

Practical Monitoring of Treatment Fidelity: Examples From a Team-Based Intervention for People With Early Psychosis

Susan M. Essock, Ph.D., Ilana R. Nossel, M.D., Karen McNamara, L.C.S.W.-C., Ph.D., Melanie E. Bennett, Ph.D., Robert W. Buchanan, M.D., Julie A. Kreyenbuhl, Pharm.D., Ph.D., Sapna J. Mendon, L.M.S.W., Howard H. Goldman, M.D., Ph.D., Lisa B. Dixon, M.D., M.P.H.

Mental health programs can address many components of fidelity with routinely available data. Information from client interviews can be used to corroborate these administrative data. This column describes a practical approach to measuring fidelity that used both data sources. The approach was used in the Recovery After an Initial Schizophrenia Episode (RAISE) Connection Program, a team-based intervention designed to

implement evidence-based practices for people experiencing early psychosis suggestive of schizophrenia. Data indicated that the intervention was implemented as intended, including program elements related to shared decision making and a range of evidence-based clinical interventions.

Psychiatric Services 2015; 66:674–676; doi: 10.1176/appi.ps.201400531

Fidelity measures serve multiple stakeholder groups: payers, trainers, supervisors, clients, and families. Payers want to know if they are getting what they are paying for. Trainers and supervisors want to know if training succeeded and whether clinical staff members are implementing interventions as intended. Clients and families want to know if services are effective and can be expected to promote outcomes that they care about (for example, in regard to school, work, friends, and health). Fidelity measures are critical to understanding how good outcomes are achieved, replicating successful programs, enhancing efficacy, and measuring performance over time.

This column describes a practical approach to measuring fidelity used in the Recovery After an Initial Schizophrenia Episode (RAISE) Connection Program, a team-based intervention designed to implement evidence-based practices for people experiencing early psychosis suggestive of schizophrenia (1,2). The project was carried out in partnership with state mental health agencies in Maryland and New York as part of the RAISE initiative funded by the National Institute on Mental Health (1–3) and enrolled 65 adolescents and young adults with early psychosis suggestive of schizophrenia across two sites (Baltimore and New York City). Each team included a full-time team leader (licensed clinician), full-time employment and education specialist, half-time recovery coach (licensed clinician), and a 20%-time psychiatrist. Teams used assertive outreach strategies and shared decision making to engage participants in care. Using a critical time intervention model (4), teams provided services for

up to two years, with the goal of helping people stabilize their psychiatric conditions; reintegrate with school, work, and family; and transition to appropriate community services and supports. Participants provided informed consent. Participating institutions' institutional review boards approved study procedures.

WHAT MAKES GOOD FIDELITY MEASURES?

Optimal fidelity measures are informed by evidence and are good proxies for the intervention components being measured, objective, and drawn from readily available information. Routine service logs or billing data support many fidelity measures (5). For example, such data have been used to document whether assertive community treatment teams are, indeed, delivering services intensively and whether clients are being served by multiple staff (5). Other easily obtainable, objective, preexisting data can address structural requirements (for example, minimum staffing and after-hours coverage) and processes of care (for example, side-effect checklists indicating that assessments were conducted). Such measures may be most useful in determining whether an implementation is minimally adequate as opposed to, for example, discriminating among exemplary programs (6).

Even when extensive administrative data exist, some topics are best addressed by self-reports from service users (7). For example, clients' ratings would be preferable to staff ratings of whether staff used shared decision making.

MEASURING FIDELITY TO A TEAM-BASED INTERVENTION

This report by the RAISE Connection Program's lead researchers and intervention developers expands on implementation findings reported elsewhere (1), describes how we monitored treatment fidelity by using measures based on the principles stated above, and provides fidelity findings.

RAISE Connection Program researchers worked with the lead developers of core treatment domains (team structure and functioning, psychopharmacology, skills building, working with families, and supported employment and education) to determine performance expectations for each domain and how to operationalize those expectations. To do so, the researchers used information commonly available electronically to programs that bill for services (hereafter, "administrative data"). [The table included in an online data supplement to this column lists performance expectations for these program domains and their operationalization into fidelity measures.] We followed this approach to enhance generalizability, even though we could not ourselves use claims data to extract data on service use because initially this research project was funded by federal research dollars that precluded the sites from billing for services. Rather, we relied on research staff to extract information from routine service logs maintained by clinical staff and from specific fields in medical records (for example, medication records). All data came from such objective sources, as opposed to reading through line by line of progress notes (1).

In addition to these measures derived from program data, researchers worked with the lead developers of the core treatment domains to identify questions to ask clients to determine whether, from clients' perspectives, intervention components had been implemented. [A figure included in the online supplement lists the questions and presents data on clients' responses.] These questions were embedded in structured research interviews that participants completed at six-month intervals after enrollment (1) and provided corroboration for fidelity measurements obtained from program data. For example, we measured the expectation that "The psychiatrist and client regularly review medication effectiveness and side effects" both by noting psychiatrists' completion of standardized side-effect-monitoring forms and by asking clients, "How much did your Connection Team psychiatrist bring up the topic of medication side effects?"

For many fidelity measures, the RAISE Connection Program had no preexisting standard to adopt for what constituted "acceptable" performance. Rather, data collected during the project were used to generate expectations based on actual performance.

FINDINGS

Both teams met or exceeded most performance targets [see table in the online supplement for a summary of fidelity data, measured across time and for the project's final complete

quarter]. Data from client interviews also indicated high fidelity to the model. The large majority of clients reported that teams paid attention to their preferences about jobs and school, made treatment decisions—including medication decisions—jointly, and responded quickly.

Tables or figures that show how individual programs compare to a standard are helpful in spotting deviations from expectations and performance outliers and also are clear reminders of program expectations, changes in performance over time, and how one's own program compares with other programs. For example, the second figure in the online supplement illustrates that both teams exceeded the performance expectation, set by consensus of the intervention developers, that at least 10% of clients meet with team members in the community, excluding visits with employment and education specialists.

We also used fidelity measures based on program data to examine differences between teams. Recovery coaches had different styles of providing services across sites. The coach at site 1 provided almost all services in a group format, and the coach at site 2 provided a mix of group and individual sessions.

Fidelity is a team-level measure, yet many fidelity measures are composed of aggregated client-level data (for example, whether a client has had an adequate trial of an antipsychotic). These measures can be used to generate exception reports (for example, lists of clients for whom no meeting has been held with a family member) that can be fed back to teams and supervisors to identify areas for improvement. By study end, we were able to provide data to teams from such exception reports and share data with supervisors.

For some expectations (for example, that employment and education specialists accompany clients to work or school when clinically indicated and desired by the client), we anticipated that only a small fraction (for example, 10%) of clients would endorse the item because the service in question may be relevant for few clients. For such measures, small but nonzero findings provide proxy measures indicating that treatment components were implemented. For other treatment components, we expected most clients to endorse the item because the component (for example, shared decision making) was relevant to all participants.

MEASURING FIDELITY EFFICIENTLY AND USING FIDELITY FINDINGS

A core challenge in measuring treatment fidelity is to do so reliably and without breaking the bank, ideally by using data already being collected for other purposes. Although research studies may rely on review of videotapes or site visits to observe program implementation, such efforts can be too labor intensive for broad implementation. Bringing model programs to scale calls for cost-effective, sustainable approaches to measuring fidelity. Increasingly, payers' contracts with programs have dollars at stake with respect to maintaining program fidelity. Fidelity data used for such purposes need to be reliable and objective, requirements that may not be met by data from summary impressions of site visitors or small samples of observations.

Routine service logs will support many fidelity measures so long as they note for each contact the client and staff involved, whether family members were present, and the location of service (for example, office versus community). Use of routine clinical forms, such as those included in the RAISE Connection Program treatment manual (8), both support the intervention and can be used to document that certain intervention components occurred.

Obtaining fidelity data from claims data and other preexisting sources minimizes the data collection and compilation burden on staff. However, as a fallback to using administrative data to measure fidelity, payers may specify data that programs are required to submit, and those submissions can be verified, in toto or at random, via site visits. Designing, building, debugging, and implementing an accompanying chart abstraction system is cumbersome for short-term use but offers an alternative when abstraction from electronic claims is not possible.

As data accrued, we were able to see that most expectations appeared reasonable, and we used early data from these teams to revise staffing and performance standards for new teams being rolled out in New York under the OnTrackNY initiative (practiceinnovations.org/CPIInitiatives/OnTrackNY/tabid/202/Default.aspx). For example, lower-than-expected rates of metabolic monitoring led to the addition of part-time nurses to OnTrackNY teams.

Even when data aren't sufficient to establish precise performance thresholds, such data allow program managers to identify outliers (for example, a team that never provides services off site). Such outliers needn't indicate poor performance, but they point to areas for further investigation and follow-up, perhaps via site visits. For program start-up, knowing that the service exists may be sufficient. If stakeholders become concerned that clients who are in need of the service aren't getting it, then a more nuanced measure would be called for.

Site visits can be costly and time consuming for routine fidelity monitoring, particularly in large systems, but they can be helpful to reinforce training and elucidate factors underpinning unusually good or poor performance on fidelity measures derived from administrative data.

Programs and their funders need to budget for fidelity measurement as core program costs. Building such data-reporting requirements into contracts helps ensure adequate budgeting. Fiscal bonuses for meeting performance expectations provide performance incentives. As noted above, although we do not always know a priori what good performance looks like, often we can define what is minimally adequate, so that we can specify and monitor accordingly. With such data, we can identify good outliers (for example, teams with high rates of engagement) and feature them in efforts to improve performance.

CONCLUSIONS

Programs can use routinely available data to determine whether many key components of an intervention have

been implemented. Fidelity data from multiple sources indicate that the RAISE Connection Program was implemented as intended for the range of expected clinical interventions, including program elements related to shared decision making.

AUTHOR AND ARTICLE INFORMATION

Dr. Essock, Dr. Nossel, and Dr. Dixon are with the New York State Psychiatric Institute and the Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York City (e-mail: se2176@columbia.edu). Dr. McNamara, Dr. Bennett, Dr. Kreyenbuhl, and Dr. Goldman are with the Department of Psychiatry, University of Maryland School of Medicine, Baltimore. Dr. Buchanan is with the Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore. Ms. Mendon is with the New York State Psychiatric Institute, New York City. Marcela Horvitz-Lennon, M.D., M.P.H., is editor of this column. This column is part of a special section on RAISE and other early intervention services.

This work was supported in part with federal funds from the American Recovery and Reinvestment Act of 2009 and the National Institute of Mental Health (NIMH) under contract HHSN271200900020C (Dr. Dixon, principal investigator), by the New York State Office of Mental Health, and by the Maryland Mental Hygiene Administration, Department of Health and Mental Hygiene.

The authors thank Robert Heinssen, Ph.D., and Amy Goldstein, Ph.D., at NIMH for their efforts to bring the RAISE initiative to fruition and Dianna Dragatsi, M.D., Jill RachBeisel, M.D., and Gayle Jordan Randolph, M.D., for their ongoing support. Jeffrey Lieberman, M.D., was the original principal investigator on the NIMH contract, and the authors thank him for his foresight in assembling the original research team.

As part of the RAISE Connection Program, Dr. Essock, Dr. McNamara, Dr. Bennett, Ms. Mendon, Dr. Goldman, and Dr. Dixon may be part of training and consultation efforts to help others provide the type of FEP services described here. They do not expect to receive compensation for this training other than that received as part of work done for their employers. Dr. Buchanan has served on advisory boards of or as a consultant to Abbott, Amgen, BMS, EnVivo, Omeros, and Roche. He is a member of the data safety monitoring board for Pfizer. The other authors report no financial relationships with commercial interests.

REFERENCES

1. Dixon LB, Goldman HH, Bennett ME, et al: Implementing coordinated specialty care for early psychosis: the RAISE Connection Program. *Psychiatric Services* 66:692–699, 2015
2. Essock SM, Goldman HH, Hogan MF, et al: State partnerships for first-episode psychosis services. *Psychiatric Services* 66:672–674, 2015
3. Lieberman JA, Dixon LB, Goldman HH: Early detection and intervention in schizophrenia: a new therapeutic model. *JAMA* 310: 689–690, 2013
4. Herman D, Conover S, Felix A, et al: Critical Time Intervention: an empirically supported model for preventing homelessness in high risk groups. *Journal of Primary Prevention* 28:295–312, 2007
5. Essock SM, Kontos N: Implementing assertive community treatment teams. *Psychiatric Services* 46:679–683, 1995
6. Wisdom JP, Knapik S, Holley MW, et al: New York's outpatient mental health clinic licensing reform: using tracer methodology to improve service quality. *Psychiatric Services* 63:418–420, 2012
7. Essock SM, Covell NH, Shear KM, et al: Use of clients' self-reports to monitor Project Liberty clinicians' fidelity to a cognitive-behavioral intervention. *Psychiatric Services* 57:1320–1323, 2006
8. Coordinated Specialty Care for First Episode Psychosis. Manual II Implementation. Bethesda, Md, National Institute of Mental Health, 2014. Available at www.nimh.nih.gov/health/topics/schizophrenia/raise/CSC-for-FEP-Manual-II-Implementation-Manual_147093.pdf