

RESEARCH AND PRACTICE NOTES

THE EXPERIENCE OF DEVELOPING A
PACKAGE OF INSTRUMENTS TO MEASURE
THE CRITICAL CHARACTERISTICS OF
COMMUNITY SUPPORT PROGRAMS FOR
PEOPLE WITH A SEVERE MENTAL ILLNESS

Brian Rush
Ellen Tate
Centre for Addiction and Mental Health
Toronto, Ontario

Ross Norman
University of Western Ontario
London, Ontario

Bonnie Kirsh
University of Toronto
Toronto, Ontario

Meghan Prosser
Centre for Addiction and Mental Health
Toronto, Ontario

T. Cameron Wild
University of Alberta
Edmonton, Alberta

Steve Lurie
Canadian Mental Health Association Metro Toronto Branch
Toronto, Ontario

Abstract: A major challenge for program evaluators is to develop and use measures of program implementation that are both comprehensive and generalizable across programs. The goal is to understand how measures of program characteristics are related to outcomes and, ultimately, to improve program effectiveness. In this brief report, we describe our experience in developing a package of instruments to measure critical characteristics of community support programs for people with a severe mental illness. We highlight three primary methodological challenges

Corresponding author: Brian Rush, Health Systems Research and Consulting Unit, Centre for Addiction and Mental Health, 33 Russell St., Toronto, ON M5S 2S1; <Brian_Rush@camh.net>

encountered: ensuring comprehensiveness, attaining a common understanding of program boundaries, and incorporating multiple perspectives. Potential solutions are discussed. We conclude that, despite these challenges, the goal of model-free measures of program characteristics is essential to enable meaningful program comparisons and to substantiate important process-outcome links.

Résumé: L'un des défis majeurs concernant l'évaluation des programmes est le développement et l'utilisation d'outils, à la fois pratiques et généralisables, permettant de mesurer la mise en oeuvre de ces programmes. Le but est de comprendre comment les caractéristiques des programmes sont associées aux résultats, afin d'améliorer leur efficacité. Ce bref rapport décrit notre expérience dans le développement d'un ensemble d'outils destinés à mesurer les caractéristiques critiques des programmes communautaires de soutien aux personnes atteintes de troubles de santé mentale sévères. Nous présentons trois défis méthodologiques auxquels nous avons été confrontés : présenter une vue d'ensemble du travail; définir les limites des programmes de façon compréhensible; et intégrer des perspectives multiples. Les solutions potentielles sont également discutées. Nous concluons que, malgré ces défis, le but de développer des mesures des caractéristiques des programmes, indépendantes des modèles existants, est essentiel pour rendre possible des comparaisons plausibles entre les programmes et de démontrer des liens importants entre les processus et les résultats.

Various approaches have been used to expand the scope of evaluation designs to better understand programs as they are implemented and to relate the program characteristics to outcomes. *Theory-based evaluation* emphasizes conceptual clarity about what components and activities of a program should predict outcomes (Chen, 1990, 1994; Chen & Rossi, 1992; Finney & Moos, 1989; Lipsey, 1990). *Fidelity* approaches have used scales or standards to assess the adherence of programs to an original model on selected parameters (e.g., Mowbray, Holter, Teague, & Bybee, 2003). Other approaches have tailored *program implementation analysis* to focus attention on unique program components and characteristics and their link to outcomes (Brekke, 1987; Cousins, Aubry, Fowler, & Smith, 2004; Yates, 1996). Each approach has advantages and disadvantages. In essence, the unanswered challenge for evaluation practitioners is to develop and use measures of program implementation that are comprehensive, explanatory, and generalizable across program contexts.

In this article, we describe our experience in meeting this challenge in the context of developing measures to assess critical characteristics of community mental health programs for people with a severe mental illness.

OUR APPROACH

We focused our attention on community support programs being implemented in Ontario that have the common objectives of providing long-term support needed to improve quality of life and to make more appropriate and, hopefully, reduced use of psychiatric hospitalization. In practice this included a number of “models” such as Assertive Community Treatment teams (ACT), other intensive case management programs, clubhouses, housing support, and consumer-run services. Our mixed-method approach consisted of several phases:

1. Identification of an initial list of program characteristics that may be predictive of improved quality of life and more appropriate hospitalization through qualitative analysis of interview data (clients, service providers, and family members) and content analysis of an exhaustive literature review;
2. Design and item development of two questionnaires: one for consumers and one for service providers based on the identified themes and sub-themes;
3. Distillation of the list of items and revision of questionnaires in stages based on pilot testing and a test-retest with consumers and service providers selected from a diverse group of programs, both geographically and organizationally.

METHODOLOGICAL CHALLENGES AND POTENTIAL SOLUTIONS

In this brief report we focus on three challenging methodological issues: establishing comprehensiveness, identifying program boundaries, and assessing multiple perspectives.

Comprehensiveness

Working from the interviews and selected literature, our item development process resulted in an initial set of over 400 items per questionnaire, which was then subjected to field-testing. Our field-testing immediately highlighted the necessary trade-offs in terms of coverage and potential response burden. On the one hand, the average length

of time for consumers to complete the face-to-face interviews was 2.5 hours (estimated range 1.5 to 5 hours), while service providers reported spending an average of 3.7 hours (estimated range 1.5 to 6 hours) for a self-administered format. Both groups reported that they found the questionnaire long. Fatigue and loss of concentration may have affected the reliability of reports, despite our test administration procedure, which allowed as many breaks as needed. On the other hand, consumers reported that the experience was useful for educating them about the full range of services that they had received or that might be available to them. The service providers said that the experience was both educational and self-reflective, as it made them think of potential services and supports people *may* need but which they were not currently providing. When asked if there were items that were less important than others, or could be eliminated, both service providers and consumers typically answered, "Everything is *potentially* important." This reluctance to nominate domains or items for removal was linked to the highly individualized nature of consumer needs and services provided.

Program Boundaries

Program evaluations are typically focused on individual organizational entities that provide a set of treatment/support activities for people with severe mental illness. We found, however, that consumers and service providers had many different conceptions of the "program" as an ensemble of activities and services.

Substantial numbers of consumers did not know the name of the program, nor the broader range or structure of services offered. If they were receiving services from more than one provider within the agency or within the community, they sometimes found it difficult to respond only in reference to the focal program under consideration. There was also ambiguity as to whether a service that is brokered through or facilitated by the focal program should be identified as part of the services provided by that program. Service providers also varied in their understanding of the program boundaries. For instance, some program staff had duties in other programs within a multi-service agency, which were confused with being part of the target program. The program structure also appeared to influence how well the program boundaries were recognized. Staff working in team structures were better able to respond to what the program did "in general," or with "most consumers" than were case managers who tended to work less frequently with other staff, and therefore had less direct

knowledge of overall program activity. Participants from programs that were predominantly peer support and had fewer paid staff interpreted the program's role with the most variability. The role blurring and absence of professional boundaries within consumer-run programs could explain the variability in responses.

Multiple Perspectives

At the outset we developed virtually parallel versions of the service provider and consumer questionnaires, anticipating that a comparison of the perspectives of consumers and service providers would provide evidence of convergent validity. However, in the first round of testing the two perspectives were markedly different. Service providers tended to report various services being provided more frequently than the aggregated consumer data would suggest. The consumers showed much more variation in their responses than did service providers. One reason for the discrepancies may at least partially reflect basic differences in the understanding of the functions of staff-consumer interactions. For instance, a service provided by a worker such as "transportation to an appointment," which may also have involved some personal counselling, may simply be seen by the consumer as "a ride."

Using exactly the same response categories became a major challenge. Consumers were asked to respond based on their individual experience, whereas service providers were asked to report on their experiences with a group of consumers. Through our field-testing we discovered that service providers needed a more sophisticated response structure to be able to reflect both the number of consumers to whom they provided a service *and* the amount of time spent on particular services with consumers. Some clients had long involvement with the program; others were new. Some had required quite intensive services for some period of time, while others received more maintenance-like services if they were comparatively stable in the community. This complexity continually raised issues of how to capture both the nature and intensity of what was being delivered.

SUMMARY

Despite the challenges we have faced and trade-offs that are necessary in the measurement process, we remain convinced of the im-

portance of developing instruments to assess critical characteristics of support programs in a sufficiently comprehensive fashion so as to enable comparisons across different types of programs and to substantiate process-outcome links. Based on the feedback from consumers and service providers, we believe that our instruments have content validity and are valuable educational tools. At a program level, this process of self-reflection could be formalized within the context of a program review or continual quality improvement. A less complex "Program Characteristic Checklist" has been prepared (Service Provider and Consumer versions), and a manual is under development to guide its use in an internal evaluation or quality improvement context. In addition, the large variability we have seen in the many programs working with us reinforces the need for measures of program characteristics that will be sensitive to their unique features rather than only those features prescribed in model-based fidelity scales. One criticism of these model-based fidelity scales is their emphasis on program structure rather than processes, in large part because structure is easier to measure (Mowbray et al., 2003).

Nevertheless, the comprehensiveness of the instruments brings with it challenges related to participant fatigue and overall response burden. Further analyses are planned to reduce the length of the instruments through other psychometric analyses (e.g., item analysis and factor analysis), and the use of technology is being explored as a way to reduce respondent burden. Given its length, additional psychometric analyses, construct validation, and assessment of predictive utility for treatment outcome may have to be undertaken within subcomponents of the instruments.

Variable understanding of program boundaries between respondents associated with the same program is cause to re-examine our assumptions about the utility of a program-focused approach. While the idea of a well-defined set of services or "program" is important for internal evaluation purposes, the research interest of establishing a link between process and consumer outcomes may require a more systems-oriented approach that examines supports received by the consumer regardless of their source in the community.

Our finding concerning the discrepancy in perceptions of service providers and consumers has recently been reported by others (Crane-Ross, Roth, & Lauber, 2000). Such differing results obtained from consumers and service providers raise important issues of validity and suggest to us that such a systems-level evaluation should

also incorporate some additional “objective” measures of services provided. These might be obtained through chart audits, service logs, or participant observation, all of which have been suggested in the literature as potential sources of data on critical characteristics (Mowbray et al., 2003).

Ultimately, the best approach in going forward with a research agenda on measuring critical program characteristics will be to reduce our comprehensive shopping list of characteristics via a well-funded, multi-level outcome study that will allow enough statistical power to identify those program attributes associated with consumer outcomes. The “acid test” will be to see which questionnaire items are most closely associated with outcomes and ultimately allow a trimming of the instrument on the basis of predictive validity criteria.

REFERENCES

- Brekke, J.S. (1987). The model-guided method for monitoring program implementation. *Evaluation Review*, 11, 281–299.
- Chen, H.T. (1990). *Theory-driven evaluation*. Newbury Park, CA: Sage.
- Chen, H.T. (1994). Theory-driven evaluations: Need, difficulties and options. *Evaluation Practice*, 15, 79–82.
- Chen, H.T., & Rossi, P. (1992). *Using theory to improve program and policy evaluations*. Westwood, CT: Greenwood.
- Cousins, J.B., Aubry, T.D., Fowler, H.S., & Smith, M. (2004). Using key component profiles for the evaluation of program implementation in intensive mental health case management. *Evaluation and Program Planning*, 27, 1–23.
- Crane-Ross, D., Roth, D., & Lauber, B.G. (2000). Consumers' and case managers' perceptions of mental health and community support service needs. *Community Mental Health Journal*, 36, 161–178.
- Finney, J.W., & Moos, R.H. (1989). Theory and method in treatment evaluation. *Evaluation and Program Planning*, 12, 307–316.
- Lipsey, M.W. (1990). Theory as method: Small theories of treatments. In L. Sechrest, E. Perrin, & J. Bunker (Eds.), *Research methodology: Strengthening causal interpretations of nonexperimental data*. Washington, DC: U.S. Department of Health and Human Services.

Mowbray, C.T., Holter, M.C., Teague, G.B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315–340.

Yates, B.T. (1996). *Analyzing costs, procedures, processes, and outcomes in human services* (Vols. 42). Thousand Oaks, CA: Sage.

Brian Rush, Ph.D., is a senior scientist and associate director at the Health Systems Research and Consulting Unit, Centre for Addiction and Mental Health, Toronto, Ontario.

Ellen Tate is the project coordinator of the Explaining Outcomes Project at the Centre for Addiction and Mental Health, Toronto, Ontario.

Ross Norman, Ph.D., is a professor in the Departments of Psychiatry and Epidemiology and Biostatistics and Director of Research at the Department of Psychiatry, University of Western Ontario, London, Ontario.

Bonnie Kirsh, Ph.D., is an assistant professor in the Department of Occupational Therapy and Graduate Department of Rehabilitation Sciences, University of Toronto, Toronto, Ontario.

Meghan Prosser is a research analyst with the Explaining Outcomes Project at the Centre for Addiction and Mental Health, Toronto, Ontario.

T. Cameron Wild, Ph.D., is an associate professor at the Centre for Health Promotion Studies and the Department of Public Health Sciences, University of Alberta, Edmonton, Alberta.

Steve Lurie, B.A., M.S.W., M.Mgt., works at the Canadian Mental Health Association Toronto Branch, Toronto, Ontario.