an analysis of covariance with pretest scores as the covariate (or even a t-test between the experimental and control group means if the pretest scores were equivalent)—the difference *between* groups, which is the crucial measure in evaluating overall impact of an experimental variable, may not have been significant. This is especially true in the several instances in which the differences between the groups were so slight. Again, the only conclusion that can be reached in this study is that the data were not presented in such a way as to justify a conclusion either of no effect or of significant effect.

The studies reviewed in this section contained a wider variety of clients and programs than those studies reviewed in the previous section. However, of the five studies, three clearly revealed little or no significant differences between the experimental and control groups and two provided inconclusive results.

Summary Analysis of Studies

Tables I.1 and I.2 provide a summary of all the studies reviewed. Six of the eleven studies dealt primarily with children as clients, three with low-income multiproblem families, one with the aging, and one with female probationers, aged 18–25. Most studies dealt with predominantly low-income subjects, although this was not uniformly the case. Both sexes and several ethnic groups were represented. Over two thousand separate cases, including a high percentage of families with multiple members, were involved. The group of studies reviewed here demonstrated a great diversity in criterion measures, ranging from subjective to objective measures that deal with several aspects of both personal and social functioning. Judgment, descriptive, and performance data were utilized and collected in a variety of ways, from psychological inventories and questionnaires, to worker and client ratings, to observed behaviors. While these measures individually could be faulty as indicators of change resulting from casework services, together they provide a wealth of information about the effects of casework services. More than one source of data was used to draw conclusions in almost all the studies. A wide variety of services was offered, although perhaps because many of the studies were conducted in the same time period, they reflect some uniformity in caseworker orientation, which is related to psychodynamic theoretical perspectives and/or “family-centered reaching-out” approaches.

Most of the studies provided at least minimally acceptable designs wherein experimental and control groups were assigned either through matching, randomization, or a combination of the two.³⁸ Frequently, however, the independent variable

³⁸ Randomization is an attempt to ensure that every potential subject has an equal chance of being assigned to an experimental or control group and is the preferable approach in attempting to avoid bias. Matching of subjects is often used when randomization is not possible. Since it would be desirable to have experimental and control groups demonstrate pretest equivalence on outcome measures (and perhaps other relevant life measures such as age) and also have an equal chance of being assigned to treatment or control groups, the optimum design would include both matching and randomization procedures. Randomization alone often produces equivalence, but this still must be examined separately by the researcher to determine whether equivalence between groups has in fact been attained.

was inadequately defined, so that the precise nature of the casework techniques used was unknown. This, however, may be less a fault of the research than, as noted earlier in this paper, of the theory and field that spawned it. There were no attempts to control for various traits and characteristics of the caseworkers (e.g., style, personality, techniques) and few attempts to examine differential characteristics of clients, especially in relation to differential responses to treatment.

Although these last flaws detract somewhat from the ability to analyze comprehensively all aspects of the results of these studies, they do not detract from the more general conclusions that can be drawn from this review. Of all the controlled studies of the effectiveness of casework that could be located, nine of eleven clearly showed that professional caseworkers were unable to bring about any positive, significant, measurable changes in their clients beyond those that would have occurred without the specific intervention program or that could have been induced by nonprofessionals dealing with similar clients, often in less-intensive service programs. In the two additional studies, the results were obfuscated by deficiencies in the design or the statistical analysis. Thus not only has professional casework failed to demonstrate that it is effective, but lack of effectiveness appears to be the rule rather than the exception across several categories of clients, problems, situations, and types of casework.

Deterioration of Clients

One of the most disturbing conclusions from the field of psychotherapy research is the finding that in a high proportion of psychotherapy studies, as many clients receiving professional services deteriorate as improve.³⁹ Averaged together and compared with a control group, the experimental group would therefore show no differences; thus the true effects of the experimental variable would be concealed. A reanalysis of the studies in this review shows a parallel phenomenon. In slightly under 50 percent of the studies, clients receiving services in the experimental group were shown either to deteriorate to a greater degree than clients in the control group or to demonstrate improved functioning at a lesser rate than control subjects.

For example, Berleman and Steiner, in examining the percentage of boys with school disciplinary records, concluded that there was no overall difference between the groups.⁴⁰ However, further analysis reveals that the percentage of boys in the experimental group with school disciplinary records was far higher (X^2 was significant beyond .01) than the percentage of boys in the control group. The study of Blenkner, Bloom, and Nielsen was already reviewed with regard to the deterioration of clients in the experimental group. That is, the experimental group subjects had a significantly higher death rate than those in the control group.⁴¹

The study by McCabe of educationally superior children revealed several areas

³⁹ Bergin, *op. cit.*

⁴⁰ Berleman and Steiner, *op. cit.*, p. 421.

⁴¹ *Op. cit.*

Table I.1 Summary of Studies Reviewed: Untreated Control Groups^a

| Author and Year | Clients | | | Caseworkers | | |
|-----------------------------------|--------------------|---|---------------------|--------------------------------------|--|------------------------------------|
| | Number | Characteristics | Method of Selection | Orientation | Major Approach | Setting for Services |
| Berleman and Steiner (1967) | E = 21 C = 26 | Black seventh-grade boys with school disciplinary problems and police records | Matching, random | Undetermined | Intensive, direct individualized, and group services | Settlement house, home, and school |
| Craig and Furst (1965) | E = 29 C = 29 | First-grade boys rated as "probable delinquents" on Glueck Social Prediction Scale | Matching | Undetermined, possibly psychodynamic | Intensive child guidance therapy | Child guidance clinic |
| McCabe (1967) | E = 42 C = 23 | Mainly "intellectually superior, socially disadvantaged" black and Puerto Rican children | Matching, random | Ego psychology | Groups, some individual services | Office |
| Meyer, Borgatta, and Jones (1965) | E = 189 C = 192 | High school girls, varied races and socioeconomic statuses, identified as "potential problems" | Random | Ego psychology, diagnostic casework | Group services, individualized services | Office |
| Miller (1962) | E = 205 C = 172 | Lower-class gang members, varied ethnic backgrounds, both sexes | Matching | Psychodynamic, group dynamics | Group and individualized services | Streets, homes, schools |
| Powers and Witmer (1951) | E = 325 C = 325 | Pre-delinquent boys aged 10–17, screened through teacher reports and test data. A variety of socio-economic classes and ethnic groups | Matching, random | Dynamic psychology | Direct individualized services | Homes, school, office |

in which experimental group members declined at a higher rate than control group members or in which control group members improved at a higher rate than experimental group members.⁴² On the overall index of functioning, 50 percent of the experimental group members declined, compared to only 38 percent of the control group members. The greatest decline was found in the black clients: eight out of fourteen in the experimental group deteriorated—presumably as a result of

⁴² Op. cit.

Table I.1 Summary of Studies Reviewed: Untreated Control Groups^a (continued)

| Length and Amount of Contact | Assessment Procedure | | Outcome |
|--|--|--|--|
| | Dependent Variable | Criterion Measures | |
| L = 5 months A = median—75 hours per client | Acting-out behavior | School disciplinary records, police records | No significant difference between E and C groups |
| L = 5 years (median 50 months) A = Unknown | Delinquency rates | Teacher's behavior reports, delinquency records | No significant difference between E and C groups |
| L = 3 years overall A = 90.5 meetings | Intellectual functioning of children, parental functioning, family functioning | Intelligence tests, school achievement, behavior rating scales, ego functioning scales, ratings of parental and family functioning | No significant difference between E and C groups |
| L = 1 contact to 3 years A = median—17 contacts | School behavior, social functioning | Client and worker ratings, school grades, school-related behaviors, teacher ratings, personality and attitude inventories | No significant difference between E and C groups |
| L = 3 years A = 3.5 contacts per week | Law-violating behavior (delinquency) | Number of court appearances | No significant difference between E and C groups |
| L = 8 years (mean of 4 years, 10 months per boy) A = 27.3 contacts per year | Frequency and seriousness of delinquency, social adjustment | Court records, police statistics, ratings of seriousness of offenses, ratings of social adjustment, psychological inventories | No significant difference between E and C groups |

^aIn this table "L" stands for length, "A" stands for amount of contact, "E" stands for experimental group, and "C" stands for control group.

treatment—whereas only one black control group member did so. The outcome pattern was reversed for Puerto Rican clients. Thus the overall effect was that the black and Puerto Rican clients canceled each other out so that no significant differences could be observed. McCabe further reported that means on both ego and family functioning indicators for black subjects tended to increase (indicating more positive outcome) to a greater extent for control group members than for experimental group members. This suggests treatment may have retarded normal improvement.

The delinquency control project by Miller also showed evidence of this phenomenon. In several areas related to trends in disapproved behavior and in illegal acts, the experimental group showed statistically significant increases rather than the hypothesized desired decreases.⁴³ However, since no figures were reported for

⁴³ Op. cit., pp. 180–183.

Table I.2 Summary of Studies Reviewed: Other-Treated Control Groups^a

| Author and Year | Clients | | Method of Selection | Caseworkers | | |
|--------------------------------------|------------------|--|-----------------------------------|--|--|-------------------------------------|
| | Number | Characteristics | | Orientation | Major Approach | Setting for Services |
| Blenkner, Bloom, and Nielsen (1971) | E = 76 C = 88 | Mentally impaired aged in need of protective services; noninstitutionalized | Random | Undetermined, probably psychodynamic, "social therapy" | Intensive direct services, use of environmental resources | Office and home |
| Brown (1968) | E = 50 C = 50 | Multiproblem families receiving AFDC | Random | Multiproblem, family centered | Intensive direct services, use of environmental resources | Office and home |
| Geismar and Krisberg (1967) | E = 30 C = 51 | Low-income multiproblem families, predominantly white | Unclear, mainly post-hoc matching | Reaching-out, family centered | Intensive direct services, use of environmental resources, multimethod | Office, home, neighborhood |
| Mullen, Chazin, and Feldstein (1970) | E = 88 C = 68 | Newly dependent public assistance recipients, mixed ethnic group, families with at least 2 members | Random | Psychodynamic | Direct individualized services | Undetermined, probably office, home |
| Webb and Riley (1970) | E = 26 C = 32 | Female probationers aged 18–25; variety of ethnic groups | Random | Psychodynamic | Direct individualized services | Office |

the control group, there is no way of knowing whether such deterioration was an effect of treatment or of other circumstances.

The Powers and Witmer study showed that although some of the clients in the delinquency program seemed to benefit from treatment, a substantial proportion actually were handicapped by it. The authors concluded that "the apparent chance distribution of terminal adjustment ratings . . . was due to the fact that the good effects of the study were counterbalanced by the poor."⁴⁴

Geismar and Krisberg revealed that 10 percent of the experimental group members in their study deteriorated in social functioning over the course of the project. A comparable breakdown was not available to examine such possible decline in the control group.⁴⁵

In three of the five studies (Berleman and Steiner, McCabe, and Powers and Witmer) control procedures made it appear likely that decline in the experimental

⁴⁴ Op. cit., p. 455.

⁴⁵ Op. cit.

Table I.2 Summary of Studies Reviewed: Other-Treated Control Groups^a (continued)

| Caseworkers Control Group Workers | Length and Amount of Contact | Assessment Procedure | | Outcome |
|--|--|--|--|--|
| | | Dependent Variable | Criterion Measures | |
| Variety of community workers, generally not social workers or not MSWs | L = 1 year A = mean of 31.8 per case | Competence, environmental protection, affect, effect on others | Ratings from structured interviews, observer ratings, clinical ratings, death and institutionalization rates | Experimental group had higher death and institutionalization rates. Also higher on "physical environment," "concrete assistance," and relief of collateral stress |
| Public assistance workers—BAs | L = 31 months A = median of 2+ per month | Family functioning | Geismar Scale of Family Functioning, Hunt-Kogan Movement Scale | No significant difference between E and C groups |
| Public assistance workers—BAs | L = 18 months A = mean of 4.4 direct contacts per month | Family functioning | Geismar Scale of Family Functioning | Major movement within E group. Major differences between E and C groups at pretest not handled statistically |
| Public assistance workers—BAs | L = up to 2 years A = median of 15 direct interviews | Individual and family disorganization, family functioning | Ratings of structured interviews with clients in 11 areas of family functioning | No significant differences in family functioning between E and C groups |
| Non-MSW probation workers | L = 1 year A = median of 6 to 9 interviews | Life adjustment | MMPI, Semantic Differential, behavior ratings | No between-group measures reported. Reported "improved" scores on 5 of 16 behavior ratings favoring E group and on 5 of 12 psychological measures favoring E group |

^aIn this table "L" stands for length, "A" stands for amount of contact, "E" stands for experimental group, and "C" stands for control group.

group was actually a result of the treatment, while in two studies (Miller and Geismar and Krisberg) there is evidence to suggest that such deterioration took place. It was not always clear that the deteriorated group was sufficient in number to offset statistically the number of clients who may have improved and thereby produce a finding of no significant difference between experimental and control groups. However, even the evidence presented here is strong enough to suggest that, as with psychotherapy, the results of casework may be for better or for worse!⁴⁶ At the least, future research should attempt to specify the influence, whether personal

⁴⁶ Charles Truax and Robert Carkhuff, *Toward Effective Counseling and Psychotherapy* (Chicago: Aldine Publishing Co., 1967).

(e.g., personality characteristics of caseworkers) or situational, that might account for this variation in effects.

Conclusion

This article has been concerned primarily with a presentation of research findings related to practice, rather than with an analysis of practice per se. But the disturbing nature of the results of these studies does suggest some areas for further questioning.

It is possible that the type of casework used in these studies really was not representative of the mainstream of casework practice. There appeared to be, for example, a disproportionate number of studies concentrating on work with children, especially with juvenile delinquents. Also since most of the studies dealt with low-income clients and few with middle-class clients, it might be argued that the high rate of failure was merely an artifact of the general inability of caseworkers to help clients when other more powerful environmental forces hold sway. And, although the nature of the problem is important, the methods used may reflect outdated forms of casework.

Most of the studies were conducted in the 1950s and 1960s and reflect the dominant modes of practice in those decades. Compared at least to the services offered in the earlier studies, the nature of casework practice has undergone many alterations, although there is as yet no controlled evidence that newer schools of casework have been able to demonstrate success in helping clients.⁴⁷

But caseworkers do have to act, even in the face of such discouraging evidence, since practice can never be painted in terms of absolute success or failure. Making judgments in the face of uncertainty of knowledge has long been a characteristic of most of the helping professions. A variety of emerging approaches to practice are available as the search for more effective methods of intervention continues.⁴⁸ Perhaps future research will be able not only to validate new methodologies, but, as in the studies in which client deterioration was evident, more clearly define the elements of the old forms that enhance effectiveness.

Nevertheless, this review of the available controlled research strongly suggests that at present lack of evidence of the effectiveness of professional casework is the rule rather than the exception. A technical research corollary to this conclusion, and a comment frequently appearing in the social work literature, is that "we also lack good scientific proof of ineffectiveness."⁴⁹ This assertion, however, taken alone, would appear to be rather insubstantial grounds on which to support a profession.

⁴⁷ Roberts and Nee, op. cit.

⁴⁸ See Joel Fischer, *Interpersonal Helping: Emerging Approaches to Social Work Practice* (Springfield, Ill.: Charles C Thomas, 1972).

⁴⁹ Mary E. MacDonald, "Reunion at Vocational High: An Analysis of 'Girl's at Vocational High'" *Social Service Review*, 40 (June 1966), p. 188. See also Cooper and Krantzler, op. cit.