Provider-Associated Measurement Error in Routine Outcome Monitoring in Community Mental Health

2 authors, including:

Emily Treichler
VA San Diego Healthcare System
25 PUBLICATIONS   **76** CITATIONS

Some of the authors of this publication are also working on these related projects:

Collaborative Decision Skills Training View project

ORIGINAL ARTICLE

CrossMark

Provider-Associated Measurement Error in Routine Outcome Monitoring in Community Mental Health

Emily B. H. Treichler[1] · William D. Spaulding[2]

Abstract
Despite the strengths of routine outcome monitoring (ROM) in community mental health settings, there are a number of barriers to effective implementation of ROM, including measurement error due to provider factors (e.g., training level) and non-target client factors (i.e., client characteristics which have no meaningful relationship to the outcome of interest). In this study, ROM data from 80 client-provider dyads were examined for sources of variance due to provider factors and non-target client factors. Results indicated that provider factors and non-target client factors accounted for between 9.6 and 54% of the variance in the ROM measures. Our findings supported past research that provider characteristics impact ROM, and added the novel finding that client gender, age, diagnosis, and cognition also impact ROM. Methods to increase accuracy and utility of ROM in community mental health are discussed.

Keywords Routine outcome monitoring · Community mental health · Serious mental illness · Measurement

Introduction

Ongoing assessment of client functioning and outcomes across domains is a key part of best practices for the care of people with serious mental illness[1] (SMI; APA Presidential Task Force on Evidence-Based Practice 2006; Drake et al. 2009). The intention of assessment is to increase the systematic nature of care in real world settings, improving treatment choices based on real-time client progress as well as enhancing the connection between research and clinical care (Gilbody et al. 2002a). In this paradigm, all clients receiving services in a given mental health agency receive ongoing assessment, and their services are adjusted based on the results. Using this kind of system in community settings theoretically results in the most accurate information about

preferred treatments, best outcomes, and cost-benefit ratios, leading administrators to be able to make informed decisions about allocation of funding, beds, and other resources (Wampold 2015; Wennberg 1990). This ideally results in decreased cost and improved standard of care in community settings.

The most common approach to implementing ongoing assessment in practice is the use of standardized assessments given routinely in clinical settings [routine outcome monitoring (ROM)], which can be used to create individualized feedback about client progress and assist providers in making treatment decisions (Hatfield and Ogles 2007). This could use complex models, like informatics systems (Spaulding and Deogun 2011), or simpler models applying clinical scales already in use in community settings (Amble et al. 2015; Gilbody et al. 2001. A wide array of research generally indicates that ROM is feasible, useful for diverse service questions, and has wide utility (Lambert 2003; Rosenheck et al. 1999; Wampold 2015; Wolpert et al. 2012). ROM appears to be particularly adept at identifying at risk clients and preventing deterioration (e.g., Boswell et al.

✉ Emily B. H. Treichler
emily.treichler@gmail.com

1   VA San Diego Healthcare System, VISN 22 MIRECC & University of California-San Diego, 3350 La Jolla Village Drive, San Diego, CA 92161, USA

2   University of Nebraska-Lincoln, Lincoln, NE, USA

---

[1] *Serious mental illness* here refers to psychotic spectrum disorders that cause significant impairment in functioning, typically including schizophrenia spectrum disorders and mood disorders with psychotic features, as well as occasionally including trauma-induced disorders and obsessive–compulsive disorders depending on presentation and functional impact.

🙋 Springer

2015; de Jong et al. 2014). These promising indicators have led some countries, like Australia, to mandate use of ROM (Burgess et al. 2015)

However, there are still many concerns about implementing ROM effectively (e.g., Hoenders et al. 2014; Miller et al. 2015; Roe et al. 2015a). These concerns are primarily focused on barriers to effective implementation, including uptake, appropriate assessment use, and accurate interpretation and decision-making based on assessment findings. One systematic review concluded that the number of confounds present in ROM studies precluded ability to make any clinically significant interpretation (Gilbody et al. 2002a). Given these difficulties with generalization and interpretation, the utility of ROM remains unclear.

Measurement error significantly contributes to problems in ROM implementation effectiveness (e.g., Gilbody et al. 2002a). Some have suggested that assessment measures currently implemented are not intended for individual interpretation, but rather population norms, and thus are unsuitable for clinical use (Hoenders et al. 2014; McHorney and Tarlov 1994). Additionally, assessments are implemented at low rates by providers in community settings, and when they are implemented, are often implemented incorrectly, their results misunderstood and unlikely to be acted upon (Gilbody et al. 2002b). If current methods lack reliability and validity, then the purpose of ROM is defeated.

Sources of Measurement Error in ROM

This study focuses on two possible sources of measurement error, with the intention of specifically describing how providers using ROM may contribute to methodological problems. These provider-related sources were chosen because of the current discrepancy between literature supporting ROM and difficulties in implementation of ROM. This gap partially stems from provider-level difficulties, like incorrect implementation and interpretation, as discussed above. Increased understanding will allow future studies to improve ROM via accommodation or amelioration of these sources of error.

The first of these sources is provider factors. Even in successful implementations of ROM, provider resistance to ROM has the potential to impair success (e.g., Roe et al. 2015b). Providers often perceive ROM as being of low value generally, and report that it does not impact their treatment planning (Hatfield and Ogles 2007; Trauer et al. 2009). Use of ROM is low among both psychologists (Hatfield and Ogles 2004; Ionita and Fitzgerald 2014) and psychiatrists (Gilbody et al. 2002c), partly due to low buy-in. Trauer et al. (2009) suggested that provider attitudes may result in less accurate data even when ROM is completed, because providers are not invested in ROM and agency administrations do not allot significant time for its completion. This hypothesis

is supported by findings that case managers spent more time on activities they believed were effective (Young et al. 1998) and clinicians who reported more positive attitudes towards evidence-based assessment were more likely to use it (Jensen-Doss and Hawley 2010).

Additionally, up to 50% of variance in provider-rated assessment is due to provider-related factors (Baldwin and Imel 2013; Boswell et al. 2016). It has been suggested that clinicians are necessarily biased participants of their clinical interventions and thus unable to assess objectively (Bilsker and Goldner 2002), which may explain inaccurate ratings. Providers may also fear negative repercussions from employers, colleagues, or clients if outcome monitoring is applied to job performance ratings (Ionita et al. 2016). Acknowledgement of these biases is why providers generally are excluded from completing these types of measures in rigorous clinical trials (Boswell et al. 2016), but similar protocols are infeasible for most community settings.

The second of these sources of measurement error is non-target client factors. Client characteristics beyond those targeted by an assessment may impact the results of ROM, typically due to bias among assessors. There is little research available on these sources of error at present. However, one study (Cradock et al. 2001) found that among people diagnosed with schizophrenia spectrum disorders, clients' records contained significantly fewer symptoms than were observed or reported during clinical interviews. Clients with longer duration of illness, who had poorer insight, or were recorded as being "noncompliant" were less likely to have symptoms recorded, and black clients and clients recorded as being "noncompliant" were less likely to have side effects recorded, suggesting that these characteristics may influence assessor judgment. Given the significant bias introduced by non-blinded clinicians on research projects (Hróbjartsson et al. 2013) it is likely that there is also significant bias introduced by non-blinded clinicians in community mental health settings.

Purpose of Study

Although ROM is an important method, there are significant difficulties in accurately and effectively implementing ROM procedures in community settings. As outlined above, there are several possible sources for these difficulties, including provider motivation and attitudes, provider or institutional bias, and non-target client characteristic confounds or biases. Compelling evidence suggests that these issues are associated with measurement error in ROM. This study seeks to examine variance in two currently used ROM measures, and detect whether this variance is due to the effect of provider factors and/or non-target client factors. Significant variance due to non-target factors will indicate the presence of measurement error.

Table 1 Client descriptive statistics

Variable	Mean	SD
Age (years)	47.85	11.34
Education (years)	12.51	2.39

Variable	n	%
Gender		
Female	30	37.50
Ethnicity		
European American	71	88.75
Diagnosis		
Psychotic disorder	33	41.30
Mood disorder	41	51.30
Anxiety disorder	19	23.80
Substance use disorder	23	28.80
Personality disorder	16	20.00

Three general hypotheses are postulated for this study: (1), that provider factors will independently contribute to ROM ratings; (2), that non-target client factors will independently contribute to ROM ratings; and (3) that client and provider variables will interact with each other to contribute to ROM ratings.

Methods

The Institutional Review Board at the University of Nebraska-Lincoln approved this study.

Participants

The archival database files of 80 adult clients with SMI participating in services at a publicly funded community mental health center (CMHC) in the Midwest were used for this study. Participants must have completed a neurocognitive assessment battery within 6 months of a community functioning assessment in order to be included. There were no other inclusion or exclusion criteria. See Table 1 for client descriptive statistics.

Seventeen providers currently working at the CMHC as direct care providers also participated in this study. Providers volunteered to complete a demographic survey about themselves and their work history for the study. Of the providers that completed the survey, 12 had completed assessments of clients in the database and were included in the final analyses. These providers previously rated the client participants using the measures described below as a regular clinical task associated with the archival database. Providers received training on completing these measures when the measures were first implemented or during new provider orientation, if the provider began working after the measures were implemented. They also received ongoing supervision on use of these measures, and feedback after completing the measures to facilitate integration of the results into their everyday clinical practice. See Table 2 for provider descriptive statistics.

Measures

Provider variables were used as predictor variables, including demographic variables of age, gender, and years of education. Additionally, variables intended to measure work experience were measured using self-report, including area of training, certification, position at CMHC, and years of experience in mental health.

Non-target client variables were also used as predictor variables. First, demographic variables were used, including age, gender, and years of education. Second, diagnostic information was used. Third, neurocognitive data were used. These data include the Controlled Oral Word Associated Test (COWA-FAS; Spreen and Strauss 1998), Letter Number Sequencing (LNS; Wechsler 2008), Trails A & B (Reitan and Wolfson 1985) and the Wide-Range Achievement Test (WRAT-R2; Jastak et al. 1993). Fourth, social cognitive data was used. These data include Hinting Tasks (Corcoran et al. 1995) and the Inventory of Competence and Control Beliefs (FKK: Krampen 1991).

Due to collinearity among predictor variables, two factors were created to capture neurocognition and provider experience, respectively. The neurocognition factor included the COWA-FAS, LNS, and Trails A & B, with 53.033% total variance accounted for. The provider experience factor included age, number of years working at the CMHC

Table 2 Provider descriptive statistics

Variable	Mean	SD
Age (years)	42.94	13.87
Experience at CMHC (years)	13.27	12.38
Experience in mental health (years)	19.83	13.66
Variable	n	%
Gender		
Female	11	61.20
Ethnicity		
European American	17	94.40
Marital status		
Married	9	50.00
Single	7	39.00
Highest level of education		
Bachelor's level degree	9	50.00
Master's level degree	4	22.22

specifically, and total number of years working in mental health, with 92.373% of the variance accounted for.

Two ROM measures were included as the dependent variables in this study: the Service Engagement Scale (SES; Tait et al. 2002) and the Multnomah Community Ability Scale (MCAS; Barker et al. 1994). The SES measures engagement in community-based mental health services, and the MCAS measures the community functioning level of people with SMI. These measures were chosen because these outcome domains are commonly used in ROM (e.g., Boswell et al. 2016), and are important outcome indicators in community mental health. At this agency, these two measures were used to (a) identify clients at risk of disengagement or relapse; (b) monitor functional improvement; and (c) identify priorities for treatment. A previous confirmatory factor analysis (B. Zolnikov, J. Peer, N. Mosier, personal communication, February 5, 2011) performed on these two scales yielded a seven-factor model: (1) Availability for Treatment, (2) Help-Seeking, (3) Treatment Compliance, (4) Treatment Participation, (5) Mood Disturbance, (6) Independent Functioning, (7) Social Skills. These factors will serve as the dependent variables for the study.

Procedure

Provider members who consented to participating completed the provider questionnaire during work hours. Provider members were informed of the use of archival data including their provider identification numbers in the study.

Archival client data for this study were gathered from a ROM database maintained at the CMHC. This database divides types of data collected into several categories. Only demographics, social cognitive test results, neurocognitive test results, and provider ratings of the MCAS and SES were extracted for this study.

Data Collation Procedure

Once these data were collected for the study, data were combined for data analysis. The demographic, social cognitive, and neurocognitive datasets were combined into one client assessment dataset. Then the provider data was matched to the provider ratings data using the provider identification numbers to create a provider dataset. Provider and client pairs were treated as dyads (e.g., Kenny 1996); therefore one set of provider data could match to multiple sets of provider ratings data to create multiple distinct dyadic units if provider members completed assessments for multiple clients. The client assessment dataset and provider dataset were combined by selecting for only those sets of client data that included neurocognitive and social cognitive assessment data within 6 months of a provider rating data entry. If there were multiple provider ratings for a particular client with assessment data within 6 months, the more recent ratings was chosen. If two sets of assessment data were within 6 months of a provider rating, then the one closer to the rating date was chosen. Data collation was completed after the data set only included data matching this criteria and only one set per provider/client dyad.

Data Transformation Procedure

Prior to data analysis, the data were checked for skewness and outliers. Outliers were winsorized, and if winsorizing did not resolve skewness, than the data was transformed using

Table 3 Parameter β weights for provider variables

	Provider experience (factor)	Provider education (years)	Provider degree	Provider position
Parameter β weights				
Availability for Treatment	–	.339**	− 1.024***	.927***
Treatment Participation	− .267*	–	–	–
Treatment Compliance	–	–	–	–
Help Seeking	− .369*	–	–	− .459**
Mood Disturbance	–	–	–	–
Social Skills	–	.270*	− .393**	–
Independent Functioning	–	–	–	–

*p < .05; **p < .01; ***p < .001

the square root transformation. No variables required further transformation to resolve skewness. Data were centered at meaningful points to increase interpretability of the data and to decrease collinearity.

Six items were deleted or combined to improve fit and increase power. Race in both clients and provider was removed due to low variability (90.8% of clients and 92.5% of provider reported being white). Mood disorder diagnosis and psychotic disorder diagnosis was combined into a major mental illness disorder diagnosis variable to increase clarity of the construct and manage collinearity between the two existing variables. Certification among provider was removed due to low response rate (81.3% of dyads were missing this item without clarity as to whether responders did not have certifications or simply left that item blank). Prior work experience in medical setting was removed due to low variability (only 7.5% of provider reported prior medical setting experience).

Analytic Procedure

A series of stepwise regressions was conducted to determine the differential impact of client and provider predictors on perceived functioning across the seven functional areas. Client predictors were divided into demographic variables, which included age, gender, years of education and diagnosis, and assessment data, which included neurocognitive and social cognitive data. Provider predictors were kept together. Each set of variables was entered together into a stepwise multiple regression with client demographic variables as the first step, client assessment variables as the second step, and provider variables as the final step. When more than one variable was significant in the complete model, another model including interactions was conducted as a fourth step. Interactions were calculated using the product of the predictors.

Results

Please see Tables 3, 4 and 5 for summaries of parameter estimates.

For *Availability for Treatment*, the final model found that fewer years of provider education ($\beta = .339$, $t = 2.954$, $p = .005$), provider degrees in psychology in comparison to social work ($\beta = -1.024$, $t = -6.180$, $p < .001$), and providers working as mental health specialists in comparison to program coordinators ($\beta = .927$, $t = 5.639$, $p < .001$), was associated with higher perceived availability for treatment $R^2 = .477$, $F(3,45) = 13.698$, $Mse = .224$, $p < .001$.

For *Treatment Participation*, the final model found that male client gender ($\beta = .282$, $t = 2.062$, $p = .045$), client personality disorder diagnosis ($\beta = -.373$, $t = -2.901$, $p = .006$), higher scores on the Externality subscale of the FKK ($\beta = -.329$, $t = -2.598$, $p = .013$), and higher provider scores on the experience factor ($\beta = -.267$, $t = -2.039$, $p = .047$) was associated with higher perceived treatment participation, $R^2 = .340$, $F(4,44) = 5.672$, $Mse = 1.886$, $p = .001$.

For *Treatment Compliance*, the final model found that older clients were perceived to have better treatment compliance, ($\beta = -.310$, $t = -2.238$, $p = .030$), $R^2 = .096$, $F(1,47) = 5.008$, $Mse = .146$, $p = .030$.

For *Help Seeking*, the final model found that more provider experience factor ($\beta = -.369$, $t = -2.493$, $p = .016$) and provider working as case managers in comparison with mental health specialists ($\beta = -.459$, $t = -3.103$, $p = .003$) were associated with higher help seeking skills, $R^2 = .195$, $F(2,46) = 5.576$, $Mse = .966$, $p = .007$.
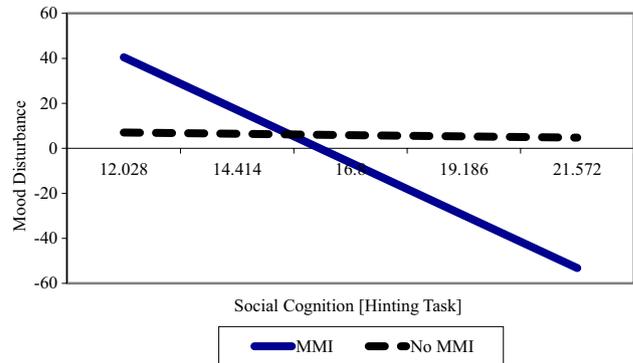
For *Mood Disturbance*, the final model found that client SMI diagnosis ($\beta = 1.486$, $t = 6.650$, $p < .001$), clients with lower scores on the Hinting Task ($\beta = -16.721$, $t = -5.948$, $p < .001$), and a positive and significant weight for the interaction between SMI diagnosis and Hinting Task scores ($\beta = 16.334$, $t = 5.836$, $p < .001$), was associated with higher perceived mood disturbance $R^2 = .540$, $F(3, 45) = 17.603$, $Mse = .964$, $p < .001$. Please see Fig. 1 for a depiction of this interaction.

Table 4 Parameter β weights for client variables

	Client gender	Client age	Client education	Client SMI	Client PD	Client neurocognition (factor)	FKK-externality	Hinting task	Hinting task X SMI
Parameter β weights									
Availability for Treatment	.282*	–	–	–	–	–	–	–	–
Treatment Participation	–	–	–	–	– .373**	–	– .329*	–	–
Treatment Compliance	–	– .310*	–	–	–	–	–	–	–
Help Seeking	–	–	–	–	–	–	–	–	–
Mood Disturbance	–	–	1.486***	–	–	–	–	– 16.721***	16.334***
Social Skills	– .328*	–	–	–	–	–	–	–	–
Independent Functioning	–	–	– .260	–	–	– .345*	–	–	–

*$p < .05$; **$p < .01$; ***$p < .001$

Table 5 Parameter estimates

	R^2	F	p
Availability for Treatment	.477	13.698	< .001
Treatment Participation	.340	5.672	.011
Treatment Compliance	.096	5.008	.030
Help Seeking	.185	5.576	.007
Mood Disturbance	.540	17.603	< .001
Social Skills	.300	6.423	.001
Independent Functioning	.219	6.445	.003

Fig. 1 Interaction between hinting task and major mental illness on mood disturbance

For *Social Skills*, the final model found that female client gender (β = − .328, t = − 2.627, p = .012), provider reporting degrees in other human or applied services in comparison to social work (β = − .393, t = − 3.133, p = .003), and fewer years of provider education (β = .270, t = 2.150, p = .037) were associated with higher perceived social skills, R^2 = .300, F (3, 45) = 6.423, Mse = 1.658, p = .001.

For *Independent Functioning*, the final model found that better client scores on the neurocognitive factor (β = − .345, t = − 2.605, p = .012) significantly contributed to and more client years of education (β = − .260, t = − 1.966, p > .05) insignificantly contributed to perception of greater independent functioning, R^2 = .219, F (2,46) = 6.445, Mse = 4.932, p = .003.

Discussion

Our hypotheses were generally supported; both provider and non-target client factors accounted for variance in ROM scores. We did not find evidence that provider and non-target client factors interacted with each other. Provider background, including type of training, years of training, years of experience, and current position, impacted the way provider rated clients on the Help Seeking, Availability

for Treatment, Treatment Participation, and Social Skills factors. The providers in this study were highly experienced, with an average of nearly 20 years working in mental health. The impact of provider experience of training varied by assessment domain; for example, providers with fewer years of education perceived clients as being more available for treatment, but providers with more years of experience perceived clients as more participatory and more appropriately help-seeking. It's clear that provider training, experience, and position held impacts perception of client behavior, particularly treatment engagement behaviors. However, this study cannot answer questions about which providers might be most accurate. Current research is mixed on this topic (e.g., Garb 1989; Hill et al. 2017; Young et al. 1998), indicating an area for future directions. In any case, ROM is often implemented across diverse providers, and so it is essential that these kinds of measurement error are minimized.

Additionally, there were multiple associations between non-target client factors and functional ratings. It is arguable that some of these associations are accurate depictions of these relationships rather than due to measurement error; for example, other research has supported that older clients tend to have better treatment adherence (e.g., Lang et al. 2010). However, some of these associations are likely to be impacted by bias. For example, the finding that men were rated as having greater treatment participation may be due to gender norms leading women in mental health settings to be assigned "difficult," "irresponsible," or "attention seeking" labels more frequently than men (e.g., Carli 1999; Ussher 2013), particularly given that other evidence indicates that men and women with SMI have statistically equivalent treatment participation (e.g., Sajatovic et al. 2011). This evidence of gender bias in staff-measured treatment participation is a novel, and particularly impactful, finding in this study. This indicates that the known bias against women in mental health may impact diagnostic decisions as well as ongoing assessment of female clients. This likely skews perception of progress and current treatment needs by not only the provider making the ratings, but also the larger system, which often depends on assigned staff and ROM results to identify treatment targets. This may help explain why more female clients have unmet treatment needs and are less likely to receive appropriate specialty mental health services, in comparison to men (e.g., Bijl and Ravelli 2000; Wang et al. 2005).

Finally, provider ratings of mood disturbance among people diagnosed with a major mental illness (i.e., schizophrenia spectrum disorder or severe mood disorder) did not vary based on social cognition, while for people with other diagnoses, better social cognition was associated with better provider-rated mood. It may be that providers assessed mood disturbance based on affect during appointments, rather than client report of mood (a common difficulty; cf. Serby 2003). If this is so, then providers may have perceived social cognition impairments and/or significant negative symptoms as indicative of neutral mood (e.g., associating neutral or blunted facial affect with neutral mood), decreasing the likelihood they would perceive people with major mental illness as having noticeably high or low mood, given the high incidence of negative symptoms in that population (e.g., Fusar-Poli et al. 2015; Reddy et al. 2015).

Implications and Recommendations

Given our findings, it appears that the MCAS and SES were not being implemented suitably in the studied agency. The R^2 in this study range from .096 to .540, with five of the seven factors having at least 20% of their variance explained by staff and/or non-target client variables. This indicates that the results of these assessments do not sufficiently measure their targets. Some of this error is likely due to provider training and supervision issues (e.g., drift, discussed below), but given the magnitude of some of the R^2s, we posit that the assessments themselves are in need of revision in order to facilitate accurate assessment. ROM is intended to lead to improved clinical decision making, resulting in better treatment outcomes. Assessments that include significant error are unlikely to result in these outcomes.

We focused on these issues with the intention of understanding sources of measurement error in ROM. Greater understanding of these sources can elucidate the accuracy of ROM in typical community settings with a SMI population and future steps necessary to facilitate improve ROM application to increase interpretability and utility. Given these findings in context with previous findings, it appears that ROM varies in accuracy depending on assessment domains targeted.

Based on our findings, we recommend that measures used for ROM focus on the most objective and quantifiable measures of outcomes to decrease error. The factors with the least measurement error used in our study are also the measures that are currently the most quantified, and perhaps the easiest to quantify. For instance, Independent Functioning contains no significant sources of measurement error from the variables used in this study. The questions used for Independent Functioning are also comparably more clearly quantified, including items about independent money management and completing activities of daily living. Conversely, Treatment Participation includes questions about the client's level of insight and active participation. The former items are easier to understand and quantify, whereas the latter may be more difficult to conceptualize and assign a Likert scale number to.

Instead, the use of specific behaviors or goals may allow mental health practitioners to monitor these areas of interest

while decreasing the amount of measurement error impacting scores. For instance, instead of using terms like "active participation," which allows for any definition depending on what a provider member or agency understands to be "active" and "participating," a future iteration of the MCAS and SES could monitor client initiated appointments or treatment plan changes, number of appointments attended out of those scheduled, or percentage of goals completed.

Our study also underlines the need for appropriate provider training to increase accuracy and fidelity of ROM implementation (Lyon et al. 2015). Although providers at the agency in this study were trained on these two measures and given ongoing supervision and feedback, this training was not created by the authors of the MCAS and SES and did not include testing to ensure understanding or monitoring to ensure fidelity. Increasing the structure of the training may decrease the impact of prior provider training and education on response styles, because each provider member would have a consistent procedure to follow. Additionally, using baseline and ongoing fidelity testing to ensure that provider are responding accurately and similarly can decrease measurement error due to confusion (i.e., affect versus mood), bias, or drift. Increasing objectivity and specificity of assessment items, as noted above, would also decrease amount of provider training needed to complete measures accurately.

We also recommend continued pursuit of improvement to ROM based on other studies' findings. Specifically, implementation of ROM should prioritize ease of use and clarity of purpose in order to increase buy-in to ROM among staff and accuracy of measurement (Ionita and Fitzpatrick 2014; Kotte et al. 2016). Any measures completed frequently should be short as short as possible and clearly applicable to provider and client needs. The ROM completion procedure should be quick and easy for staff to integrate into their daily schedule. For example, integrating assessments into electronic medical records systems (Jensen et al. 2015) and use of machine learning (Wampold 2015) to automatize assessment both have significant promise to increase ease of use. Finally, use of client-reported measures would decrease provider-related error. Client self-assessments can be completed with provider support without adding significant bias (Gelkopf et al. 2015). However, client-reported outcome measures have their own set of difficulties and research has yet to elucidate a clear benefit to its use (Kendrick et al. 2016). One option is to assign provider objective functional ratings (e.g., number of sessions attended) and assign clients subjective ratings (e.g., perceived symptom severity).

Limitations

The use of primarily archival data limited the sample size and range of variables available. For instance, it was not possible to independently assess client functioning to identify which providers were more or less accurate. Additionally, the use of a single site decreased the generalizability of the results. The approach of ROM is likely to be different across agencies, including types of agencies, funding sources, and locations.

Conclusions

This study considered sources of measurement error in two specific ROM measures, given inconclusive evidence whether ROM is useful in community settings for decision-making or improved outcomes. There were several sources of measurement error, which varied depending on the domain being monitored. Both provider factors (i.e., experience level, education, and current position) and non-target client characteristics (i.e., gender, age, diagnosis, and cognitive ability) predicted ROM scores. Increasing objectivity of measures and increasing provider training may improve these issues, given that objective domains of measurement showed less measurement error. Additionally, increasing ease of use, and integrating client-rated outcome measures may also improve accuracy of ROM in community mental health.

Compliance with Ethical Standards

References

Amble, I., Gude, T., Stubdal, S., Andersen, B. J., & Wampold, B. E. (2015). The effect of implementing the Outcome Questionnaire-45.2 feedback system in Norway: A multisite randomized clinical trial in a naturalistic setting. *Psychotherapy Research, 25*(6), 669–677. https://doi.org/10.1080/10503307.2014.928756.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271–285.

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook*

of psychotherapy and behavior change (pp. 258–297). Hoboken, NJ: Wiley.

Barker, S., Barron, N., McFarland, B. H., & Bigelow, D. A. (1994). A community ability scale for chronically mentally ill consumers: Part I. Reliability and validity. *Community Mental Health Journal, 30*, 363–383.

Bijl, R. V., & Ravelli, A. (2000). Psychiatric morbidity, service use, and need for care in the general population: Results of The Netherlands Mental Health Survey and Incidence Study. *American Journal of Public Health, 90*(4), 602–607.

Bilsker, D., & Goldner, E. M. (2002). Routine outcome measurement by mental health-care providers: Is it worth doing? *The Lancet, 360*, 1689–1690.

Boswell, J. F., Constantino, M. J., Kraus, D. R., Bugatti, M., & Oswald, J. M. (2016). The expanding relevance of routinely collected outcome data for mental health care decision making. *Administration and Policy in Mental Health, 43*, 482–491.

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25*(1), 6–19. https://doi.org/10.1080/10503307.2013.817696.

Burgess, P., Pirkis, J., & Coombs, T. (2015). Routine outcome measurement in Australia. *International Review of Psychiatry, 27*(4), 264–275. https://doi.org/10.3109/09540261.2014.977234.

Carli, L. L. (1999). Gender, interpersonal power, and social influence. *Journal of Social Issues, 55*(1), 81–99.

Corcoran, R., Mercer, G., & Frith, C. (1995). Schizophrenia, symptomatology, and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research, 17*, 5–13.

Cradock, J., Young, A. S., & Sullivan, G. (2001). The accuracy of medical record documentation in schizophrenia. *Journal of Behavioral Health Services & Research, 28*(4), 456–465.

de Jong, K., Timman, R., Hakkaart-Van Roijen, L., Vermeulen, P., Kooiman, K., Passchier, J., & Busschbach, J. V. (2014). The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. *Psychotherapy Research, 24*(6), 629–639. https://doi.org/10.1080/10503307.2013.871079.

Drake, R. E., Bond, G. R., & Essock, S. M. (2009). Implementing evidence based practices for people with schizophrenia. *Schizophrenia Bulletin, 35*(4), 704–713.

Fusar-Poli, P., Papanastasiou, E., Stahl, D., Rocchetti, M., Carpenter, W., Shergill, S., & McGuire, P. (2015). Treatments of negative symptoms in schizophrenia: Meta-analysis of 168 randomized placebo-controlled trials. *Schizophrenia Bulletin, 41*(4), 892–899. https://doi.org/10.1093/schbul/sbu170.

Garb, H. N. (1989). Clinical judgement, clinical training, and professional experience. *Psychological Bulletin, 105*(3), 387–396.

Gelkopf, M., Pagorek-Eshel, S., Trauer, T., & Roe, D. (2015). Routine outcome measurement in mental health service consumers: Who should provide support for the self-assessments? *Evaluation and Program Planning, 50*, 43–46. https://doi.org/10.1016/j.evalprogplan.2015.02.001.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2001). Routinely administered questionnaires for depression and anxiety: Systemic review. *BMJ, 322*, 406–409.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002a). Outcomes research in mental health: Systematic review. *The British Journal of Psychiatry, 181*, 8–16.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002b). Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome- a systematic review. *Psychological Medicine, 32*, 1345–1356.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002c). Psychiatrists in the UK do not use outcomes measures. *British Journal of Psychiatry, 180*, 101–103.

Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research & Practice, 35*(5), 485–491.

Hatfield, D. R., & Ogles, B. M. (2007). Why some clinicians use outcome measures and others do not. *Administration and Policy in Mental Health and Mental Health Services Research, 34*, 283–291.

Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivilighan, D. M., & Gelso, C. I. (2017). Therapist expertise in psychotherapy revisited. *The Counseling Psychologist, 45*(1), 7–53.

Hoenders, R. H. J., Bos, E. H., Bartels-Velthuis, A. A., Vollbehr, N. K., van der Ploeg, K., de Jonge, P., & de Jong, J. T. V. M. (2014). Pitfalls in the assessment, analysis, and interpretation of routine outcome monitoring (ROM) data: Results from an outpatientclinic for integrative mental health. *Administration and Policy in Mental Health and Mental Health Services Research, 41*(5), 647–659. https://doi.org/10.1007/s10488-013-0511-7.

Hróbjartsson, A., Thomsen, A. S. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., … Brorson, S. (2013). Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *Canadian Medical Association Journal*. https://doi.org/10.1503/cmaj.120744.

Ionita, G., & Fitzpatrick, M. (2014). Bringing science to clinical practice: A Canadian survey of psychological practice and usage of progress monitoring measures. *Canadian Psychology/Psychologie Canadienne, 55*(3), 187–196. https://doi.org/10.1037/a0037355.

Ionita, G., Fitzpatrick, M., Tomaro, J., Chen, V. V., & Overington, L. (2016). Challenges of using progress monitoring measures: Insights from practicing clinicians. *Journal of Counseling Psychology, 63*(2), 173–182. https://doi.org/10.1037/cou0000122.

Jastak, J. F., Jastak, S. R., & Wilkinson, G. S. (1993). *The Wide-range achievement test, revised 2*. Wilmington: Guidance Associates of Delaware.

Jensen, R. E., Rothrock, N. E., DeWitt, E. M., Spiegel, B., Tucker, C. A., Crane, H. M., … Crane, P. K. (2015). The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Medical Care, 53*(2), 153–159. https://doi.org/10.1097/MLR.0000000000000289.

Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes towards standardized assessment tools. *Journal of Clinical Child and Adolescent Psychology, 39*(6), 885–896.

Kendrick, T., El-Gohary, M., Stuart, B., Gilbody, S., Churchill, R., Aiken, L., … Moore, M. (2016). Routine use of patientclient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. In The Cochrane Collaboration (Ed.), Cochrane database of systematic reviews. Chichester: Wiley https://doi.org/10.1002/14651858.CD011119.pub2.

Kenny, D. A. (1996). Models of non-independence in dyadic research. *Journal of Social and Personal Relationships, 13*(2), 279–294. https://doi.org/10.1177/0265407596132007.

Kotte, A., Hill, K. A., Mah, A. C., Korathu-Larson, P. A., Au, J. R., Izmirian, S., … Higa-McMillan, C. K. (2016). Facilitators and barriers of implementing a measurement feedback system in public youth mental health. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(6), 861–878. https://doi.org/10.1007/s10488-016-0729-2.

Krampen, G. (1991). *Inventory on competence and control beliefs*. Gottingen: Hogrefe.

Lambert, M. J. (2003). Is it time for clinicians to routinely track patientt outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*(3), 288–301. https://doi.org/10.1093/clipsy/bpg025.

Lang, K., Meyers, J. L., Korn, J. R., Lee, S., Sikirica, M., Crivera, C., Dirani, R., & Menzin, J. (2010). Medication adherence and hospitalization among patients with schizophrenia treated with antipsychotics. *Psychiatric Services, 61*, 1239–1247.

Lyon, A. R., Dorsey, S., Pullmann, M., Silbaugh-Cowdin, J., & Berliner, L. (2015). Clinician use of standardized assessments following a common elements psychotherapy training and consultation program. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(1), 47–60. https://doi.org/10.1007/s10488-014-0543-7.

McHorney, C. A., & Tarlov, A. R. (1994). The use of health status measures for individual patient level applications: Problems and prospects. *Quality of Life Research, 3*(1), 43–44.

Miller, S. D., Hubble, M. A., Chow, C., & Seidel, J. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy, 52*(4), 449–457.

Reddy, L. F., Horan, W. P., & Green, M. F. (2015). Motivational deficits and negative symptoms in schizophrenia: Concepts and assessments. In *Behavioral neuroscience of motivation* (pp. 357–373). Cham: Springer. https://doi.org/10.1007/7854_2015_379.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Therapy and clinical interpretation*. Tucson, AZ: Neuropsychological Press.

Roe, D., Drake, R. E., & Slade, M. (2015a). Routine outcome monitoring: An international endeavour. *International Review of Psychiatry, 27*(4), 257–260. https://doi.org/10.3109/09540261.2015.1070552.

Roe, D., Gelkopf, M., Gornemann, M. I., Baloush-Kleinman, V., & Shadmi, E. (2015b). Implementing routine outcome measurement in psychiatric rehabilitation services in Israel. *International Review of Psychiatry, 27*(4), 345–353. https://doi.org/10.3109/09540261.2015.1025722.

Rosenheck, R., Fontana, A., & Stolar, M. (1999). Assessing quality of care: Administrative indicators and clinical outcomes in post-traumatic stress disorder. *Med Care, 37*, 180–1888.

Sajatovic, M., Micula-Gondek, W., Tatuoka, C., & Bialko, C. (2011). The relationship of gender and gender identity to treatment adherence among individuals with bipolar disorder. *Gender Medicine, 8*(4), 261–268.

Serby, M. (2003). Psychiatric resident conceptualizations of mood and affect within the mental status examination. *American Journal of Psychiatry, 160*(8), 1527–1529. https://doi.org/10.1176/appi.ajp.160.8.1527.

Spaulding, W., & Deogun, J. (2011). A pathway to personalization of integrated treatment: Informatics and decision science in psychiatric rehabilitation. *Schizophrenia Bulletin, 37*, 129–137.

Spreen, O., & Strauss, E. A. (1998). *Compenium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Oxford.

Tait, L., Birchwood, M., & Trower, P. (2002). A new scale (SES) to measure engagement with community mental health services. *Journal of Mental Health, 11*(2), 191–198.

Trauer, T., Callaly, T., & Herrman, H. (2009). Attitudes of mental health staff to routine outcome measurement. *Journal of Mental Health, 18*(4), 288–297. https://doi.org/10.1080/09638230701879177.

Ussher, J. M. (2013). Diagnosing difficult women and pathologising feminity: Genderbias in psychiatric nosology. *Feminism & Psychology, 23*(1), 63–69.

Wampold, B. E. (2015). Routine outcome monitoring: Coming of age—With the usual developmental challenges. *Psychotherapy, 52*(4), 458–462. https://doi.org/10.1037/pst0000037.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: Results from the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 629–640. https://doi.org/10.1001/archpsyc.62.6.629.

Wechsler, D. (2008). *Wechsler adult intelligence: Technical and interpretive manual* (4th ed.). San Antonio: Pearson.

Wennberg, J. E. (1990). Outcomes research, cost containment, and the fear of health care rationing. *The New England Journal of Medicine, 323*(17), 1202–1204.

Wolpert, M., Ford, T., Trustam, E., Law, D., Deighton, J., Flannery, H., & Fugard, R. J. B. (2012). Client-reported outcomes in child and adolescent mental health services (CAMHS): Use of idiographic and standardized measures. *Journal of Mental Health, 21*(2), 165–173. https://doi.org/10.3109/09638237.2012.664304.

Young, A. S., Grusky, O., Sullivan, G., Webster, C. M., & Podus, D. (1998). The effect of provider characteristics on case management activities. *Administration and Policy in Mental Health, 26*(1), 20–32.