Community Mental Health Journal, Vol. 20 No. 1, Spring 1984

Statewide Mental Health Outcome Evaluation: A Perspective of Two Southern States

Harold L. McPheeters

ABSTRACT: The states of Georgia and Tennessee implemented statewide outcome assessment programs in a number of their community mental health programs. Georgia used a Role Functioning Scale which was developed for this purpose. The system was implemented statewide and tied to quality assurance. Tennessee used the Quality of Life Questionnaire and the Program Impact Monitoring System developed in Oregon and implemented the system on a volunteer basis in six community programs. Both systems had some successes and encountered some difficulties which are described. Fiscal constraints have curtailed both systems, but suggestions are made for statewide outcome assessment projects based on these experiences.

BACKGROUND

Outcome measurement as part of evaluation is new to most mental health programs. Until recently, the focus of program evaluation activities in mental health has been on the numbers of clients served, the numbers of services delivered, and the costs, but not on the outcomes. Most efforts at the measurement of outcomes have taken place in relation to research studies, special demonstration projects, or the interests of highly motivated program administrators in local mental health agencies. This is partly because of disagreements about what should be the outcomes of various kinds of mental health programs and partly because there have been few incentives or rewards based on having outcome data. Most of the payment and reward systems have been based on the number of services delivered rather than on the results obtained from the services (McIntyre, Attkisson and Keller, 1974).

However, in the past four years, several state mental health agencies have piloted the use of some kind of statewide system of outcome measurement. The purpose of this article is to describe and compare the experiences of two states in the development and use of outcome measures. The two states are Georgia and Tennessee, both of which have been in the vanguard of states that have been concerned about outcome measurement and that have un-

Copies of the Role Functioning Scale and the Quality of Life Questionnaire may be obtained from the author. Address: Southern Regional Education Board, 1340 Spring Street, N.W., Atlanta, GA 30309.

dertaken statewide programs. Georgia and Tennessee are two of the states in the 14-state region served by the Southern Regional Education Board and have been active participants in regional activities to improve the program evaluation capability of mental health programs in the South. Both states have strong state-level mental health agencies with the evaluation and data management capability to undertake such efforts. In both states, the central office has a close relationship with the local community mental health centers. However, the relationship in Georgia is officially closer, because the local mental health programs are administered through the local health departments, which are sister agencies of the Division of Mental Health in the State Department of Human Resources. In Tennessee, the local agencies are private non-profit organizations in their local communities. This puts the Tennessee Department of Mental Health and Mental Retardation in something of a consultant/technical assistant role, although in both states there is a strong expectation that the local agencies will participate in program evaluation activities of the state office because of the need to be accountable for the state funds that come to the community programs.

Both states have had well-established management information systems that collect and analyze information about the services and programs of the community agencies, and both states had come to the point of perceiving the need for some uniform system to measure outcomes. In both states some local mental health programs had established local outcome assessment systems for at least some of their programs. In many places, these assessments were based on goal attainment scales (Kiresuk and Sherman, 1968) or various community adjustment scales. In a number of places the Global Assessment Scale (Spitzer, Gibbon and Endicott, 1975) was used. In most local agencies these were systems that involved ratings by staff clinicians, although some used family and/or client ratings—especially if the scales were client satisfaction scales. However, in all cases the data were gathered, analyzed, and used locally. While the central offices received courtesy copies of some of the reports of local outcome studies, there was no way the data could be compiled and used at the state level, since the data collection systems were developed locally and were idiosyncratic to the local needs. There was no uniformity that would allow for statewide analysis.

The variability of methods of assessing outcomes has been a major obstacle to the widespread use of outcome measurement. In addition to the fact that mental health professionals are often reluctant to be held responsible for client outcomes, which they feel are influenced by many factors other than what they do in therapy, the various professionals have very different notions of what they perceive to be the outcomes of their services. Psychiatrists tend to favor mental status; psychologists focus on interpersonal functioning; social workers opt for social functioning, and vocational counselors insist on vocational adjustment. Others think in terms of resolution of problems, reduction of symptoms, attainment of goals, or client satisfaction. Which is to

be used in a uniform system of outcome assessment? In addition, there are disagreements about who is to do the ratings—clinicians, family members, clients, or independent raters. There are also questions about whether the assessments should be done in person, by telephone, or by mail (Hargreaves, McIntyre and Attkisson, 1975).

Before any statewide uniform system of outcome assessment could be implemented, there had to be some agreement about which of these approaches would be most appropriate and useful for the kinds of decisions that must be made by state and major program administrators. In most state mental health agencies, including Georgia and Tennessee, there is a need to be primarily concerned with the social functioning of clients, especially those with major or chronic mental illness. It is difficulties in social functioning that bring most such patients to the public services and it is social performance that determines whether the clients are able to remain in the community or must return to the mental health centers or state hospitals for further public services. Thus, both states felt the need to choose an outcome assessment system that included social and interpersonal behavioral measures as specific dimensions. However, within these limitations there were still many choices of methods and instruments (Hargreaves, McIntyre and Attkisson, 1975).

THE GEORGIA MODEL

The outcome assessment program in Georgia grew from a desire of the state mental health agency's leaders to have a system that would not only allow the managers to assess the outcomes, but which would also help to direct the system. At that time, the Georgia mental health program had adopted the Balanced Service System (Gerhard and Dorgan, 1977) for managing its community mental health programs and its mental hospitals. The Balanced Service System was a conceptual scheme for organizing the activities of community mental health programs. It divided the services into eight major functional categories:

- Identification
- Crisis stabilization
- Growth
- Sustenance
- Case Management
- Prevention
- General Health
- Ancillary

Each of these categories of services was related to the functioning of clients (or of the agency). This conceptual system seemed especially relevant to the public services where problems in social functioning are the major reasons why clients come for services and are among the major concerns that legislators and the general public feel should be addressed by public mental health programs.

The leaders in Georgia felt that a measurement system based largely on social functioning could be useful both at the time of admission in providing a baseline of social functioning for each client and a series of levels of social functioning that might be set by the therapist as the goals for the client to attain. The system could be used as a quality assurance system to assure that payments were being made for appropriate services. Then an assessment at the time of release from service might provide the measure of outcome that had resulted from the services.

The instrument developed was the *Role Functioning Scale* for adults and independent adolescents. The inputs to the system came from workers at all levels in the state mental health program. This Scale was made up of four subscales, each of which had seven levels of functioning—beginning with limited functioning at level 1 and progressing to optimal functioning at level 7—that the clinician was supposed to rate. The four subscales were:

Working; Productivity
(in the client's most expected role, i.e., homemaker, student, wage earner)
Independent Living; Self-Care
(managing of household, eating, sleeping, hygiene care)
Immediate Social Network Relationships
(close friends, spouse, family)
Extended Social Network Relationships
(neighborhood, community, church, clubs, agencies, recreational activities)

Recognizing that clients sometimes came for services because of personal distress which did not impair their social functioning in any significant way, there was also a *Global Personal Distress Scale,* which also had seven levels of subjective, self-reported feelings which might be situational or symptomatic, or a combination.

The *Global Role Functioning Index* was made up of the tools from the four Role Functioning Scales, because the agency felt that the major thrust of public services should be to improve the social functioning of clients, rather than to help clients with subjective and personal distress which did not affect their social functioning. The Global Index was calculated by adding the scores of the numbers of the levels that had been assigned by the evaluator on each of the four Role Functioning Subscales. Thus, the Global Role Functioning Scale had the following scores and descriptors:

4	Severely limited
5–8	Markedly limited
9–12	Limited
13–16	Marginal

17-20 Moderate
21-24 Adequate
25-28 Optimal

The raters were the clinicians assigned to the clients, and the clients were rated at the time of admission to the service and at the time of release, or yearly if the clients remained in treatment that long. The scale was given a limited amount of validation study before it was implemented throughout the system. The implementation was accomplished after orientation sessions had been held to acquaint clinicians with the Scale and its use. The period between the initial announcement of the system and its full implementation was a brief three months.

As the system evolved and was ready for implementation, it was seen as a means to provide quality assurance as required by federal legislation and also as a way of providing assurances to the Medicaid program that the services were appropriate to the needs of the clients. The leaders of the Division of Mental Health negotiated arrangements with the Medicaid administrators whereby the ratings on the Role Functioning Scale would be accepted by Medicaid as justification for various kinds of services. It was assumed that each clinician would see the scale as a means of helping set service goals and assessing how well the client was moving toward those goals. The scores were tabulated and analyzed at the local program level; so they were readily available to clinicians.

Early in the implementation phases, it became apparent that there were problems in using the Role Functioning Scale with many alcohol and drug abuse clients and certain other clients. The state office undertook a program of technical assistance to the local programs, but soon found itself with too limited resources to provide all the assistance that was needed.

Also, it soon became apparent that there was serious resistance among clinicians at all levels in the agency. It is difficult to determine just what created the most resistance, but there were three major elements:

- Despite the fact that the system had been developed with the inputs of clinicians from all levels of the agency, there was a feeling that the system had been imposed upon them by Central Office and that it had been done too fast and without sufficient preparation. There were feelings that the system should have been implemented on a trial basis in selected programs in order to better ascertain its validity and reliability and that more gradual implementation would have allowed for working out some of the "bugs" in the system.
- There was widespread resistance to some of the concepts, and particularly to the language of the Balanced Service System. Many clinicians were not comfortable with the focus on social functioning of the Balanced Service System, but in addition the concept of the Balanced Ser-

vice System used a sociological "language" and vocabulary that was new and strange to most mental health clinicians.

- There was resistance to the idea of having the reimbursement system of Medicaid and the quality assurance system tied to the Role Functioning Scale. This resistance became marked when clinicians discovered that they could not be reimbursed for certain kinds of services because the services were deemed to be inappropriate for the level of functioning which clinicians had assigned to the clients.

At about the time the Role Functioning Scale was implemented, there was turnover in the top level of the leadership in the Division, which resulted in an altered commitment to the Balanced Service System and to the Role Functioning Scale. The new administration was not actively opposed to either concept, but the administration was consumed with other concerns, so that there was little time left to pursue the further implementation and refinement of the Scale.

The Role Functioning Scale continues to be used and is still related to the quality assurance program. However, it has received little further attention from the Central Office, which is trying to cope with fiscal cutbacks in federal and state funds. There have been no wide-scale efforts to validate the scale in practice, although individual mental health centers have done such studies as well as studies to test the inter-rater reliability of the instrument. In the local programs, there is some staff grumbling about the paperwork that it requires, and there is a degree of cynicism about its use. Some of this cynicism results from the fact that the levels of functioning are felt to be too few and too general to be of any real clinical utility. There is also concern about its validity and reliability.

The Central Office has made no substantial effort to use the findings for program evaluation or for program planning. There is a feeling among the agency's leaders that the system still has real potential for being used in assessment of the outcome of programs, but that further work would have to be done to refine the system and determine its validity and reliability. There is the feeling that some of the local programs make good use of the scale in their local operations and that they have good training programs and supervision of staff in the use of it, but that this is not universally true. There is a belief that to attempt to make the system useful at this time would require a considerable effort to back up and overcome some of the resistances that are still remembered.

THE TENNESSEE MODEL

In Tennessee, the motivation for a statewide outcome assessment system grew from the interest of the Assistant Commissioner for Planning and Evaluation

and her evaluation staff. The state was moving strongly to deinstitutionalize the chronically mentally ill from its mental hospitals and needed a way to assess how these clients were functioning in the communities. The state was also interested in other areas of outcome measurement, but this was its primary motivation. This area of the outcomes for the chronically mentally ill calls particularly for measures of social functioning, since social functioning is a major area of deficiency and an area that brings the most public criticism if it is not satisfactorily managed.

While searching for a system of outcome measurement that would meet this need, the Assistant Commissioner for Planning and Evaluation learned of the Program Impact Monitoring System that had been developed and piloted by Dr. Gerry Brodsky and Douglas Bigelow in the Oregon Division of Mental Health (Brodsky and Bigelow, 1980). She was impressed with it and sent a staff person to visit Oregon to learn more about it and to see it in operation. She then discussed what she had learned with the director of the Tennessee Department of Mental Health and Mental Retardation and other key staff and arrived at the consensus that this was the program they should replicate in Tennessee. Unfortunately, that was also a time of revenue shortfalls and cutbacks in the state of Tennessee, so it appeared that such a system would have to wait.

About this time the National Institute of Mental Health issued a request for a proposal to replicate the Oregon Program Impact Monitoring System in one of the state mental health agencies of the South and to convene a Learning Community of representatives from each of the other states of the South to observe and learn about the progress and problems in implementing such a system. The Mental Health Program of the Southern Regional Education Board became the administrator of the contract, and issued invitations to each of the 14 Southern state mental health agencies to become the state in which to replicate the Oregon system. Tennessee gave an early response and was chosen to be the replication state.

The Oregon Program Impact Monitoring System was based on a quality-of-life concept that held that an individual's quality of life results from the interaction between an individual's needs and abilities and his/her environment's opportunities and performance requirements. This results in both levels of satisfaction and levels of performance in a variety of domains of living. This quality-of-life concept can be applied to any citizen, not only those with disabilities.

From this concept, Brodsky and Bigelow developed the Quality of Life Questionnaire (QLQ), which had four major scales:

Personal Adjustment
(psychological adjustment—both distress and well-being, tolerance of anxiety, need satisfaction, independence)

Interpersonal Adjustment
(friend role, spouse role, parent role, social support)
Adjustment to Productivity
(work at home, employability, work on the job, school, use of time)
Civic Adjustment
(legal problems, alcohol/drug problems, use of community resources)

The QLQ is a 150-item instrument that has been validated with various population groups. In the Program Impact Monitoring System, as developed and implemented in Oregon, the instrument is administered by trained interviewers who interview the client at the time of admission to service and again at 90 days. There is a well-developed training program for the interviewers, and there is a system for assuring the quality of their work.

While the system had only recently been implemented in Oregon and had not yet had time to develop much outcome data for analysis, Tennessee was impressed that the number and specificity of the items on the questionnaire gave promise of the ability to discriminate specific areas of functioning in which clients were or were not showing changes. This appeared to be more clinically and programmatically useful than the more global instruments that were often used in other programs.

The Tennessee Department of Mental Health and Mental Retardation made the decision to use the QLQ instrument and the procedures developed in Oregon almost exactly as they were being used in Oregon. The agency moved quickly to solicit local community mental health centers that might be willing and able to participate in the implementation of the system. Ultimately, six mental health centers volunteered and were chosen with the agreement that they would concentrate on the chronically mentally ill, although these clients were not the exclusive focus of the effort. The staff from the Oregon project provided technical assistance in the design of the Tennessee system and in the selection and training of the interviewers. In Tennessee, as in Georgia, there was a very short time, about three months, between the decision to proceed and the implementation of data gathering in the mental health centers. However, the system was elective for the individual centers, and it was phased in for two of the six centers that were not yet ready to start when the overall implementation began.

In addition to these differences, the Tennessee Department was able to employ a person to give full-time assistance to the implementation of the system. He provided technical assistance and monitoring for the local programs. In Tennessee, the decision was made to have the data from the interviewers submitted to the Central Office for processing and analysis rather than having it done locally. The processing was done by computer and reported back to the local centers. Central processing also provided information for overall program evaluation.

Because of the research/demonstration nature of the implementation, the

system was undertaken with considerable rigor in terms of procedures, sampling, and analysis. There was significant attrition of cases in some of the programs, primarily because of the difficulties in contacting and interviewing clients once they were living in the community. However, there was relatively less attrition among the chronically mentally ill who were involved in day treatment and psychosocial rehabilitation programs, and they were the main focus of concern of the state agency.

The results showed that clients did indeed show improvements between the time of admission and the time of follow-up, but it also showed that they still fell short of the scores for levels of functioning of persons in the general community (Pokorny and Waters, 1981).

More important from a clinical program perspective, the individual scales lent themselves to analysis of which kinds of clients had improved in which areas of functioning. This provided clues for the staff for areas in which programs needed to be retailored. The data also lent themselves to analysis of which types of treatment programs yielded what kinds of results.

Local programs have found the data to be useful in providing feedback to clinicians and in identifying areas in which they *might* change the emphases of their services as well as providing assurances that their efforts are indeed yielding results for their clients.

There were some areas in which individual items on the QLQ seemed to need reworking in order to be more discriminating for certain kinds of clients, but the overall instrument seemed to be equally appropriate for persons of both sexes and for both black and white clients.

The system in Tennessee also experienced turnover of a key person, the Assistant Commissioner of Planning and Evaluation. Her position was eliminated in a reorganization, and she left for another state. However, the procedures and the determination to implement the system had by then been so well internalized in the remaining staff that the effort continued.

Since the effort of the replication study, the gathering of data has continued, but the activity has been considerably toned down, largely because of fiscal constraints that have made it difficult for the local programs to maintain the support for the interviewers, but also because Central Office staff have been required to diminish the amount of time they have been able to devote to this activity in order to cover other evaluation responsibilities. There is a feeling that the results may be too costly in terms of time and money for the value they have provided to program administrators.

The problem of time has been a serious concern in two ways: (1) the actual amount of time required by interviewers to gather the data (each administration of the instrument requires about four hours of arranging, travel, and direct administration) and (2) the length of time before the staff receives feedback from the computer. (The need to have a sufficient number of post-test clients sometimes delays the analysis for several months, during which

clinicians have moved, programs have changed, and few persons are left who are concerned.)

Another concern in Tennessee is that the data from the analyses are still not crisp enough to provide clear guidance for program planning. This is partly the result of having too few cases in the early phases, and partly because the scales (e.g., the productivity scale) are still not sharp enough in certain discriminations. However, this is one of the reasons administrators are reluctant to pursue further data gathering in a time when mental health programs are suffering fiscal retrenchment. Staff feel that the instrument probably should be revised and shortened to make a better-targeted package.

Another problem that became evident as the project went along was the high attrition rate of clients, once they returned to the community. This was true especially for chronic patients who were seen only in aftercare programs, but it was a serious problem throughout. This led to worsening of the time delay before sufficient cases could be analyzed and the reports sent back to the clinical programs.

DISCUSSION

The attempts to develop statewide outcome assessment programs in Tennessee have both been significant experiments. Both have continued past the initial implementation phases, but both have significantly diminished in their efforts to establish a statewide system of outcome measures.

From the analysis of these two efforts there seem to be some guidelines for the further development and implementation of statewide outcome measures:

- The development of a statewide system of outcome measures requires high-level commitment of the state mental health director and his/her key staff. It also requires resources and staff with the time to devote to developing, implementing, monitoring, and analyzing the system. It helps to have a sparkplug type of person in a leadership position in Program Evaluation.
- The system should be checked for validity and reliability before it is implemented, and periodically for various new client groups.
- The system should be implemented gradually and on a voluntary basis until the "bugs" have been worked out, and it is ready to extend to programs with lower levels of enthusiasm for outcome measurement.
- Training and ongoing technical assistance regarding problems must be provided to the persons who gather and analyze the data.
- The system must have some way of feeding back the data to clinicians and program administrators in a reasonable time frame. This was a particular problem in Tennessee.
- The system must have credibility and face validity for the clinicians. This

was frequently a problem in the Georgia effort, because the language and concepts of the Balanced Service System did not find wide acceptance among clinicians.

- The system should not be tied to reimbursements at the start. This is likely to cause resistance at best and sabotaging of the data at worst.
- The instrument should have sufficiently sharp distinctions to provide for program guidance. The Georgia Role Functioning Scale was too global to meet this need. The QLQ instrument also needs some refinement to better meet this requirement.
- The process of data gathering should be relatively inexpensive in both time and money so that sufficient numbers of clients can be included for meaningful analysis.
- If the follow-up post-test is to be done in the community, special attention must be given to the problem of attrition. This is known to be a problem in community studies of the mentally ill.

SUMMARY

The state mental health agencies in Georgia and Tennessee have both undertaken activities to develop statewide mental health outcome assessment programs within the past four years. Both have had their successes, and both are now languishing. This is partly the result of the fiscal constraints under which mental health agencies are presently operating, so that there is insufficient staff assigned to working with such systems. However, they are also languishing because of problems inherent in the systems that were implemented.

The Georgia Role Functioning Scale was an ambitious effort to implement a statewide system that applied to all clients and was conceived to be both a quality assurance mechanism and an outcome assessment program. It incurred substantial resistance from clinicians at the start, because of its universal implementation before the program was fully ready and because of its tie to reimbursements.

The Tennessee Program Impact Monitoring System used a validated and well-developed system already in use in Oregon. However, despite some significant successes, the program has encountered problems of costs, time delays, and attrition that have made administrators cautious about proceeding.

From these two experiences, some guidelines are offered for the future development of statewide mental health outcome assessment systems.

REFERENCES

Brodsky, G. and Bigelow, D.A. *The Oregon Program Impact Monitoring System: Final Report,* State of Oregon, Mental Health Division, Salem, OR, 1980.

Gerhard, R.J., and Dorgan, R.E. *The Balanced Service System,* The Community Mental Health Institute, Washington, DC, 1977.

Hargreaves, W.A., McIntire, M. A., Attkisson, C. C. and Siegel, L. M. Outcome measurement instruments for use in community mental health program evaluation. Part IV, *Community Mental Health Program Evaluation,* National Institute of Mental Health, 1975, 13-22.

Kiresuk, T.J. and Sherman, R.E. Goal attainment scaling: a general method for evaluating community mental health programs. *Community Mental Health Journal,* 1968 *4,* 443-453.

McIntyre, M.H., Attkisson, C.C. and Keller, T.W. Components of program evaluation capability in community mental health centers. Part I *Community Mental Health Program Evaluation,* W. A. Hargreaves et al (Eds) National Institute of Mental Health, *1974,* 3-50.

Pokorny, L. J. and Waters, R. M. *The Tennessee Mental Health Outcome Project: Final Report,* Tennessee Department of Mental Health and Mental Retardation, Nashville, TN, 1981.

Spitzer, R. L., Gibbon, M. and Endicott, J. Evaluating the effectiveness of services. Part IV *Community Mental Health Program Evaluation,* W. A. Hargreaves et al (Eds), National Institute of Mental Health, 1975, 28-29.