Strategy in Evaluating the Effectiveness of Community Mental Health Programs

DAVID J. VAIL, M.D.

THE PROBLEMS of evaluating the effectiveness of community mental health programs were the subject of a recent study in Minnesota. The study was based on an analysis of statistical reports received by the Minnesota Department of Public Welfare for the fiscal year 1958–59. These reports had been submitted to the central agency in accordance with standard reporting procedures set forth by the Public Health Service. The material was gathered from outpatient clinics, or community mental health centers, in the Minnesota mental health system.

In this paper the terms "clinics" and "centers" or "mental health centers" are used interchangeably. In actuality, the term "mental health centers" is most appropriate, since the Minnesota program is designed by law to include educational, consultative, rehabilitative, and preventive services in addition to the basic clinical ones.

Method

During the fiscal year 1958–59 eight clinics or centers were reporting comparable data. Each center was assigned as a code name the name of an Ivy League college. This was done in order to maintain confidence and to establish an atmosphere of objectivity.

The number of full-time professional personnel (FTPP) was calculated for each operation. This was easily done by adding, for each type of personnel, the products of number of personnel, amount of time, and number of months spent, and dividing the sum by 12. For example:

a. Psychiatrist, half time for 6 months, full time for 6 months
b. Social worker, full time for 12 months
c. Psychologist, 0.4 time for 5 months

Calculation:

a. 1×½ × 6= 3
   1× 1 × 6= 6
b. 1× 1 ×12=12
c. 1× 0.4× 5= 2
   —
   23

23÷12=1.92 FTPP

Psychometrists and group workers were not counted because only two centers had them. Therefore, the compilations do not include reports of psychometric testing, group therapy, or group work.

The number of unduplicated patients treated per year was recorded directly from reports, as was the number of interviews held for any purpose. The number of man-hours spent in activities outside the clinic setting itself, such as in speaking engagements, consultations, and teaching sessions, was compiled from the reports by a series of tedious but basically simple arithmetical calculations.

So that some kind of total calculation of time spent in both intraclinic and extraclinic activities could be made, interviews were assigned values of 1 hour, 40 minutes, 30 minutes, and 20 minutes each and the total hour-value arrived at was added to the extraclinic time spent in each activity.

Once these figures were determined, a further series of calculations was performed in order to ascertain the amount of product, whether it be number of interviews, number of patients,

Dr. Vail is director, division of medical services, Minnesota Department of Public Welfare, St. Paul.

or total time, per full-time professional person, that is, per unit of producer.

Results

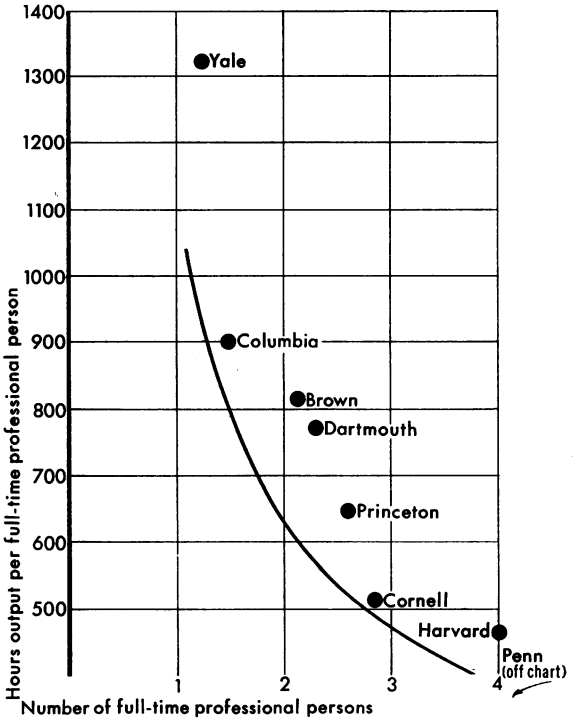
One of the main reasons for measuring quantity of output of mental health centers is the frustration inherent in any attempt to measure value. It is a great deal simpler to measure quantity. But it is impossible to avoid value questions altogether. In an industrial operation there is an easily measurable end product, let us say the number of cars coming off the assembly line. What is our end product? The answer would seem to be the number of patients successfully treated. But how do we define "success"? Or "treated," for that matter? For such reasons as this we elected to study a number of different parameters, all of them centering basically around the amount of work done. The total number of interviews for any purpose was used, rather than interviews with patients only.

Since, by deliberate design of the Minnesota community mental health program, centers devote from 25 to 40 percent of staff time to extraclinic pursuits, we had to find some way of bringing such activities into the context of work done; hence the procedure of totaling hours spent inside and hours spent outside the clinic. The difficulty of assigning a uniform and average time value per interview is too obvious to dwell on. The multiple calculations help to meet this problem, but there is no apparent way within the limits of human energy to resolve the possibility, indeed the probability, that average interview time differs from one clinic to the next. For example, in one clinic, Yale, interview time may be significantly briefer than in the other clinics.

It was a very short, logical step from calculation of average interview time to consideration of quantity of work, by whatever gauge measured, per unit of producer. This is cumbersome phrasing, but it is well to avoid at the outset a value word like "efficiency." We cannot speak of efficiency until we can measure the amount of desirable end product, that is, successful treatments or cures.

When work per unit of producer is plotted against the number of producers, the resulting

Total hours of interview time per full-time professional person, each interview calculated as 40 minutes



curve is on a typical hyperbola. The chart shows this. This hyperbola is obtained, although with differing rank order arrangements, regardless of the gauge used. Thus number of patients per FTPP, number of interviews per FTPP, extraclinic hours per FTPP, and total time, by whatever method computed, per FTPP, when plotted against the number of producers, all produce the same curve. The total time, shown on the chart, with interviews calculated at 40 minutes each, is typical.

What this means in plain language is that, as the size of the staff increases, the amount of actual, recordable work done by each individual appears to decrease. This is in accord with Parkinson's law (*1*) and should come as no surprise to anyone who has been committeed to death in organizations. The finding is also consistent with information on input-output relationships in the general behavior systems theory (*2*).

A word about rank order. Rank order for a variety of scales is shown in the table. Yale is persistently at the top and Penn, persistently

at the bottom. For Yale this may be explained by unusual productivity or by some difference in average of duration in interviews; for Penn, by unusual nonproductivity. Since different rank order relationships are produced by the use of different scales or combinations of scales, it would appear that, with the limited material available, it may not be possible to make valid or meaningful statistical correlations.

The Quality-Quantity Problem

Objections to the foregoing methods are obvious. Even assuming that recording of data is accurate, that all patients and interviews and, more particularly, involvement in extraclinic activities are being properly reported, what do the results mean? The fact that one group puts in more time on the job, sees more patients, gives more consultations, or makes more speeches than the other groups has no necessary relationship with the quality of its work or with its significance in or impact on the community.

Residue time is presumably taken up with conferences, and these are not recorded or reported. We surmise that if the staff spends time in considering and discussing cases there can be better mutual understanding of the therapeutic tactics and strategy to be pursued in each case and that performance will be improved. Unfortunately, there is no way of proving this.

The clinic program director faces a serious dilemma. He wishes to increase the overall output of his team by adding new members. And yet he knows that, if he does so, in terms of work per unit his team will become less productive. A great deal depends on how sharply the curve for amount of work accomplished drops. The rate at which it does drop or, more accurately, the way in which the drop is made less sharp would appear to be the test of effective administration. "Administration" is used in a broad sense to include the proper selection of staff, assignment of cases, and maintaining of morale. Hypothetically, it would seem that no drop at all is necessary. On the other hand, there is a quality of pessimism inherent in the problem, and we are content to assume that the curve will inevitably be a falling one rather than a straight horizontal line. Analogously, in baseball there is no reason in theory why a batter cannot hit safely each time he comes to bat; empirically, we know that it is not humanly possible for him to hit safely more than 4 times out of 10 for any protracted period.

Comment

Increasing the size of the mental health center staff does something more than add to the total output of the center or subtract from the output per producer. It makes available a greater variety of personalities for administering therapy. This is on condition that the personalities brought into the team are significantly different from one another. We refer here to what might be called the versatility factor. If personalities of new members are identical to those of current members, no

Rank order of eight mental health centers, Minnesota, fiscal year 1958–59

Rank order	Interviews [1]	Number patients [1]	Extraclinic hours [1]	Length and number of interviews [1]				Number full-time professional persons [2]
				1 hour	40 minutes	30 minutes	20 minutes	
1	Yale	Brown	Columbia	Yale	Yale	Yale	Yale	Columbia.
2	Brown	Yale	Yale	Brown	Columbia	Columbia	Columbia	Brown.
3	Dartmouth	Harvard	Dartmouth	Columbia	Brown	Brown	Dartmouth	Yale.
4	Cornell	Princeton	Princeton	Dartmouth	Dartmouth	Dartmouth	Brown	Dartmouth.
5	Princeton	Columbia	Harvard	Cornell	Princeton	Cornell	Princeton	Princeton.
6	Columbia	Dartmouth	Brown	Princeton	Cornell	Princeton	Cornell	Cornell.
7	Harvard	Cornell	Cornell	Harvard	Harvard	Harvard	Harvard	Harvard.
8	Penn	Penn	Penn	Penn	Penn	Penn	Penn	Penn.

[1] Per full-time professional person.
[2] In ascending order. Actual numbers are: Columbia, 1.42; Brown, 2.1; Yale, 2.2; Dartmouth, 2.25; Princeton, 2.57; Cornell, 2.9; Harvard, 4; and Pennsylvania, 4.1.

progress along these lines is possible. This is a hazard of the classical small team, whose members must get on well together, that is, they must like each other, which is another way of saying they must be like each other. To the extent that they are like each other, versatility is diminished. That this is no insignificant factor is seen when we realize that the number of possible relationships within an organization increases both factorially and exponentially with the number of members and the number of attitudes. The actual formula appears to be

$$N=(p!)^q$$

where

N=number of relationships

p=number of members

q=number of attitudes

"Organization" in this sense must be extended to include not only the team members but the patient and his family as well. One advantage of group and family therapy is that in the deliberate utilization of the personalities of the group members and their capacities for relationship with each other, one maximizes to a fantastic degree the versatility factor. In this same respect proper selection of staff and optimal assignment of cases are viewed as critical functions of the program director.

A final word about evaluation of results. Let us define therapy in the first instance as a process of change through relationship which we must assess. A change from what to what? Postulations as to change in basic personality or character structure, if indeed this is ever really accomplished, appear to be meaningless and to resolve finally to questions of value. If I am beat and you are square, there is no way of reconciling our contempt for one another's way of life. If I am neurotic and you are stupid, the same problem exists. When patient and therapist differ and one changes to become like the other, if the victor happens to be the therapist he can chalk up a successful case to his credit. But we have not really solved anything.

For this reason, the most satisfactory measure of results may, in the final analysis, be a symptomatic and experiential one: work with conditional responses and physiological periodicity and refinements in concepts of phenomenology show some promise in this connection. Certainly, for statistical purposes we have no measure which we can use at this time.

There is an interesting story, possibly apocryphal, of General Grant at the Battle of the Wilderness. At the close of the fighting one of his aides came to him with the report that the battle was over and the Union forces appeared to be victorious. Grant is said to have asked one question: "How many prisoners did we take?" For him at this time and under these circumstances, this was evidently the datum which would give him the most significant information. We are faced with the sad fact that we have no such yardstick. Until we do have one, the question of evaluation must be held in abeyance.

REFERENCES

(1) Parkinson, C. N.: Parkinson's law. Houghton-Mifflin Co., Boston, 1957.

(2) Miller, J. G.: Information input overload and psychopathology. Am. J. Psychiat. 116: 695–704 (1960).

Correction

In "Public Health and Medical Aspects of the Roseburg, Oreg., Disaster," published in the August 1961 issue, p. 727, the force of the explosion was erroneously given at 26 kilotons of TNT bombs. The correct force is 13 tons.