ISSUE BRIEF

Coordinated Specialty Care for People with First Episode Psychosis: Assessing Fidelity to the Model

AUTHORS:
Susan Essock, Ph.D., Department of Psychiatry, Columbia University
Donald Addington, M.D., Department of Psychiatry, University of Calgary

Technical Assistance Material Developed for SAMHSA/CMHS under Contract Reference: HHSS283201200002I/
Task Order No. HHSS28342002T

When implemented with high fidelity to the model, Coordinated Specialty Care (CSC) is associated with superior outcomes for individuals experiencing early psychosis (also called First Episode Psychosis, or FEP). In practice, teams charged with implementing CSC services may adhere closely to the CSC components set forth by the National Institute of Mental Health (NIMH) (Heinssen, Goldstein, Azrin, 2014), or they may deviate substantially from the model. How is a patient, family member, team leader, trainer, administrator, or payer to know whether a team is delivering high-fidelity CSC?

In this issue brief, we review the importance of measuring implementation fidelity, list the core components of CSC to monitor via fidelity measures, suggest potential data sources for measuring these core components, and review ways data from these sources can be aggregated into measures of CSC fidelity. Throughout, unless otherwise noted, comments refer to programs operating in the United States.

While aspects of this brief will be of interest to all of the stakeholders listed earlier, the brief will be of particular relevance for individuals who have responsibility for overseeing one or more FEP programs such as state FEP coordinators. The executive summary provides a comprehensive overview of the material that is included in the body of the brief with much greater detail found in the full text.

EXECUTIVE SUMMARY

The superior outcomes associated with CSC can be expected with high-fidelity implementation of the NIMH CSC model, so measuring implementation fidelity to the model is important.

The first steps in measuring implementation fidelity are to identify the core program components that are expected to be used in the CSC intervention, how those core components will be operationalized into measurable performance indicators, and the criteria to determine if fidelity expectations are being met. The various experts who have published approaches to measuring the fidelity of FEP programs largely agree on what they consider to be the core components to be measured, even if they propose different approaches to measuring and summarizing them.

Various data sources and data-collection strategies, alone or in combination, can be used to measure fidelity. Strengths and weaknesses associated with several different strategies are summarized in the table on page 3.

METHOD	STRENGTHS	WEAKNESS
Administrative data (claims, EMRs)	There is no need to sample, as data are present for all clients. Costs of data collection are covered by other entities.	Reports must be programmed to extract and summarize the administrative data. Not all monitoring entities have access to administrative data.
Record abstraction	Random samples can produce unbiased estimates of what is being monitored. Data abstraction protocols can be specific enough to be used by non-clinicians with minimal training.	Charts may not include the data being sought. Records may be difficult to access or read. If records must be reviewed on site, travel costs can mount. Data abstraction protocols can be difficult to develop and reliably implement, particularly with regard to maintaining an audit trail.
Staff self-report	This approach is relatively inexpensive.	This strategy entails a risk of gaming in the absence of external checks. Data provided by sites still needs to be aggregated and reported across sites. This approach can be labor-intensive for clinical staff, who must provide data that is detailed enough to accurately assess performance expectations.
Telefidelity (structured telephone interviews)	This strategy can be deployed over large distances. This strategy can be implemented using a small team of interviewers.	Feedback is less personalized. Qualitative aspects of the program (e.g., youth-friendliness) cannot be assessed. This strategy provides less opportunity to evaluate the broader administrative context than on site visits. Context is important for sustaining fidelity to the model. Interviewers must be trained to score reliably. Turnover of trained interviewers may disrupt the ability to conduct interviews.
Site visits	This strategy can provide a forum for clinical strategizing and it can address concerns other than fidelity (e.g., training). This strategy enables the assessment of qualitative aspects of the program (e.g. whether facilities are youth-friendly).	This strategy is expensive. It requires large numbers of experts trained to score reliably. Turnover of trained site visitors may disrupt the ability to conduct site visits. Activities observed at a site visit may not be typical of everyday practice.
Client self-report	This strategy enables the assessment of certain client perceptions of program components that would otherwise be difficult or impossible to assess (e.g. the expectation that treatment decisions will be made jointly by the client and the staff).	This approach measures only a subset of performance expectations. Under this strategy, infrastructure must be created to allow data collection and summarization. Some clients may not be reachable or may decline to participate.

Regardless of the data sources a program monitor may use to measure fidelity, collecting, summarizing, and reporting data comes with costs. The list of what people would *like* to measure can get very long, with associated expense from increased complexity. Therefore, program administrators should use the minimum set of measures required to determine whether a CSC team is functioning as anticipated, and they should budget for such reporting and performance monitoring.

If you are a state administrator responsible for contracting with CSC sites, it is prudent to include in the contract the data-reporting requirements the sites will be required to meet. It is important that the sites understand that the costs specified in the contract include their costs for providing data and that performance monitoring will involve viewing and reporting client-specific data.

In addition to specifying performance expectations and identifying data sources, fidelity protocols must specify how to analyze fidelity information in such a way that inferences can be made about whether CSC programs have implemented the intervention as intended. Programs may opt for simple pass/fail ratings (i.e., met the standard or didn't meet it); a 3-point scale (below expectation/meets expectation/exceeds expectation); or a more finely demarcated scaling. As a complement or alternative to setting an overall threshold for determining what constitutes a passing score, contracts and/or program administrators may specify particular components that, if failed, result in an overall failing score.

Setting a threshold for a pass-fail judgement requires the identification of an acceptable standard as to what constitutes minimally adequate performance. In the absence of such data, programs may choose to monitor trends in scores prior to setting a threshold for passing or to set the threshold so low that, in effect, teams are being told "You must do this at least some". As data accrue from multiple teams, performance expectations can be tuned accordingly.

> **Clients, family members, team leaders, trainers, administrators, and payers all have a stake in knowing whether a given team is delivering high-fidelity CSC.**

Clients, family members, team leaders, trainers, administrators, and payers all have a stake in knowing whether a given CSC team is delivering high-fidelity CSC. How best to measure fidelity depends on local circumstances, and on such considerations as the purpose of the fidelity assessment, regulatory requirements, resources (time and money), availability of administrative data for electronic summarization, and performance expectations to which the CSC teams are being held by their funding sources. Selecting among the options for measuring CSC intervention fidelity means assessing the strengths and weaknesses of various approaches to determining performance expectations, how adherence to each expectation will be measured, and whether and how the results of the fidelity assessment will be shared across CSC sites. States and other payers will, no doubt, increasingly specify their performance expectations for CSC teams and how adherence to these expectations will be measured. Their actions will be informed, in part, by the information sources and summarization methods reviewed in this document and by the accumulating data on CSC team performance and client outcomes.

THE WHY'S AND HOW'S OF MEASURING PROGRAM FIDELITY

Why measure fidelity? Diverse stakeholders want to know, "Has this particular CSC program been implemented as planned?" and "Will participation in the program result in the expected outcomes for the client?" Payers want to know if they are getting what they are paying for; trainers and supervisors want to know if the training they provided has resulted in clinical staff implementing the interventions as intended over time; clients and families want to know if the services in which they are investing their time, effort, and finances can reasonably be expected to promote the outcomes they care about—in school or at work, in their friendships, and with regard to their overall health. (Essock, Nossel, McNamara, et al., 2015).

> **Payers want to know if they are getting what they are paying for.**

The payer for CSC services has powerful tools to assure high-fidelity implementation. The payer can devise contract terms that are generous enough to allow for high-fidelity implementation and that are enforceable, with sanctions for non-performance that are punitive enough to incentivize high-fidelity interventions (Essock, 2016). Implementing evidence-based practices at high fidelity means *doing* the practices so that service recipients, program staff, and payers can expect to achieve the outcomes associated with the intervention.

A core challenge to measuring intervention fidelity involves identifying feasible ways to monitor ongoing program performance to determine whether the intervention is being implemented as intended. At one extreme, approaches to measuring fidelity can be so burdensome that they aren't feasible; at the other extreme, they can be so simplistic that they fail to capture the quality of the clinical care being delivered. Depending on what data are readily available and what are more inexpensive/reliable/easy to obtain, programs may implement different approaches for measuring fidelity. Different settings likely have different sources for the same data. For example, when examining a team's staffing, some settings will have human resources departments that can summarize electronically maintained administrative data to report what staff, of what disciplines, are assigned to a team over time; other settings may track such information by hand.

Whatever approach to fidelity measurement ultimately is selected, the first steps in measuring implementation fidelity are to identify the core program components that are expected to comprise the CSC intervention (e.g., multidisciplinary team including a medication prescriber, provision of supported employment/supported education services, etc.); how those core components will be operationalized into measurable performance indicators; and the criteria to determine if fidelity expectations are being met. The various experts who have published approaches to measuring the fidelity of FEP programs largely agree on what they consider to be the core components to be measured, even if they propose different approaches to measuring and summarizing them.

Experts largely agree on the core components of CSC.

Identifying core components of CSC services. The NIMH Recovery After Initial Schizophrenia Episode (RAISE) initiative included the development of an implementation manual that enumerates the core components listed below (NIMH CSC Implementation Manual II, https://www.nimh.nih.gov/health/topics/schizophrenia/raise/csc-for-fep-manual-ii-implementation-manual_147093.pdf ; Heinssen, Goldstein, Azrin, 2014).

- Multidisciplinary team
- Small caseload
- Meets as a team
- Staffed with individuals with training and expertise in treating people with FEP
- Intake occurs promptly
- Uses shared decision making
- Provides recovery-focused treatment
- Uses personalized treatment plans
- Provides medication management
- Promotes skill building
- Works with families
- Provides supported employment/supported education services

Although the manual in the first link above lists 12 core program components, CSC programs may decide to lump or split individual components to meet their own purposes. For example, the first three components listed could be combined into a component called "Intensive interdisciplinary teams." Moreover, not every component is supported by the same level of evidence. For example, there is strong evidence for using antipsychotic medication and employment support services, while the evidence base regarding what constitutes a "timely intake" is less precise. Where the evidence base is sparse, CSC teams, payers, trainers, or administrators may opt to specify performance expectations that are based on local circumstances and clinical insights. Such measures may need to be modified as the evidence base and local circumstances change.

As the evidence base evolves, so do performance expectations. Researchers and administrators will note that the collection of data that address the 12 core components can be used to strengthen the evidence base. As programs mature and more outcomes data become available, researchers and administrators increasingly will be able to relate variations in the fidelity of CSC components to variations in outcomes. Over time, information about the relationships between the fidelity measures and outcomes can be used to set performance expectations across programs and/or over time (Essock, Nossel, McNamara, 2015). Similarly, performance data can be reviewed to identify CSC teams that are performance outliers (e.g., teams that make no off-site visits; teams that never prescribe clozapine) even without evidence-based standards for

specifying ideal performance. Identifying high-end performance outliers (e.g., teams that have high rates of family involvement, or high rates of metabolic monitoring) can help identify best practices for achieving particular performance expectations that can be helpful to share with lower-performing sites.

> **Structural components of a CSC program, such as the number of staff members, may be the easiest components to measure reliably.**

Operationalizing Performance Measures. Structural components of a CSC program, such as the number of staff members, their full-time equivalents, and the number of clients served, may be the easiest components to measure reliably (Mowbray, Holter, Teague, et al., 2003).

Performance expectations that necessitate measuring processes of care, such as the time between referral and intake or percentage of clients on a CSC team who have had an adequate trial of an antipsychotic medication, may also be relatively straightforward to measure. This is especially true when claims data or data from electronic medical records include information on dates and types of service provided. Proxies for processes of care, such as the use of shared decision-making, may be more difficult to discern from administrative data or staff self-reports, but they lend themselves well to "ask-the-client" measures that can be attached to satisfaction surveys (Essock, Nossel, McNamara, et al, 2015). If a process of care, such as the use of shared decision-making, cannot be measured directly, proxies for direct measurement may include determining whether staff members have participated in training on the topic.

Sometimes the presence of a treatment component can be inferred, based on some other observable measure. For example, the OnTrackNY CSC program uses clients' voluntary participation in treatment as a proxy for engagement in treatment via pursuing client-centered goals (Bello, Lee, Malinovsky, et al., 2017). The rationale for using this proxy is the belief that, if clients are opting to engage with the treatment team, then the team must be focusing on goals valued by the clients. These more remote proxies for the treatment component being measured may be useful for training and quality assurance purposes, even though such measures may be too imprecise for regulatory or certification purposes.

Setting Performance Expectations. In addition to identifying program components, their associated performance expectations, and the method and data sources used to measure them, measuring program fidelity also entails deciding how to summarize the data from the various data sources. Because the core components of CSC services are stated very generally (e.g., "multidisciplinary team"), programs must determine and adopt performance expectations that map to each core component (e.g., "Teams will have at least X number of FTE psychiatrists/prescribers for every Y number of clients served"). They also will need to identify data sources to measure each performance expectation. Although there is broad agreement about the core components to be measured, FEP programs may choose very different ways to measure them, including, in some instances, splitting or lumping the core components and associated performance expectations (e.g., Lester, Birchwood, Jones-Morris, et al., 2006; Melton, Blea, Hayden-Lewis, et al., 2013; Essock, Nossel, McNamara, 2015; Addington, Norman, Bond, et al., 2016; Hetrick, O'Connor, Stavely, et al., 2017).

The following links provide examples of performance expectations for FEP programs by showing those used in the Practice Guidelines for Oregon Early Assessment and Support Alliance, the RAISE Connection Program, (page 2-29), and the First Episode Psychosis Fidelity Scale (FEPS-FS).

Although there is broad agreement about the core components to be measured, FEP programs may choose very different ways to measure them.

DATA SOURCES AND DATA COLLECTION STRATEGIES FOR MEASURING IMPLEMENTATION FIDELITY

Six primary information sources, alone or in combination, can be used to develop fidelity measures. These information sources are: administrative data; data abstracted from health records; program self-report with spot checks to verify accuracy; data gathered from or confirmed by site visits; data from telephone or video-conferenced interviews (telefidelity); and client interviews.

Administrative data can reliably be used to describe the structural aspects of care.

Administrative data. Administrative data have long been considered the most reliable source of information on the structure of health care services (Mowbray, Holter, Teague, et al., 2003). As electronic health records (EHRs) become more widely used, obtaining clinical data such as the number of visits and type of therapy will become easier. Currently, EHRs can be accessed most easily in large organizations with integrated information systems, such as HMOs or the health care facilities administered by the Department of Veterans Affairs (Rollins, McGrew, Kukla, et al., 2016). Accessing such information from widely dispersed community-based mental health programs operated by multiple providers in multiple states is more challenging—in part because they may employ different types of electronic health records. They also may be funded by different payers, with the associated difficulties that come with summarizing client-level data across different payers. The more local the comparisons to be made, the more likely that administrative data can be used to fill at least some data needs for assessing fidelity, especially assessing structural aspects of CSC programs.

Administrative data may include data collected to meet other reporting requirements, such as the following: billing, data required for all admissions and discharges, human resource data such as the number of staff members and staff full-time equivalents (FTE) on the payroll and assigned to the CSC program, staff job descriptions, and records of staff members' participation in training activities. Programs and program administrators with access to client-level claims data have straightforward, minimally-gameable ways to generate reports that address performance expectations. Access to such data allows programs and program administrators to look at issues such as engagement in services (claims over time), working with families (claims indicating presence of a collateral), team-based intervention (claims from more than one team member), and caseload (number of clients actually receiving services in a given month). Because claims data typically are submitted electronically and are subject to audit, they usually are both timely and accurate.

In many settings, CSC staff members routinely are required to submit service logs that note the client and staff person involved, whether family members were present, and the location of the service (e.g., office versus community). Such service logs can be summarized to measure whether performance expectations are being met. Sites that use standardized forms to monitor

things like weight and medication side effects can use the presence or absence of such forms as indicating whether the team measured weight and asked about medication side effects. Noting the presence or absence of a completed form is more objective, and much less labor-intensive, than reading progress notes.

Programs with access to EHRs may have an automated way to measure performance expectations. They may be able to summarize electronically, for example, the percentage of clients who have had an adequate trial of at least one antipsychotic medication and whether weight is assessed regularly. Such data can have clinical utility by creating exception reports to feed back to the teams that list the team's clients who have not yet had an adequate trial with an antipsychotic or not had weight measured within the reporting timeframe.

With regard to monitoring program staffing, human resource systems often have electronic records noting filled positions over time which can be used to determine whether a team has the appropriate staff in place and whether vacancies are filled promptly. Staff training records may be kept electronically, allowing a program monitor to review whether staff have participated in required training.

Administrative materials such as program brochures, client and family information materials, and curricula for groups such as family or client psychoeducation can be useful in documenting that programs are meeting requirements to provide at least some services after hours, to include family treatment groups, and to be available in emergencies. However, more reliable measures would be those obtained from service data confirming that the services described in the program's brochures or procedural manuals actually were being provided. Such service records could be used to confirm, for example, that teams were meeting with families, that some routine services were occurring outside the regular work day, or that clients were seen during off-hours in urgent situations.

> **Service records can be used to confirm that a program is meeting requirements to provide services after hours, include family treatment groups, and be available in emergencies.**

Data abstraction from health records. Data abstraction from health records, either by program staff or site visitors, is a standard source of data for fidelity assessments (Astrachan, Essock, Kahn, et al., 1995); (Bond, Evans, Salyers, et al., 2000); Wisdom, Knapik, Holley, et al., 2012). This strategy has been widely used and tested for primary care research with high inter rater reliability (Liddy, Wiens, Hogg, 2011). Practical procedures for conducting reliable audits of health records include attending to rater training, designing a chart abstraction template, selecting charts, assessing reliability and validity, and budgeting (Hogg, Johnston, Dahrouge, et al., 2010).

A random selection of particular health records (e.g., all admissions in the past 12 months; all clients who were hospitalized 6-12 months prior to the audit; all clients discharged in the past 12 months) can be reviewed and data abstracted to assess the performance expectation in question. For example, if the performance expectation is that all admissions will include a risk

assessment for suicidality, then one would review a random sample of clients admitted during the period in question to determine whether an assessment of suicidality was documented as part of the clinical intake. If the expectation is that the CSC team will continue to work with clients who are hospitalized, then one would review a random sample of clients who had been hospitalized during the period in question to determine if one or more team members was in contact with the client, or the client's family, while the client was hospitalized.

Program self-report with spot checks to verify accuracy. Some approaches to fidelity measurement require programs to provide data on their performance, with the caveat that their self-reports may be audited to confirm accuracy. Such self-reported data may be, in essence, raw data to be summarized centrally. It may include, for example, client-level data showing, for the time period in question, the number of clients admitted, the medications prescribed, or the number of clients hospitalized. Alternatively, programs may be required to submit summary data on such factors as the case load at the end of the reporting period, or the percentage of clients who are prescribed an antipsychotic drug on the last day of the reporting period. OnTrackNY CSC teams are required to submit data quarterly on each client as well as program-level data (e.g., staff vacancies), and those data are used by OnTrack's central office to assess fidelity (see http://www.ontrackny.org/Resources).

Some programs may have access to EHRs or electronic service logs to aid them in providing this data, while others may need to review a paper record. These self-reported data may be compared directly with performance expectations and/or fed into scaled approaches to measuring fidelity to obtain a scaled score. Such self-reported data have been compared with data collected via site visits and used to scale fidelity to assertive community treatment (ACT) (McGrew, White, Stull, & Wright-Berryman, 2013). In that instance, the investigators concluded that the self-report method could be used to complement more intensive methods of fidelity assessment, using both self-reports and interviews.

Another ACT study looked at costs and acceptability of three methods of completing fidelity scales: site visits, phone interviews, and self-reports (Rollins, Kukla, Salyers, et al., 2017). In that study, program staff members said they preferred site visits over phone interviews or self-reports; however, the authors noted that the cost of site visits was higher than the cost of obtaining data by phone or from self-reports.

Site visits allow program monitors to observe, first-hand, aspects of programs that may be difficult to summarize quantitatively.

Data gathered from or confirmed by site visits. Since at least the early 1990s, payers have used structured site visits as a means of monitoring the quality of behavioral health care and assuring that performance expectations set forth in contracts for behavioral health services are occurring as intended (Astrachan, Essock, Kahn, et al., 1995). Site visits allow program monitors to observe, first-hand, aspects of programs that may be difficult to summarize quantitatively. For example,

CSC programs are expected to have welcoming spaces for youth and young adults. Such an expectation may be difficult to quantify yet easy to observe (in the spirit of, "I know it when I see it").

Site visits also afford the opportunity to review a random selection of charts to estimate the extent to which various performance expectations that cannot be addressed via administrative data are being met. Even entities that rely heavily on performance measures obtained from administrative data (e.g., the Joint Commission's use of CMS's Healthcare Effectiveness Data and Information Set (HEDIS) Measures), may opt to perform site visits for ongoing accreditation/certification purposes. (Essock, Olfson, Hogan, 2015)

A site visit also may allow for face-to-face interviews with managers, clinicians, clients, and families. A site visit can help provide an overall feel for a program and the challenges it faces. It can serve as a vehicle for clinical problem-solving, in addition to measuring performance. A site visit can also support other agendas that might be in the purview of the reviewers, such as reviewing funding issues, addressing inter-organizational relationships, supporting the clinical training of staff members, and creating strategies for quality improvement. Some technical support and assistance programs delivered through third-party contracts conduct on-site fidelity assessments and use the results to support quality assurance and quality improvement.

Semi-structured interview guides that cover the questions relevant to specific staff members can be used to interview managers and line staff to determine whether particular program components are in place. For example, if a program expectation is that members of the CSC team respond to urgent situations after hours, a site visitor might say, "Show me your on-call schedule" or "Show me the last 6 times a client or family member contacted someone on call"). While a program's policy and procedure manual may state that particular program components exist, what exists in theory (in the manual) may not always exist in practice. Site visits can help determine what exists in practice when other, more efficient ways of assessing practice are unavailable. When site visits overlap with training visits, the cost of staff members' time and travel may be offset against a budget that already covered some on-site consultations. For a centralized agency focused primarily on conducting fidelity reviews in a large state, the costs of travel, accommodation, and staff time make site visits expensive.

Site visits also can be very helpful as a means of "spot auditing" a random subset of the data reported by a CSC program (e.g., "You report that 27 of the 32 clients active as of January 1 had an adequate trial of at least one antipsychotic; please show me how you determined that by identifying the 27 and when they had such a trial"). Clearly, site visits can serve many purposes, including staff training, keeping central staff members in touch with the CSC team's real-world challenges, and assessing program fidelity. They also enable the site visitor to spot-check the accuracy of data provided by the CSC program.

Site visits also can serve as primary sources for collecting some or all fidelity data. For example, if teams are expected to work with clients during urgent situations and to provide some routine services outside regular business hours, a site visitor can ask to see the team's on-call system, including when it was used to respond off hours during a time period of the site visitor's choice, and to show what routine services were provided during nights and weekends during the specified time period. Collecting data via site visits requires a clear protocol for data collection and staff members who are trained to implement that protocol. Because the information collected during the site visit is used to characterize routine team functioning, a challenge in creating site visit protocols is to arrive at data that reflect team performance when the site visitor is not present. Addington and colleagues have described their methodology for using site visits to collect fidelity data in the manual for the First Episode Psychosis Fidelity scale (FEPS-FS) https://cumming.ucalgary.ca/feps-fs/.

> **The more automated and routinized the monitoring process, the less disruption there will be when current staff members leave or new ones are hired**

Factors to consider when determining where to get data during a CSC site visit include: the data collection burden (the more a staff member's time is consumed providing data, the less time available to provide clinical services to clients); cost (startup costs, maintenance costs); reliability of information obtained (asking someone a question is easy, but the person's response may not reflect reality); ease of use (each manual link in a system has a way of bogging down); and sustainability (limitations in any of the just-named factors have cumulative effects that can make continuing monitoring more challenging, yet continuing monitoring is essential to characterize current performance).

Monitoring systems need to be able to withstand changes in CSC staff; hence, the more automated and routinized the process, the less disruption there will be when current staff members leave, or new ones are hired. On the other hand, monitoring systems need to be flexible enough to modify data collection and scoring systems when new evidence emerges or when local circumstances change.

Telefidelity interviews. The use of telephone or video conferencing to conduct interviews with CSC staff members (telefidelity interviews) as a substitute for on-site interviews has been investigated and found to be reliable and valid for some purposes. When the Dartmouth Assertive Community Treatment Scale (DACTS) was used to assess a state-wide program of 23 teams (McGrew, Stull, Rollins, et al., 2011), there was a high level of consistency between telephone raters and on-site fidelity raters. Following the initial use of the DACTS, two key changes to the fidelity schedule were made. First, subjective questions were modified to elicit discrete, concrete behaviors. Second, several of the ratings cited in the assessment of the state-wide program were based on program reports of structural components of the program, such as staffing. These principles were applied during the adaptation of the FEPS-FS for the 10 percent Mental Health Block Grant (MHBG) study, which is discussed further below (https://tenpercenteval.samhsa.gov/).

Only clients can address whether, from their perspective, particular performance expectations are being met.

Client interviews. Interviews with clients can be used as primary data sources or to corroborate other sources of information. (Essock, Nossel, McNamara, et al., 2015) While data from CSC teams can provide an enormous amount of information on how a team is functioning, only clients can address whether, from their perspective, particular performance expectations are being met.

For example, the NIMH manual for CSC teams (https://www.nimh.nih.gov/health/topics/schizophrenia/raise/csc-for-fep-manual-ii-implementation-manual_147093.pdf). While staff members may be able to document their participation in shared decision-making training and to report that they use it, a program monitor can be more confident that shared decision-making is occurring routinely if clients have high rates of endorsing statements such as, "Decisions about your treatment were joint decisions between you and the team" (Essock, Nossel, McNamara, 2015). The RAISE Connection Program used such "Ask the client" questions and published the questions and the responses from clients (Essock, Nossel, McNamara, 2015; page 1 of online supplement, https://ps.psychiatryonline.org/doi/suppl/10.1176/appi.ps.201400531/suppl_file/appi.ps.201400531.ds001.pdf).

High rates of endorsement by clients of survey items addressing shared decision-making and responsiveness to clients' concerns allow their CSC teams, and their program monitors, to feel confident that these aspects of the intervention are being carried out as intended. For some survey items, such as "The team helps me figure out my finances," one would not expect a high rate of endorsement, because not all clients need help with their finances. So long as at least some clients endorse this survey item, this indicates that their team does provide the service. Because client surveys are becoming more prevalent, program administrators may be able to insert key questions into routine satisfaction surveys given to clients. A client's signature on a list of goals and objectives on his or her treatment plan may be suggestive of shared decision making, although clients in many settings have been asked to sign treatment plans for many years, including plans to which they may have had little input much less be the result of shared decision-making.

Combinations of the above-noted approaches. The approaches to measuring program performance discussed above suggest multiple means can be used effectively to assess intervention fidelity, alone or in combination with other approaches.

Regardless of the data sources and data collection methods used, the data and corresponding measures need to be reliable, feasible and valid. Reliability means that the data are reproducible (e.g., that independent raters would make similar ratings). Feasibility means that the cost and effort to implement and sustain the approach can be carried out with the available resources. Validity refers to accuracy, that the approach truly measures the construct in question.

Table of common pros and cons for various data-collection strategies for measuring fidelity

METHOD	STRENGTHS	WEAKNESS
Administrative data (claims, EMRs)	There is no need to sample, as data are present for all clients. Costs of data collection are covered by other entities.	Reports must be programmed to extract and summarize the administrative data. Not all monitoring entities have access to administrative data.
Record abstraction	Random samples can produce unbiased estimates of what is being monitored. Data abstraction protocols can be specific enough to be used by non-clinicians with minimal training.	Charts may not include the data being sought. Records may be difficult to access or read. If records must be reviewed on site, travel costs can mount. Data abstraction protocols can be difficult to develop and reliably implement, particularly with regard to maintaining an audit trail.
Staff self-report	This approach is relatively inexpensive.	This strategy entails a risk of gaming in the absence of external checks. Data provided by sites still needs to be aggregated and reported across sites. This approach can be labor-intensive for clinical staff, who must provide data that is detailed enough to accurately assess performance expectations.
Telefidelity (structured telephone interviews)	This strategy can be deployed over large distances. This strategy can be implemented using a small team of interviewers.	Feedback is less personalized. Qualitative aspects of the program (e.g., youth-friendliness) cannot be assessed. This strategy provides less opportunity to evaluate the broader administrative context than on site visits. Context is important for sustaining fidelity to the model. Interviewers must be trained to score reliably. Turnover of trained interviewers may disrupt the ability to conduct interviews.
Site visits	This strategy can provide a forum for clinical strategizing and it can address concerns other than fidelity (e.g., training). This strategy enables the assessment of qualitative aspects of the program (e.g. whether facilities are youth-friendly).	This strategy is expensive. It requires large numbers of experts trained to score reliably. Turnover of trained site visitors may disrupt the ability to conduct site visits. Activities observed at a site visit may not be typical of everyday practice.
Client self-report	This strategy enables the assessment of certain client perceptions of program components that would otherwise be difficult or impossible to assess (e.g. the expectation that treatment decisions will be made jointly by the client and the staff).	This approach measures only a subset of performance expectations. Under this strategy, infrastructure must be created to allow data collection and summarization. Some clients may not be reachable or may decline to participate.

COSTS OF COLLECTING DATA

Regardless of what data sources a program monitor may use to measure fidelity, collecting, summarizing, and reporting data comes with associated costs. Moreover, if the methods chosen to measure fidelity are too labor-intensive or difficult to routinize, they risk failing or becoming unreliable. The list of what people would like to measure can get very long; the longer the list, the more costly to implement and sustain. Therefore, program administrators may wish to ask: What is the minimum set of measures that will capture whether a CSC team is functioning as anticipated? If that minimum list is longer than is feasible to collect and verify throughout a team's operation, one strategy is to require reporting of the full list but to audit only a random subset to verify the accuracy of the data being reported (much as a teacher hands out a list of potential exam questions but only uses a subset on the exam, thereby encouraging the student to prepare for each question). (Essock and Norquist, 1988)

CSC budgets need to include resources for performance monitoring. If you are a state administrator responsible for contracting with CSC sites, it is prudent to include in the contract the data-reporting requirements the sites will be required to meet. It is important that the sites understand that the costs specified in the contract include their costs for providing data, and that performance monitoring will involve viewing and reporting client-specific data.

SCORING DATA TO MAKE INFERENCES ABOUT THE QUALITY OF SERVICES PROVIDED

In addition to noting performance expectations and identifying data sources, fidelity protocols must specify how to score fidelity information in such a way that clear inferences can be made about whether or not CSC programs have implemented the intervention as intended. Programs may opt for simple pass/fail ratings (i.e., met the standard or didn't meet it); a 3-point scale (below expectation/meets expectation/exceeds expectation) (Wisdom, Knapik, Holley, et al., 2012); or a more finely demarcated scale (e.g., FEPS-FS uses a 5-point scale) (Addington, Norman, Bond, et al., 2016).

The FEPS-FS scale has been designed to set an overall score of 80 percent of potential points as acceptable. In the development study for the FEP-FS that led to creation of the scale, this cut-off score differentiated the results of programs that met fidelity standards according to both experts and a separate fidelity assessment process. The results of the ongoing evaluation of the MHBG 10 percent set-aside will provide further information to help establish treatment standards and norms. Two of the FEPS-FS treatment access indicators, the percentage of expected incident cases recruited, and the percentage seen within two weeks, have been incorporated into national standards for providing early psychosis services in England (NICE, 2015).

As a complement or alternative to setting an overall threshold for determining what constitutes a passing score, contracts and/or program administrators may specify particular required components that, if failed, result in an overall failing score. For example, depending on local rules and expectations, a CSC team that has no prescriber as part of the team, or that does not work with families, or that does not provide off-site services may be scored as failing, regardless of how well the team does on other measures.

Setting a pass-fail threshold requires the identification of an acceptable, defensible standard as to what constitutes a minimally adequate performance. In the absence of such a standard, programs may choose to monitor trends in raw scores prior to setting their own thresholds for passing. For example, administrators may be reluctant to set an expectation level for the percentage of clients who have had the team's employment/education specialist meet with client's employers or school staff. Their expectation may simply be that, in each reporting period, this has happened at least once. As data accrues from multiple teams, performance expectations can be tuned accordingly.

MULTIPLE PURPOSES FOR WHICH FIDELITY FINDINGS CAN BE USED AND THE IMPLICATIONS OF EACH

The results of the assessment of fidelity to specified performance expectations can be used for several purposes. Examples include quality improvement (Margolies, Broadway-Wilson, Gregory, et al., 2015), regulatory oversight (Lester, Birchwood, Jones-Morris, et al., 2006), implementation (McHugo, Drake, Whitley, et al., 2007), and research (Bond, 2004).

Quality Improvement. Quality improvement can be defined as "the attainment or process of attaining a new level of performance or quality" (Lowe and Barnett, 1994). Two of the many models for improving quality of care that have been adopted in the U.S. mental health care system include Measurement-Based Quality Improvement (MBQI) and Evidence-Based Practice Implementation (EBPI) (Hermann, Chan, Zazzali, et al., 2006). MBQI is used by many hospitals and health plans and focuses on measuring specific indicators such as wait times and 30-day readmission rates that are useful for assessing a health system. EBPI focuses on the implementation of evidence-based practices and is more often undertaken by state health authorities. It focuses on evidence-based practices such as CSC for first episode psychosis. Both use indicators (Mainz, 2003). An expert committee of The International Early Psychosis Association has published a set of set of key indicators and fidelity scales that can be used for the assessment of first episode psychosis services (Addington, Birchwood, Jones, et al., 2018).

Ideally, the assessment of evidence-based practices relies on the use of a specific set of indicators that define and separate the evidence-based practice from other practices. A fidelity scale is a set of indicators for a specific evidence-based practice that are reliable, valid, feasible and related to outcomes (Bond, Becker, & Drake, 2011). Fidelity measures targeting provider behavior have been incorporated into fidelity scales and have been used for quality improvement of psychiatric rehabilitation programs (Bond, Evans, Salyers, et al., 2000). As measures are developed and validated for assessing the fidelity of CSC services, measuring desired change can be used to improve the quality of the program. For these measures to be effectively used for quality assurance, they must be deemed by staff members to be credible and consistent with the CSC program model.

States may use fidelity measures to help assure that the services they are purchasing are being delivered as intended.

Regulatory oversight. States and other payers may use fidelity measures to help assure that the services they are purchasing are, indeed, being delivered as intended. CSC teams that fail to meet performance expectations may be at risk of losing their funding. For example, Oregon has issued practice guidelines for their early-intervention teams, called the Oregon Early Assessment and Support Alliance (EASA). The guidelines specify, for each program element, a performance target, how adherence to the target is measured, and the data source for the measure (see EASA Reporting Regulations). Similarly, New York State has established OnTrackNY teams to serve people with FEP and it has issued performance expectations describing the required program components, how they are measured, and data sources.

These approaches are consistent with the NIMH recommendations for measuring fidelity contained in Appendix 12 of the NIMH implementation manual for Coordinated Specialty Care for First Episode Psychosis. This document specifies performance expectations for each program component and explains how those performance expectations can be operationalized. Not surprisingly, these performance expectations emphasize performance of the expected processes of care (e.g., that intake assessments be completed within X number of days of referral). In such instances, performance measurement and fidelity measurement can be the same thing—performing the expected processes means implementing the intervention with fidelity to those processes.

States and other payers may opt to require, as a condition of funding, that CSC programs provide data necessary to assess compliance with program expectations, including fidelity measurement. While states and other payers may be able to compute some performance measures without collecting primary data from a program (for example, data on the number of active clients over time or rates of hospitalization may be obtainable from claims data), other data may need to come from the program itself, hence the contractual requirement that programs must provide the required data. This is a "softer" approach than tying specific fiscal bonuses or penalties to achieving particular fidelity targets. Such "pay for performance" or "value-based payment" incentives and penalties have long been a component of managed behavioral health care contracts (Astrachan, Essock, Kahn, et al., 1995).

As CSC programs become more common, payers and program managers will have greater access to performance data, and this information will be helpful in establishing and reviewing performance expectations. Such data also can be used when tying payments to performance. Because nationwide reporting systems in the United States pertaining to mental health outcomes are largely voluntary, measurement specification has largely fallen to accrediting entities. These more global measures (e.g., rates of metabolic screening for clients who are prescribed an antipsychotic drug) may characterize population-level performance. The same data can also be used to create "exception reports" for clients on a CSC team who are in the

measure's denominator but not in the numerator (e.g., individuals prescribed an antipsychotic drug for whom there was no indication of metabolic screening) (Essock, Olfson, Hogan, 2015).

Other contractual elements also may help promote access to high-quality CSC. For example, in its contracts with Medicaid managed care vendors, New York State requires that, when a managed care vendor identifies a member with first episode psychosis, the member be referred to a CSC team meeting the New York State Office of Mental Health's criteria. New York also requires that the managed care vendor have in place a means of monitoring whether or not this occurs (https://www.omh.ny.gov/omhweb/bho/docs/first-episode-psychosis.pdf). The state notes that all OnTrackNY teams meet the state's requirements, thereby encouraging referral to these CSC teams.

Oregon was an early adopter of establishing a network of early intervention services. Oregon's state-wide implementation includes financial support, a technical support unit to provide consultation, provider training, and quality monitoring. It employs locally developed practice guidelines in addition to uniform Fidelity Checklists of performance expectations (Melton, Blea, Hayden-Lewis, et al., 2012). Other states have followed Oregon's example, many of them motivated by the federal government's addition of funds to each state's Mental Health Block Grant for treatment for individuals with FEP. A state survey conducted by the NASMHPD Research Institute, Inc. (NRI) identified 261 FEP programs at varying stages of implementation across the country. In response to the survey, 45 states indicated they were collecting outcome and performance measures for these programs (Lutterman, Kazandjian, 2017). Many of the FEP programs identified the components of CSC that they deliver. The NRI report did not collect information on whether and how program fidelity was being measured.

In England, the National Institute for Health and Care Excellence has issued a guidance for access and wait time standards for early intervention services (Hollis, Kendall, Shiers, 2013; NICE, 2015). The standards are being monitored using a mandatory national on line self-report rating the Early Intervention in Psychosis Network Self-Assessment tool (Royal College of Psychiatrists, 2016a). This requires submission of data on recruitment based on expected population incidence and the time from referral to first assessment. The report also asks for information on the availability of interventions in a NICE-approved care package with the following components: Cognitive behavioral therapy for psychosis, supported employment, care coordination, family interventions, antipsychotic medications, physical health monitoring, weight management, and physical health interventions for smoking.

FEP programs in England are required to submit data to a national data repository, the Mental Health Services Data set. A new draft set of English standards for early intervention

in psychosis services has broadened the earlier standards for access. It describes nearly 300 specific standards covering about 30 domains, ranging from access to care and client referral, to commissioning and financial management (Royal College of Psychiatrists, 2016b). These standards have been categorized into three types:

- **Type 1:** Failure to meet these standards would result in a significant threat to patient safety, rights of dignity and/or would breach the law.
- **Type 2:** Standards that a service would be expected to meet.
- **Type 3:** Standards that are desirable for a service to meet.

In England, administrators or payers may require that a CSC team pass all Type 1 standards to obtain an overall passing score, regardless of the team's performance on other standards.

Implementation. Programs and regulators may measure fidelity to assist in implementing programs. Fidelity measures can serve as benchmarks and targets that can help new programs, and their payers, determine the extent to which the programs are being implemented as intended. Such measures may be implemented in stages depending upon the maturity of the program. For example, fidelity standards about case load may be relevant only after a program has been accepting clients for at least X number of months. Similarly, standards having to do with the percentage of clients who have had an adequate trial of an antipsychotic medication are relevant only once the program has been in operation long enough for clients to accrue such trials.

Problems implementing evidence-based practices in routine clinical practice are common.

Problems in implementing evidence-based practices in routine clinical practice are common (Grol & Grimshaw, 2003). Empirically supported programs may not be implemented at all or in the real world may not deliver the outcomes expected, based on research findings (Greenhalgh, Robert, Macfarlane, et al., 2004). A Consolidated Framework for Implementation Research (CFIR) has been proposed that provides a pragmatic structure for understanding and promoting implementation in health care (Damschroder, Aron, Keith, et al., 2009). Within this framework, the identification of core elements, the flexibility to adapt the service to local conditions, the presence of an agreed-upon evaluation framework, and an economic justification are important ingredients for successful implementation. Fidelity measures play a central role in CSC program implementation by providing descriptions of the core elements, evidence of successful implementation, and targets for improvement where implementation is just starting or the program is in the process of meeting challenges (Schoenwald, Garland, Chapman, et al., 2011) (McHugo, Frake, Whitley, et al., 2007) (Drake, Bond, & Essock, 2009). Repeated assessment of fidelity showed significant improvements over two years following implementation of supported employment programs (Bond, McHugo, Becker, et al., 2008). During implementation of CSC programs, fidelity can be tracked to show improvement over time (Bond, Drake, Becker, et al., 2016). Fidelity tools have been used to assess early implementation of first episode psychosis treatment in New York State (Essock, Nossel, McNamara, et al., 2015) and to confirm implementation in the United Kingdom (Birchwood, Lester, McCarthy, et al., 2014) and Denmark (Melau, Albert, Nordentoft, 2018).

In order to assess implementation of first episode psychosis services funded by the Federal Mental Health Block Grant, the Substance Abuse and Mental Health Services Administration (SAMHSA), the National Institute of Mental Health (NIMH), and the Office of the Assistant Secretary for Planning and Evaluation (ASPE) are supporting the national Mental Health Block Grant (MHBG) 10% Set-Aside Early Intervention Study for addressing first episode psychosis (FEP). (https://tenpercenteval.samhsa.gov/). This 3-year study focuses on how the MHBG 10% set-aside has impacted access to essential services within CSC programs. The study should help policymakers understand how the delivery of CSC services relates to individual client outcomes. The objectives of this study are to:

- Identify and describe CSC program services being offered nationally;
- Assess program fidelity to the CSC model and client outcomes; and
- Assess the local environmental and contextual factors related to the implementation of CSC programs.

The 3-year study began in September 2016 and includes 36 sites from across the United States. Study sites are participating in fidelity assessments and a process evaluation as well as submitting client outcomes up to four times over 18 months. The study also includes a national online survey of all sites that receive MHBG 10% Set-Aside Funds for addressing first episode psychosis. The fidelity to the CSC model is being assessed with the First Episode Psychosis Fidelity Scale, which assesses the 12 CSC domains listed above. Fidelity is assessed using data from three sources; administrative data, such as numbers of admissions and discharges and program staffing; reviews of a random selection of health records; and interviews with program managers and staff, using a semi-structured protocol. The modifications to the scale and data collection methods used to make the scale ratings have not yet been published.

Research. Research studies have long used fidelity measures to confirm adherence to a particular model of psychotherapy (Moncher and Prinz, 1991). They have also been used to assess more complex multi-component programs, such as assertive community treatment (Essock and Kontos, 1995; McGrew, Bond, Dietzen, et al., 1994). A published review of fidelity measures emphasized the importance of establishing the reliability and validity of scales used in both research and in practice (Schoenwald and Garland, 2013).

One limitation of research studies comparing complex programs, such as assertive community treatment programs, with treatment as usual, has been the frequent failure to assess the fidelity of the treatment as usual arm of the research study (Burns, 2009). The NIMH Recovery After an Initial Schizophrenia Episode (RAISE) Early Treatment Program (ETP) study provides an example of the ambiguity that can occur in the absence of measuring the intervention received by participants in both the experimental and control conditions. Participants in the treatment as usual arm of the RAISE-ETP study were found to have a lower-than-expected rate of hospitalization; however, without information on the fidelity to the CSC services received by individuals in each treatment condition, this finding is difficult to interpret.

That challenge can be overcome by assessing the services in both arms of a study using a fidelity measure that uses the same definitions of service components to assess both arms of the study. For example, Essock and Kontos (1995) used administrative data to compare team-based service delivery and intensity of services for clients served by ACT teams, with clients assigned to standard services. Similarly, NCQA and the Joint Commission rely on administrative data to make inferences about the quality of care across program settings, and one can examine whether these measures differ for clients served in various treatment settings (e.g., by FEP teams or usual care). Simon and colleagues have developed an algorithm to identify individuals with FEP using claims data (Simon, Coleman, Yarborough, et al, 2017). Systems that use this algorithm to identify individuals with FEP can examine, for example, the extent to which performance measures reported for NCQA and other purposes differ for individuals with FEP served by specialized FEP services versus treatment as usual.

If one wishes to conduct site visits for the purpose of assessing fidelity, the FEPS-FS was specifically developed with versions that can be used to assess fidelity delivered by both team-based services and treatment as usual treatment arms (Addington, Norman, Bond, et al., 2016). The RAISE-ETP study assessed fidelity to the individual psychosocial treatments using established strategies to record and measure clinicians' adherence to Individual Resiliency Training (IRT) (Browne, Edwards, Penn, et al., 2016; Kane, Robinson Schooler, et al., 2016). Fidelity assessment of an individual clinical practice to this level of detail may not be economically viable in routine practice (Rollins et al., 2017).

VARYING PURPOSES MAY SUGGEST DIFFERING MEASUREMENT APPROACHES

The needs of clinical and health services research have driven the development of many of the well validated and widely used measures of symptoms, functioning, and health services. Changes in social expectations and technology are broadening the application of measures from research to clinical practice and policy. The changes include a focus on patient-centered care (Stewart, 2001). The growing evidence for the benefit of measurement-based care supports the use of reliable and valid measures in clinical practice (Harding, Rush, Arbuckle, et al., 2011). There is also increasing attention to funding and accountability for evidence-based practices in public mental health settings that can be addressed via objective measurements of intervention fidelity (Raghavan, Bright, Shadoin, 2008). The Improving Access to Psychological Therapies program provides public access to outcomes of publicly funded evidence-based practices (Clark, Canvin, Green, et al., 2018).

The purpose for which a measure is being used should inform the choice of the measure, so different measures may be used to assess implementation, clinical outcomes, research outcomes, or quality improvement (Proctor, Silmere, Raghavan, et al., 2011). That said, there are obvious synergies to be had if a measure can meet more than one need.

CONCLUSION

Clients, family members, team leaders, trainers, administrators, and payers all have important interests in knowing whether a given Coordinated Specialty Care team is delivering high-fidelity CSC to its clients. How best to measure fidelity depends on local circumstances and on considerations such as the purpose of the fidelity assessment, regulatory requirements, resources (time and money), availability of administrative data for electronic summarization, and performance expectations to which the CSC teams are being held by their funding sources. Choices of what fidelity assessment strategy to use must always be made within these constraints.

Selecting among the various strategies that are available requires an assessment of the strengths and weaknesses of each approach from the local perspective, how adherence to each expectation is to be measured, and whether and how the results of the fidelity assessment will be shared across CSC sites. States and other payers, likely, will increasingly specify their performance expectations for CSC teams and how adherence to each expected outcome will be measured. Their actions will be informed, in part, by the materials reviewed in this document, as well as the accumulating data on CSC team performance and client outcomes.

REFERENCES

Addington, D., Birchwood, M., Jones, P., Killackey, E., McDaid, D., Melau, M., & Nordentoft, M. (2018). Fidelity scales and performance measures to support implementation and quality assurance for first episode psychosis services. *Early Intervention Psychiatry*. doi:10.1111/eip.12684.

Addington, D.E., Norman, R., Bond, G.R., Sale, T., Melton, R., Mckenzie, E., & Wang, J. (2016). Development and testing of the First-Episode Psychosis Services Fidelity Scale. *Psychiatric Services*, 67(9), 1023-1025.

Astrachan, B.M., Essock, S., Kahn, R., Masi, D., McLean, A.A., & Visotsky H. (1995). The role of a payor advisory board in managed mental health care: The IBM approach. *Administration and Policy in Mental Health and Mental Health Services Research*, 22, 581-595.

Bello, I., Lee, R., Malinovsky, I, Watkins, L.,Nossel, I., Smith, T., Ngo, H., Birnbaum, M., Marino, L., Sederer, L., Radigan, M., Gu, G., Essock, S. & Dixon, L.(2017). OnTrackNY: The development of a coordinated specialty care program for Individuals experiencing early psychosis. *Psychiatric Services*, 1;68(4):318-320. doi: 10.1176/appi.ps.201600512.

Birchwood, M., Lester, H., McCarthy, L., Jones, P., Fowler, D., Amos, T., & Marshall, M. (2014). The UK national evaluation of the development and impact of early intervention services (the National EDEN studies): Study rationale, design and baseline characteristics. *Early Intervention Psychiatry*, 8(1), 59-67. doi:10.1111/eip.12007 [doi].

Bond, G.R. (2004). Supported employment: evidence for an evidence-based practice. *Psychiatric Rehabilitation Journal*, 27(4), 345-359.

Bond, G.R., Becker, D.R., & Drake, R.E. (2011). Measurement of fidelity of implementation of evidence-based practices: Case example of the IPS Fidelity Scale. *Clinical Psychology Science and Practice, 18*(2), 126-141.

Bond, G.R., Drake, R.E., Becker, D.R., Noel, V.A. (2016). The IPS Learning Community: A longitudinal study of sustainment, quality, and outcome. *Psychiatric Services, 67*(8), 864-869. doi:10.1176/appi.ps.201500301.

Bond, G.R., Evans, L., Salyers, M.P., Williams, J., & Kim, H. W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Administration and Policy in Mental Health and Mental Health Services Research, 2*(2), 75-87.

Bond, G.R., McHugo, G.J., Becker, D.R., Rapp, C.A., & Whitley, R. (2008). Fidelity of supported employment: lessons learned from the National Evidence-Based Practice Project. *Psychiatric Rehabilitation Journal, 31*(4), 300-305. doi:J86464U37M737859 [pii];10.2975/31.4.2008.300.305 [doi].

Browne, J., Edwards, A. N., Penn, D. L., Meyer-Kalos, P. S., Gottlieb, J. D., Julian, P., & Kane, J. M. (2016). Factor structure of therapist fidelity to individual resiliency training in the Recovery After an Initial Schizophrenia Episode Early Treatment Program. *Early Intervention Psychiatry*. doi:10.1111/eip.12409.

Burns, T. (2009). End of the road for treatment-as-usual studies? *British Journal of Psychiatry, 195*(1), 5-6. doi:195/1/5 [pii];10.1192/bjp.bp.108.062968 [doi].

Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): An analysis of public data. *Lancet*, 391(10121), 679-686. doi:10.1016/S0140-6736(17)32133-5.

Damschroder, L.J., Aron, D.C., Keith, R.E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science, 4*, 50. doi:1748-5908-4-50 [pii];10.1186/1748-5908-4-50 [doi].

Drake, R.E., Bond, G.R., Essock, S.M. (2009). Implementing evidence-based practices for people with schizophrenia. *Schizophrenia Bulletin, 35*(4), 704-713. doi:sbp041. [pii];10.1093/schbul/sbp041 [doi].

England, M.J., Applebaum, P.S., Bonder, S., & Weisner, C. (2007). Improving the quality of health care for mental and substance-use conditions. The National Academies Press Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders: Washington, DC. Retrieved from: http://www.nap.edu/openbook.php?isbn=0309100445.

Essock, S.M. (2016). What to do when the managed care firm says no. *JAMA-Psychiatry, 73(11), 1109-1110. doi:10.1001/jamapsychiatry.2016.2409.*

Essock, S.M., & Kontos, N.J. (1995). Implementing assertive community treatment teams. *Psychiatric Services*, 46, 679-683.

Essock S.M., & Norquist G.S. (1988). Toward a fairer prospective payment system. *Archives of General Psychiatry*, 45(11):1041-1044.

Essock, S.M., Nossel, I. R., McNamara, K., Bennett, M. E., Buchanan, R. W., Kreyenbuhl, J. A., Mendon, S. J., Goldman, H. H., & Dixon, L. B. (2015). Practical monitoring of treatment fidelity: Examples from a team-based intervention for people with early psychosis. *Psychiatric Services,* 66 (7) 674-676. PMCID: PMC4490109.

Essock, S.M., Olfson M., Hogan M.F., (2015). Current practices for measuring mental health outcomes in the USA: International overview of routine outcome measures in mental health. *International Review of Psychiatry*, 27(4):296-305.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O. (2004). Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Quarterly, 82*(4), 581-629.

Grol, R., & Grimshaw, J. (2003). From best evidence to best practice: Effective implementation of change in patients' care. *Lancet, 362*(9391), 1225-1230. doi:S0140-6736(03)14546-1 [pii];10.1016/S0140-6736(03)14546-1 [doi].

Harding, K. J., Rush, A. J., Arbuckle, M., Trivedi, M. H., & Pincus, H. A. (2011). Measurement-based care in psychiatric practice: A policy framework for implementation. *Journal of Clinical Psychiatry, 72*(8), 1136-1143. doi:10.4088/JCP.10r06282whi.

Heinssen, R., Goldstein, A., & Azrin S., (2014). *Evidence-based treatments for first episode psychosis: Components of coordinated specialty care*. National Institute of Mental Health. Retrieved from: https://www.nimh.nih.gov/health/topics/schizophrenia/raise/nimh-white-paper-csc-for-fep_147096.pdf.

Hermann, R.C., Chan, J.A., Zazzali, J.L.,& Lerner, D. (2006). Aligning measurement-based quality improvement with implementation of evidence-based practices. *Administraion and Policy in Mental Health and Mental Health Services Research, 33*(6), 636-645.

Hetrick, S.E., O'Connor, D.A., Stavely, H., Hughes, F., Pennell, K., Killackey, E., & McGorry, P.D. (2017). Development of an implementation guide to facilitate the roll-out of early intervention services for psychosis. *Early Intervention Psychiatry*. doi:10.1111/eip.12420.

Hogg, W., Johnston, S., Dahrouge, S., Liddy, C., Russell, G.,& Kristjansson, E. (2010). Conducting chart audits in practices-based primary care research. *Canadian Family Physician*, 56(5), 495-496.

Hollis, C., Kendall, T., & Shiers, D. (2013). *Psychosis and schizophrenia in children and young people: Recognition and management*. *National Clinical Guideline,* Number 155. Retrieved from National Institute for Health and Care Excellence (NICE), London, England: https://www.nice.org.uk/guidance/cg155.

Kane, J.M., Robinson, D.G., Schooler, N.R., Mueser, K.T., Penn, D.L., Rosenheck, R.A.,& Heinssen, R. K. (2016). Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the NIMH RAISE Early Treatment Program. *American Journal of Psychiatry, 173*(4), 362-372. doi:10.1176/appi.ajp.2015.15050632 [doi].

Lester H., Birchwood M., Jones-Morris N., England, E., Rogers, H. & Sirvastava, N (2006) *EDEN: Evaluating the development and impact of early intervention services (EISs) in the West Midlands*. Manchester, United Kingdom, National Primary Care Research and Development Centre, Retrieved from: http://www.netscc.ac.uk/hsdr/files/project/SDO_ES_08-1304-042_V01.pdf.

Liddy, C., Wiens, M., & Hogg, W. (2011). Methods to achieve high interrater reliability in data collection from primary care medical records. *Annals of Family Medicine, 9*(1), 57-62. doi:10.1370/afm.1195.

Lowe, H. J., & Barnett, G. (1994). Understanding and using the medical subject headings (mesh) vocabulary to perform literature searches. *Journal of the American Medical Association, 271*(14), 1103-1108. doi:10.1001/jama.1994.03510380059038.

Lutterman, T., & Kazandjian, M. (2017). Snapshot of state plans for using the community mental helath block grant ten percent set-aside to address first episode psychosis. Retrieved from: https://www.nasmhpd.org/sites/default/files/Snapshot_of_State_Plans-Rev5._508%20compliant.pdf.

Mainz, J. (2003). Defining and classifying clinical indicators for quality improvement. *International Journal for Quality in Health Care, 15*(6), 523-530. doi:10.1093/intqhc/mzg081.

Margolies, P.J., Broadway-Wilson, K., Gregory, R., Jewell, T.C., Scannevin, G., Jr., Myers, R.W., & Dixon, L. B. (2015). Use of learning collaboratives by the Center for Practice Innovations to bring IPS to scale in New York State. *Psychiatric Services, 66*(1), 4-6. doi:1917350 [pii];10.1176/appi.ps.201400383 [doi].

McGrew, J.H., Bond, G.R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of mental health program model. *Journal of Consulting and Clinical Psychology, 62*(4), 670-678.

McGrew, J.H., Stull, L.G., Rollins, A.L., Salyers, M.P., & Hicks, L.J. (2011). A comparison of phone-based and on-site assessment of fidelity for assertive community treatment in Indiana. *Psychiatric Services*, 62(6), 670-674. doi:62/6/670 [pii];10.1176/appi.ps.62.6.670 [doi].

McGrew, J. H., White, L.M., Stull, L.G., & Wright-Berryman, J. (2013). A comparison of self-reported and phone-administered methods of ACT fidelity assessment: A pilot study in Indiana. *Psychiatric Services*, *64*(3), 272-276. doi:1555236 [pii];10.1176/appi.ps.001252012 [doi].

McHugo, G.J., Drake, R.E., Whitley, R., Bond, G.R., Campbell, K., Rapp, C.A., & Finnerty, M.T. (2007). Fidelity outcomes in the National Implementing Evidence-Based Practices Project. *Psychiatric Services, 58*(10), 1279-1284. doi:58/10/1279 [pii];10.1176/appi.ps.58.10.1279 [doi].

Melau, M., Albert, N., & Nordentoft, M. (2018). Programme fidelity of specialized early intervention in Denmark. *Early Intervention Psychiatry*. doi:10.1111/eip.12549.

Melton, R., Blea, P., Hayden-Lewis, K.A., Penkin, A., Sale, T., & Sisko, A. (2013). *Practice Guidelines for Oregon Early Assessment* and Support Alliance (EASA) (1). Retrieved from Oregon: http://www.easacommunity.org/PDF/Practice%20Guidelines%202013.pdf.

Moncher, F.J.,& Prinz, R.J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266.

Mowbray, C.T., Holter, M.C., Teague, G.B., & Bybee, D. (2003). Fidelity criteria: Development measurement and validation. *American Journal of Evaluation, 24*(3), 315-340.

National Institute for Health and Care Excellence (NICE). (2015). *Psychosis and schizophrenia in adults. Quality Statement 1: Referral to early intervention in psychosis services*. (QS 80). London, England. Retrieved from: https://www.nice.org.uk/guidance/qs80/chapter/quality-statement-1-referral-to-early-intervention-in-psychosis-services.

National Institute of Mental Health Coordinated Specialty Care Implementation Manual II. Retrieved from: https://www.nimh.nih.gov/health/topics/schizophrenia/raise/csc-for-fep-manual-ii-implementation-manual_147093.pdf.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65-76. doi:10.1007/s10488-010-0319-7.

Raghavan, R., Bright, C.L., Shadoin, A. L. (2008). Toward a policy ecology of implementation of evidence-based practices in public mental health settings. *Implementation Science, 3*, 26.

Rollins, A.L., Kukla, M., Salyers, M.P., McGrew, J.H., Flanagan, M. E., Leslie, D.L., & McGuire, A.B. (2017). Comparing the costs and acceptability of three fidelity assessment methods for Assertive Community Treatment. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(5), 810-816. doi:10.1007/s10488-016-0785-7 [doi];10.1007/s10488-016-0785-7 [pii].

Rollins, A.L., McGrew, J.H., Kukla, M., McGuire, A.B., Flanagan, M.E., Hunt, M.G., & Salyers, M.P. (2016). Comparison of Assertive Community Treatment fidelity assessment methods: Reliability and validity. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(2), 157-167. doi:10.1007/s10488-015-0641-1 [doi];10.1007/s10488-015-0641-1 [pii].

Royal College of Psychiatrists. (2016a). *Early Intervention in Psychosis Network Self-Assessment tool guidance*. College Centre for Quality Improvement (CCQI): London, England, 1-18. Retrieved from: www.rcpsych.ac.uk/pdf/EIPN Self Assessment Guidance.pdf.

Royal College of Psychiatrists. (2016b). *Standards for Early Intervention in Psychosis Services*. London, England. Retrieved from. www.rcpsych.ac.uk/pdf/Standards%20for%20Early%20Intervention%20in%20Psychosis%20Services%20V6%206.pdf.

Schoenwald, S.K., & Garland, A.F. (2013). A review of treatment adherence measurement methods. *Psychological Assessment, 25*(1), 146-156. doi:2012-21628-001 [pii];10.1037/a0029715 [doi].

Schoenwald, S.K., Garland, A.F., Chapman, J.E., Frazier, S.L., Sheidow, A.J.,& Southam-Gerow, M.A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 32-43. doi:10.1007/s10488-010-0321-0 [doi].

Simon, G.E., Coleman, K. J., Yarborough, B. J., Operskalski, B., Stewart, C., Hunkeler, E., & Beck. A. (2017). Incidence and presentation of first-episode psychosis in a population-based sample." *Psychiatric Services*, 68(5), 456–461. http://doi.org/10.1176/appi.ps.201600257).

Stewart, M. (2001). Towards a global definition of patient centred care. *British Medical Journal, 322*(7284), 444-445.

Wisdom, J.P., Knapik, S., Holley, M.W., Van Bramer, J., Sederer, L.I., Essock, S.M. (2012). New York's outpatient mental health clinic licensing reform: Using tracer methodology to improve service quality. *Psychiatric Services*, 63(5):418-420.

RECOMMENDED CITATION

Essock, S. M. and Addington, D. E. (2018). Coordinated Specialty Care for People with First Episode Psychosis: Assessing Fidelity to the Model. Alexandria, VA: National Association of State Mental Health Program Directors.

ACKNOWLEDGMENTS

This report was prepared for the Substance Abuse and Mental Health Services Administration (SAMHSA) under Contract No. HHSS283201200002I/Task Order No. HHSS28342002T with SAMHSA, U.S. Department of Health and Human Services (HHS). Jeanette Miller served as the Contracting Officer Representative. Monique Browning served as the Task Lead.

DISCLAIMER

The views, opinions, and content of this publication are those of the authors and do not necessarily reflect the views, opinions, or policies of SAMHSA or HHS. Nothing in this document constitutes a direct or indirect endorsement by SAMHSA or HHS of any non-federal entity's products, services, or policies, and any reference to a non-federal entity's products, services, or policies should not be construed as such.