Contextualizing the road to recovery: A novel method of assessing outcome trajectories in clinical trials

9 authors, including:

Emily Treichler
VA San Diego Healthcare System
26 PUBLICATIONS **81** CITATIONS

William Hochberger
VA San Diego Healthcare System/University of California, Dan Diego
14 PUBLICATIONS **82** CITATIONS

Gregory A Light
University of California, San Diego
231 PUBLICATIONS **8,646** CITATIONS

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

Project   NEUROPHYSIOLOGICAL BIOMARKERS OF BEHAVIORAL DIMENSIONS FROM CROSS-SPECIES PARADIGMS View project

Project   Collaborative Decision Skills Training View project

Psychological Services

Contextualizing the Road to Recovery: A Novel Method of Assessing Outcome Trajectories in Clinical Trials

Emily B. H. Treichler, Lauren Cardoso, Yifeng Du, John Nungaray, William C. Hochberger, Yash B. Joshi, Joyce Sprock, Amy N. Cohen, and Gregory A. Light

Contextualizing the Road to Recovery: A Novel Method of Assessing Outcome Trajectories in Clinical Trials

Emily B. H. Treichler
Veterans Affairs Desert Pacific Mental Illness Research,
Education, and Clinical Center (MIRECC), San Diego,
California, and University of California, San Diego

Lauren Cardoso
University of California, San Diego

Yifeng Du
University of California, San Diego, and Columbia University

John Nungaray
University of California, San Diego

William C. Hochberger
Veterans Affairs Desert Pacific MIRECC, San Diego,
California, and University of California, San Diego

Yash B. Joshi and Joyce Sprock
University of California, San Diego

Amy N. Cohen
Veterans Affairs Desert Pacific MIRECC, Los Angeles,
California, and University of California, Los Angeles

Gregory A. Light
Veterans Affairs Desert Pacific MIRECC, San Diego,
California, and University of California, San Diego

In clinical trials, standardized assessment conducted by research staff facilitates identification of treatment benefit. Narrative notes completed by clinicians offer a novel source to characterize and contextualize outcomes. In this study, we examine qualitative analysis of clinical notes as a method to augment quantitative outcome measures and supply meaningful context in clinical trials. Two hundred eighty-four clinical progress notes from 19 participants with schizophrenia or schizoaffective disorder assigned to receive either auditory-targeted cognitive training or treatment as usual were included. Qualitative analysis of weekly progress notes written by clinicians involved in ongoing care of the participants was used to identify overall outcome trajectories and specific changes in program participation, social functioning, and symptom severity. Trajectories were compared with the parent study's 2 primary outcome measures. Qualitative analysis identified personalized and complex trajectories for individual participants. Approximately half the participants improved overall. Most participants displayed improved program participation and social functioning, whereas most participants experienced symptom deterioration. Engagement in targeted cognitive training did not impact change in trajectories. Qualitative trajectories were congruent (e.g., both indicated improvement) with the 2 primary outcome measures for 26–36% of the participants depending on the comparison. Including qualitative analysis of clinician progress notes provides useful context and identifies underlying processes not captured in quantitative data. However, they cannot replace quantitative outcome measurement. Better alignment

with clinician- and patient-targeted outcomes may strengthen clinical trials. Qualitative analysis of routinely collected data can benefit research and programmatic decision making in usual care settings.

Impact Statement
This study found that qualitative analysis of routinely collected narrative data in clinical settings offers a novel and impactful method to understand and contextualize outcomes in clinical trials. Researchers should consider integrating nonresearch clinician perspectives and routinely collected qualitative data into clinical trials to facilitate assessment of complex clinical outcomes and better align clinical trial outcome measurement with outcomes targeted by clinicians and patients.

Keywords: psychosis, cognitive remediation, clinical notes, content analysis, narrative analysis

In clinical trials, hypothesized benefits of treatment are measured using standardized assessments conducted by research staff that capture domains including symptom severity and cognition. This approach maximizes internal validity and informs treatment efficacy. Even in effectiveness trials that prioritize evidence of generalized benefits in usual care settings, standardized assessments tend to prevail, given their psychometric characteristics and ease of translation across settings. In contrast, clinical trials rarely assess routinely collected data in usual care settings that may inform relevant outcome trajectories, specifically narrative progress notes written by clinical staff.

Including qualitative analysis of routinely collected data may elucidate the broader context surrounding, and the trajectory preceding, the final benchmark assessed in a clinical trial (i.e., the primary outcome variable). The perspective offered by clinician progress notes facilitates assessment of outcomes in a wider context, typically with more data points throughout the trial period per participant. For example, although cognitive performance can be assessed with high validity, cognitive function outside the research context often diverges from these measures (Sabbag et al., 2011; Silberstein & Harvey, 2019; Treichler et al., 2019). Because frontline clinicians directly observe patients during nonresearch interactions over weeks to years, their progress notes may reveal subtle and complex changes in clinical status, treatment-related behavior, and psychosocial functioning not reflected in standardized assessments. Harnessing notes written by frontline clinicians may inform programmatic decision making by demonstrating whether interventions generalize to outcomes that are relevant to and easy to discern by stakeholders. If so, this would facilitate identification of ecological validity and increase stakeholder buy-in to effective treatments. Therefore, we posit that the potential use of progress notes in clinical trials might provide insight and useful context to researchers regarding patients' overall recovery trajectories and facilitating programmatic decision making about promising interventions.

To our knowledge, there is no literature incorporating qualitative analysis of progress notes into clinical trials. Broadly, a limited number of nontrial mental health studies have included qualitative analysis of progress notes, including examining forensic mental health nurses' perceptions of aggressive behavior and studying ways that people with serious mental illness support their families (Berring, Pedersen, & Buus, 2015; Haselden et al., 2018). Clinical notes of varying types have also been incorporated into medical research (e.g., March et al., 2016; Miller & Velanovich,

2010; Ye & Fabbri, 2018). Qualitative analysis is a promising approach for clinical trials because it allows researchers to take into consideration the full context of complex phenomena (Hsieh & Shannon, 2005), in this case outcome trajectories for people with schizophrenia and schizoaffective disorder receiving usual care and participating in a clinical trial. Outcome trajectories in these disorders are highly individualized and nonlinear (e.g., Liu, Choi, Reddy, & Spaulding, 2011; Verma, Subramaniam, Abdin, Poon, & Chong, 2012), meaning that trajectories may not be fully captured by quantitative measurement. Progress notes are a particularly appropriate way to capture these complex trajectories, given that clinicians with expertise in mental health are already collecting these data. Additionally, progress notes often contain explicit discussion of topics like symptom severity, social functioning, and treatment engagement and essential domains for research objectives and clinical decision making as well as being domains of interest to patient, clinician, and administrative stakeholders.

Therefore, exploration of this novel approach to assess outcomes of a clinical trial has the potential to be impactful. However, previous work suggests that progress notes collected about people with schizophrenia may underestimate symptom severity, particularly for patients with more severe illness, those who are black, and those perceived as noncompliant (Cradock, Young, & Sullivan, 2001). Analysis of clinician progress notes written during a clinical trial offers the opportunity to assess the usefulness of these notes for research and more broadly, for use in programmatic decision-making.

This study evaluated the innovative approach of incorporating qualitative analysis of clinician progress notes into a clinical trial of auditory-targeted cognitive training (TCT), a computerized cognitive intervention for people with schizophrenia and related disorders that significantly improves neurocognitive performance and functional capacity (Fisher, Holland, Merzenich, & Vinogradov, 2009; Fisher, Mellon, Wolkowitz, & Vinogradov, 2016, 2017). Specifically, TCT uses six computer-based exercises developed by Brain HQ by Posit Science Corporation that target auditory perception, auditory processing speed, and auditory memory. Exercises apply a personalized algorithm so that regardless of individual learning rate, participants are continuously and consistently challenged. The complete TCT program is 40 hr total, usually broken down into hour segments over 10 weeks.

In the parent study, we reported that people with schizophrenia and schizoaffective disorder who require long-term, high-intensity

care showed improvements in verbal learning and memory, and hallucination symptom severity (Thomas et al., 2018). In the current study, we use content and narrative analysis of clinical progress notes to achieve three study aims: (a) to characterize overall outcome trajectories; (b) to identify specific trajectories in domains of interest to researchers, clinicians, and patients: program participation, social functioning, and symptom severity; and (c) to determine level of agreement between this analysis and those of the parent study. Because of our overall pragmatic approach to inquiry, focused on using qualitative analysis to better understand these complex research questions and contextualize quantitative data, content and narrative analysis approaches were selected because they allow for triangulation of the qualitative data and can accommodate a priori thematic categories.

Method

The Institutional Review Board of the University of California, San Diego approved this study (IRB130874).

Participants

A subsample of 19 participants (TCT = 9, treatment as usual [TAU] = 10) was randomly selected from the 46 participants enrolled in the parent study. Participants in this study were current patients at a locked, community-based, long-stay residential program in Southern California. This program is part of a stepped-care treatment plan in which individuals come from an acute inpatient stay, reside in the program to achieve sufficient psychiatric stability and independent living skills, and then transition to a lower level of care. All participants in this study were under public or private conservatorship. The participants had diagnoses of schizophrenia or schizoaffective disorder based on an interview using the Structured Clinical Interview for *DSM–IV–TR* (*Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, text revision; First, Spitzer, Gibbon, & Williams, 2002). Study exclusion criteria included inability to provide assent; insufficient English fluency; history of head injury with loss of consciousness >30 min; neurological illness or severe systemic illness; and current mania. Participants completed baseline assessments and were randomized to either TCT or TAU using stratified random assignment by age, ethnicity, and gender. Participants in the TAU group completed study assessments but no TCT or other (e.g., placebo) sessions. All participants participated in usual care provided by the residential program.

The subsample included 11 women and eight men. Their average age was 35.47 years (*SD* = 12.07). Eleven participants were White, two were Black or African American, three were Hispanic or Latino, one was Asian, one was Native American, and four had a mixed race/ethnic background. Mann-Whitney *U* tests were used to examine potential differences by group in age and baseline symptom and cognitive scores, whereas a Fisher's exact test was used to identify any differences in gender, race, and ethnicity by group. All results were not significant, indicating that the TCT and TAU subsamples used in this study were statistically equivalent across these variables.

Procedure

For a complete description of the TCT trial, please see Thomas et al. (2018). The two primary outcome measures from the parent study were the Measurement and Treatment Reearch to Improve Cognition in Schizophrenia (MATRICS) consensus cognitive battery verbal learning and memory subscale t-scores (MCCB; Nuechterlein et al., 2008) and the Scale for the Assessment of Positive Symptoms (SAPS; Andreasen, 1984) global hallucination severity score. The MCCB is a validated assessment of neurocognition for adults with schizophrenia; the verbal learning and memory subscale has strong construct, concurrent, and divergent validity, high test-rest reliability, and no evidence of ceiling or floor effects (Nuechterlein et al., 2008). The SAPS is a rating scale based off a semistructured clinical interview that has strong construct and concurrent validity and high test-rest reliability (Andreasen, 1984).

The progress notes assessed in this study are a part of the residential programs' standard care. The frontline clinicians who authored notes were not involved with the parent TCT study. These clinicians were functionally blinded to study condition, meaning that condition was never communicated to staff and study activities occurred in a separate ward from the program, minimizing the likelihood that clinicians were aware of participants' study assignment. The decision to analyze these notes was made after the notes were written, meaning no intentional or unintentional clinician priming occurred. Twenty-two total clinicians contributed to these notes. Of these, 14 contributed to only one participant's set of notes, and the other eight contributed to two sets each. Two clinicians contributed to notes for a TAU participant and a TCT participant. For 11 of the participants, all notes were written by the same clinician; three participants had two different clinician writers and four participants had three different clinician writers.

Weekly progress notes are written by master's-level clinicians working toward licensure under the supervision of a licensed clinician. Each note, capturing 1 week, includes observations of patient behavior and presentation, the ability to complete activities of daily living, program participation, progress toward recovery, and any other notable incidents including physical aggression, or self-harm. Notes inform treatment decisions and client privileges.

Weekly notes were collected from 1 week prior to beginning the study until the week the study was completed. Total time in the study varied (e.g., by number of TCT sessions completed) between 10 and 20 total weekly notes to assess, with a total of 284 notes across all participants. The subsample size of 19 was selected because it was expected to provide sufficient data to achieve qualitative outcome saturation, given that these progress notes are both thick and rich (Fusch & Ness, 2015). We defined these notes as thick because of their sheer number (284) and length (one to two typed pages each), and rich based on the complex content including many different aspects of each patients' presentation and progress, use of specific anecdotes from staff and quotes from patients, and detailed examples of progress or deterioration throughout the week (e.g., attendance in specific activities).

Analytic Plan

The qualitative analysis applied two independent techniques, which were then triangulated to create a more nuanced depiction of patient trajectories over the course of the study (Humble, 2009).

This was in keeping with our overall pragmatic approach, based on understanding a research problem thoroughly through approaches like methodological and investigator triangulation, both used here. The first technique was content analysis, a method of organizing qualitative data through cultivation of themes that represent meaningful portions of the data (Elo & Kyngäs, 2008; Hsieh et al., 2005). This content analysis used the deductive method, meaning that the themes were identified a priori. Three themes explicitly targeted in these notes (i.e., built into the template) that are of importance to researchers, clinicians, and patients alike were chosen: program participation, symptom severity, and social functioning. See Table 1 for operational definitions of these themes.

Two independent coders (Lauren Cardoso and Yifeng Du) coded each note using NVivo (QSR International, Burlington, Massachusetts) qualitative analysis software. Each participant's progress note set was analyzed together as a set, one note (and so 1 week) at a time, to identify trajectories over time. For each progress note, segments were coded to a primary theme, with week-by-week subthemes (i.e., Week 1, . . . Week 20) and valence indicators used to compare week-to-week changes. Each coder established baseline valence (positive, negative, or neutral) for each participant during Week 1 coding and then identified valence (improvement, deterioration, or maintenance) for each subsequent week based on whether that week's progress note suggested any change from the prior week and, if so, whether that change was positive (improvement) or negative (deterioration). No change was coded as maintenance. Coders met with a coding supervisor (Emily B. H. Treichler) routinely to discuss divergences between coders. Percent agreement statistics were calculated to measure interrater reliability rather than the kappa coefficient because of bias introduced to the kappa coefficient in a number of circumstances, including when the sample is not random and when the class prevalence is nonuniform, both of which apply to these data (Brennan & Silman, 1992; Erdmann, De Mast, & Warrens, 2015). Percent agreement was high for all sets of data. The final set of notes had 99.6% agreement.

For each of the three themes (program participation, symptom severity, and social functioning), valence ratings for each week were used to calculate week-by-week net scores. For each domain, an improvement comment received a score of 1, a deterioration comment received a score of -1, and a maintenance comment received a score of 0. For example, if a participant had two program participation comments coded as improvement and one program participation comment coded as maintenance in a given week, their net score for that week would be 2. Overall trajectories were calculated by adding all of the net scores for a given theme.

The second technique was narrative analysis, a method of assessing perspectives illustrated through language both literally and symbolically to understand context and longitudinal perspective contained within the data (Sosulski, Buchanan, & Donnell, 2010; Stern, Doolan, Staples, Szmukler, & Eisler, 1999). Narrative analysis was conducted by a single coder (Emily B. H. Treichler) and confirmed by a secondary reviewer (Lauren Cardoso). The coder wrote a summary of each progress note set, with specific attention to describing change over time in treatment areas including program participation, symptom severity, and social functioning. These summaries also included metadiscussion about the progress notes including any variance because of different clinician authors. The coder assigned an overall trajectory to each note set: improvement, marginal improvement, maintenance (baseline low), maintenance (baseline high), maintenance (baseline mixed), marginal deterioration, and deterioration. Marginal improvement captured participants who showed improvement in one or more areas but struggled significantly enough in other areas to prevent fitting an improvement trajectory, whereas marginal deterioration captured participants who deteriorated in one or more areas but had enough stability or improvement in other areas to prevent fitting a deterioration trajectory. Maintenance (baseline low), maintenance (baseline high), and baseline (mixed) were intended to capture participants whose overall status did not change over the course of the study but differentiated those who continued doing well (baseline high) throughout the study, those who continued to experience significant difficulties (baseline low), and those who had both significant strengths and significant difficulties (baseline mixed). The secondary reviewer had deep knowledge of these progress notes because of completing the content analysis and received all analytic notes. These statements were discussed until consensus between both researchers was reached. The independent coders were not aware of participant study condition. One participant was removed from content analysis because significant variability in reporting by different clinicians impaired the ability to make week-by-week interpretation. This participant was retained for narrative analysis because authorship variability could be appropriately accounted for.

Lauren Cardoso is a licensed master's-level clinician and research team member with significant clinical and research experience with this population and treatment setting. Yifeng Du was an undergraduate student at the time of analysis with no clinical experience and little research experience with this population and setting and is now a master's student in clinical psychology. Emily B. H. Treichler is a clinical psychologist with significant clinical and research experience with this population and treatment setting and background in qualitative methods. Combining these three perspectives allowed for balance between the expert knowledge of the population, setting, and clinical language used from Lauren Cardoso and Emily B. H. Treichler and an inexperienced eye to question and offer new perspectives on the notes from Yifeng Du. Amy N. Cohen and Gregory A. Light provided mentorship in the

Table 1

Operational Definitions of Themes in the Content Analysis

Theme	Definition
Program participation	Discussion of progress toward goals and program engagement. Specific groups attended each week were not coded.
Social functioning	Comments regarding clients' social interactions (e.g., initiating conversation, avoid engaging with other patients).
Symptoms	References to mental health symptoms including hallucinations, delusions, anxiety, and depression. Does not include low mood because of nonmental health causes.

analytic process. Amy N. Cohen is a clinical psychologist and qualitative methods expert, whereas Gregory A. Light is a content expert in TCT and the primary investigator of the TCT trial.

Results

Overall Trajectories

Table 2 reports overall trajectories assigned by narrative analysis. Approximately half of participants (53%) improved at least marginally.

Participants Whose Status Improved Overall

Of the eight participants who had improvement trajectories, three experienced consistent improvement across domains and time. For example, Participant 16's first progress note included many concerns and few strengths.

> She presented with a labile mood ranging from pleasant to tearful to isolative this past week. She displayed a delusional thought process this past week as evidenced by responding heavily to internal stimuli and having an increase in delusional thought content. She struggled with staff prompting or redirection secondary to her increased agitation or yelling at staff.

By week 11, Participant 16's emotional dysregulation and interpersonal aggression had decreased, although her symptoms continued to impair her functioning.

> She presented with a flat affect and a pleasant mood as she had an increase in eye contact and a minor increase in social interactions this past week with both staff and peers. Participant 16 displayed a tangential thought process this past week, as staff observed an increase in thought blocking, as well as slow, drawn-out speech. She also rapidly changed topics when engaged with staff and peers.

In the final week, week 17, her mood and social functioning both improved.

"She displayed a pleasant mood and a congruent affect secondary to being visible on the milieu, having positive range of facial expression, and having good social interactions with her peers and staff. [. . .] Participant 16 needed additional prompting and reality testing to her increase in paranoia; however, she appeared receptive to this reality testing."

Of the other participants, one significantly improved in some areas but less in others. Three had inconsistent trajectories indicating a two-steps-forward-one-step-back progression. Participant 19 continued to experience significant symptoms and engage in inappropriate behavior, but his willingness to engage in treatment

improved over time. He had a recorded breakthrough moment when he advocated for a new medication in Week 17 after reporting an intention to stop using all medications in Week 8.

> His strengths were in meeting with staff for check-in and to discuss issues and symptoms of psychosis or aggressive behavior. [. . .] He has the ability to advocate for himself but at times lacks the motivation and requires staff redirection and encouragement to facilitate engagement (Participant 19, Week 20).

Two participants were coded as marginal improvement. These participants' notes did not show much variability week to week, but both had some evidence of improvement in one or more areas, including participating in an optional vocational apprenticeship program.

Participants Whose Status Was Maintained Overall

There were six participants whose overall status was maintained during the course of the study. Among these, three were largely stable at baseline and remained so over time. Participant 13 began the study in a relatively high phase of the program (phases are completed sequentially and facilitate increased independence before discharge) and was discharged to a board-and-care facility in the final week of these notes. Throughout the notes, clinicians perceive Participant 13 as symptomatically stable, engaging in prosocial behavior and actively working toward discharge, with high treatment engagement.

> Participant 13 did not exhibit any symptoms of psychosis on the unit over the past week. [. . .] Participant 13 is attending specialized groups with an emphasis on coping skill development to assist him with his symptoms in treatment and after discharge [. . .] Participant 13's strength is his commitment to be successful in treatment and complete the program in a timely fashion (Participant 13, Week 6).

Two of the remaining three maintenance participants were coded as baseline mixed. They were primarily stable and engaged in care but also experienced important areas of difficulty (e.g., inappropriate social behavior) that did not improve. The final participant was baseline low and consistently displayed severe negative symptoms, guardedness, and low engagement in treatment. However, these symptoms did not worsen and did not present risk high enough to prevent him from being discharged to a lower level of care.

Participants Whose Status Deteriorated Overall

For each of the three participants who deteriorated over time, there was a specific change cascaded into other domains, leading

Table 2
Overall Trajectories

Group	Improvement	Marginal improvement	Maintenance (baseline high)	Maintenance (baseline mixed)	Maintenance (baseline low)	Marginal deterioration	Deterioration	Total
TAU	4	1	2	2	0	0	1	10
TCT	4	1	1	0	1	0	2	9
Total	8	2	3	2	1	0	3	19

Note. TCT = targeted cognitive training; TAU = treatment as usual.

to generalized deterioration. These changes were impaired sleep, development of a romantic relationship with another patient, and increased symptom severity, respectively. In week 1, Participant 10 is highly engaged in treatment, safe and appropriate on the unit, and exhibits no severe symptoms. Beginning in Week 2, the clinician writer notes sleep concerns:

> Though he has regularly struggled with getting out of bed, especially in the mornings, Participant 10 has successfully met requirements for participation in groups and activities.

By Week 4, the clinician reports that Participant 10's hypersomnia is impacting his program participation. These sleep issues become chronic, and although Participant 10 completes the minimum program requirements by attending afternoon activities, his overall functioning is notably impaired in his final week:

> Participant 10's level of social interaction has varied as Participant 10 has tended to sleep often when not on pass or engaged in groups or activities. [. . .] He has presented as disheveled. [. . .] Participant 10 has been medication compliant, though has required several prompts to take them due to him sleeping much of the day. He has refused all breakfasts and one lunch this week secondary to sleeping (Participant 10, Week 13).

Program Participation

The majority of program participation trajectories (78%) were positive, indicating an overall improvement trajectory (see Figure 1). Improvement comments captured engagement in voluntary and value-based activities:

She was able to self-advocate to participate in dual diagnosis this past week, where she was given positive feedback for her positive participation in the outing. Participant 14 was able to participate in a pass with her mother this past weekend where she was able to attend and return without incident (Participant 14, Week 2).

For the three participants whose participation deteriorated overall (17%), low program participation hindered progress through the program's phases and in two cases prevented discharge despite participants reaching or exceeding the typical 6 month length of stay:

> Participant 9 was minimally observed on the milieu as she prefers to spend time in her room. Participant 9 would need to be consistent in her program to be advanced to the next phase [of the program] (Participant 9, Week 12).

Symptom Severity

Most participants (83%) had deterioration symptom severity trajectories, meaning that their symptoms worsened over the course of the study (see Figure 2). Participants who deteriorated often had complex trajectories, indicating improvement in some areas (e.g., receptiveness to reality testing) and worsening in others (e.g., depression). Clinicians reported that these participants made efforts to engage in treatment, but these efforts did not prevent symptom deterioration:

> She presented with a tangential thought process secondary to having scattered thoughts, appearing anxious and worrisome, and she appeared to have an increase in paranoia this past week. Participant 16
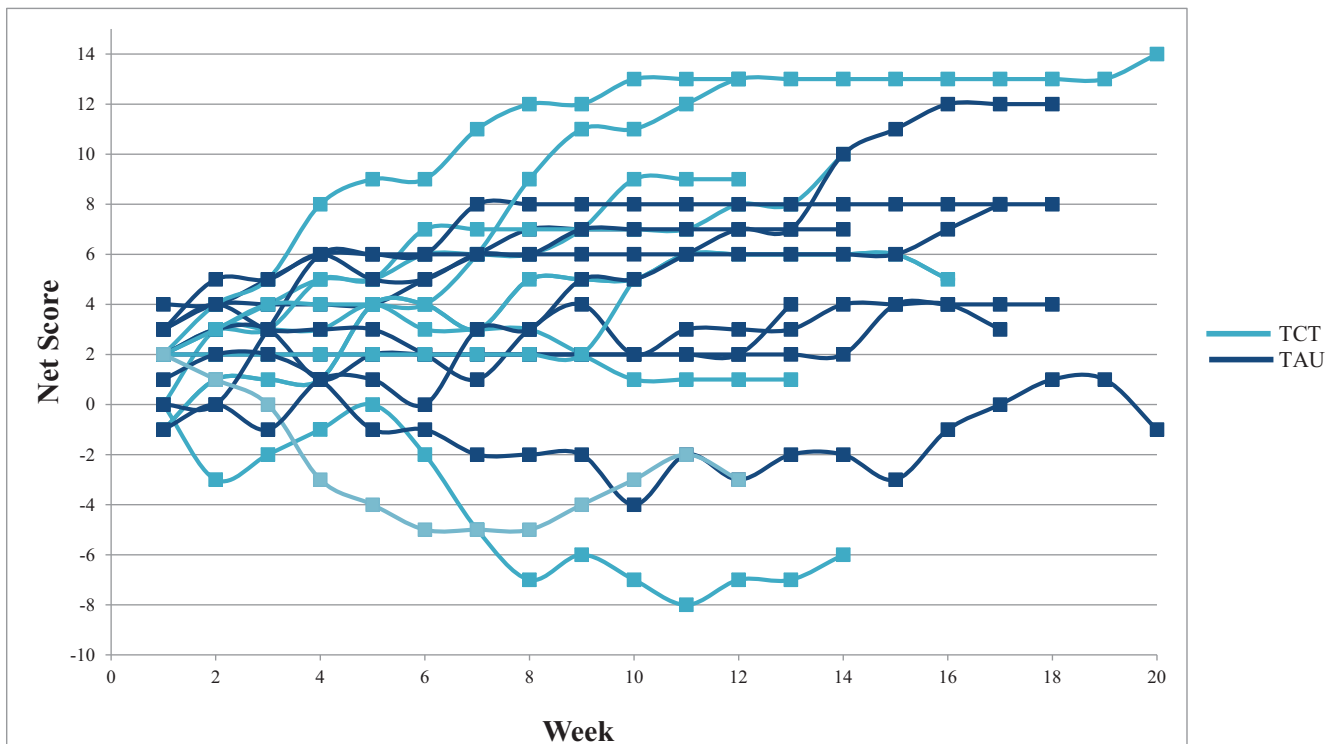

Figure 1. Program participation trajectories. See the online article for the color version of this figure.
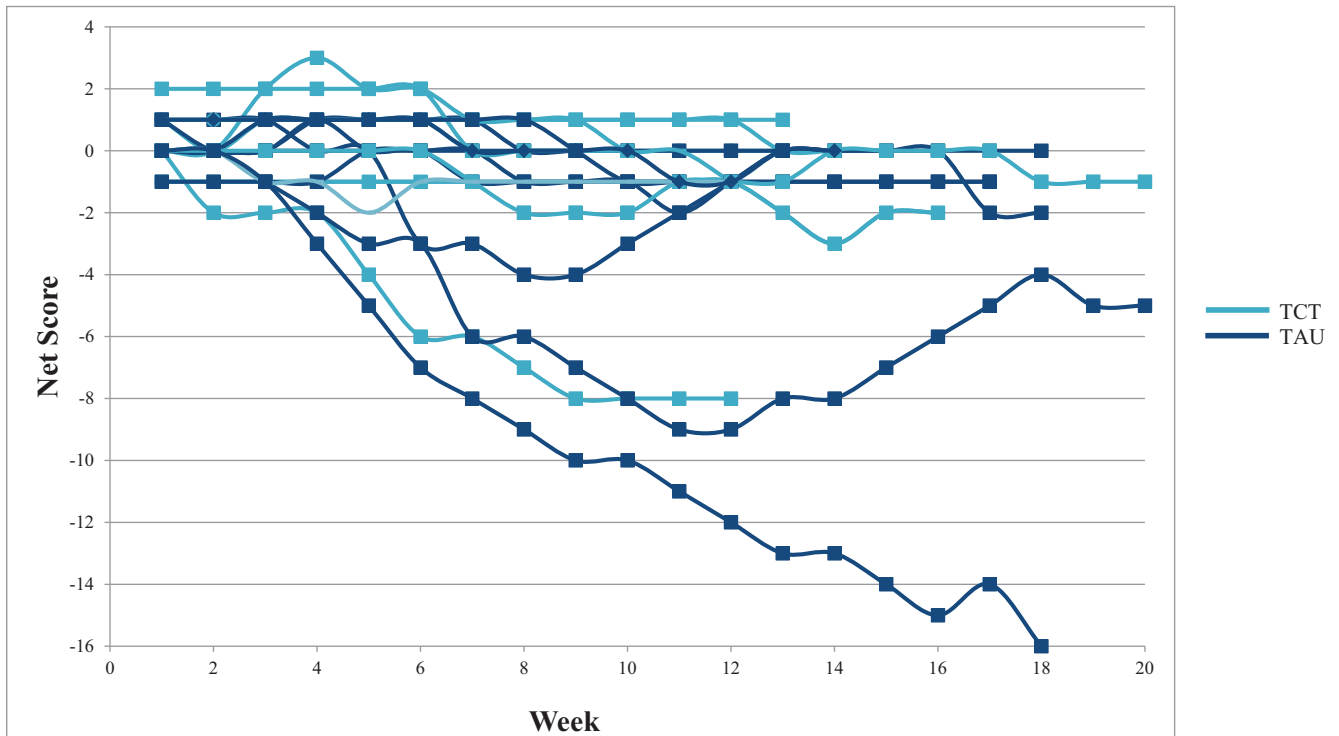
Figure 2.   Symptom severity trajectories. See the online article for the color version of this figure.

needed additional prompting and reality testing this past week secondary to her increase in paranoia; however, she appeared receptive to this reality testing. She was able to endorse concerns to nursing regarding side effects of her medications as well as endorse having an increase in symptoms of psychosis (Participant 16, Week 17).

Even among participants whose symptoms deteriorated overall, there was evidence of targeted improvements in specific symptoms or coping strategies. For example, Participant 3 began experiencing severe anxiety symptoms including panic attacks in Week 4:

> Participant 3 also attended the Mental Health Awards Dinner with staff, and though he maintained safe behavior, he reported to staff that he was feeling anxious and possibly having a panic attack when entering the room for dinner. He was able to stay outside with staff for 45 minutes and calm down before entering for dinner. He reported never feeling like that before but was able to maintain safety and immediately reached out to staff when needed.

These symptoms decreased and no longer impacted Participant 3's behavior by Week 13:

> Participant 3 presented in a positive and happy mood with bright affect, as he was excited for a pass this week; he successfully attended the 10-hour pass and excitedly talked about it with staff upon return. Participant 3 endorsed feeling much better and not as anxious anymore.

Two participants (11%) had maintenance trajectories, and a single participant improved over the course of the study. This participant, Participant 15, had difficulty managing anger, which

led to at least three verbal altercations with other patients and staff over Weeks 5–7. Participant 15 consistently expressed remorse for these behaviors and motivation to find better ways to express anger. There were no further altercations after Week 8, and the clinician noted that Participant 15 actively worked on his symptoms and coping skills, resulting in improved prosocial behavior and ability to leave the unit for outings.

Social Functioning

The majority of participants (78%) had improvement social functioning trajectories (see Figure 3). Improvement in social functioning was typically characterized by more frequent and appropriate social behavior toward staff and other patients, including making friends with peers and utilizing staff for facilitated coping skill use and goal attainment; successful outings with others; and working on social/interpersonal goals.

> This week Participant 7 was seen socializing with her peers and was pleasant and polite when staff engaged with her. Participant 7 displayed some insight into the fact that she ignored certain red flags in certain relationships that she had in the past. This week Participant 7 went on two buddy passes and returned without incident. [. . .] Participant 7 was seen engaging in her coping skills of talking through her feelings with staff and reading in her room (Participant 7, Week 16).

For two participants (11%), social functioning deteriorated over time. Both of these participants tended to isolate, and this tendency
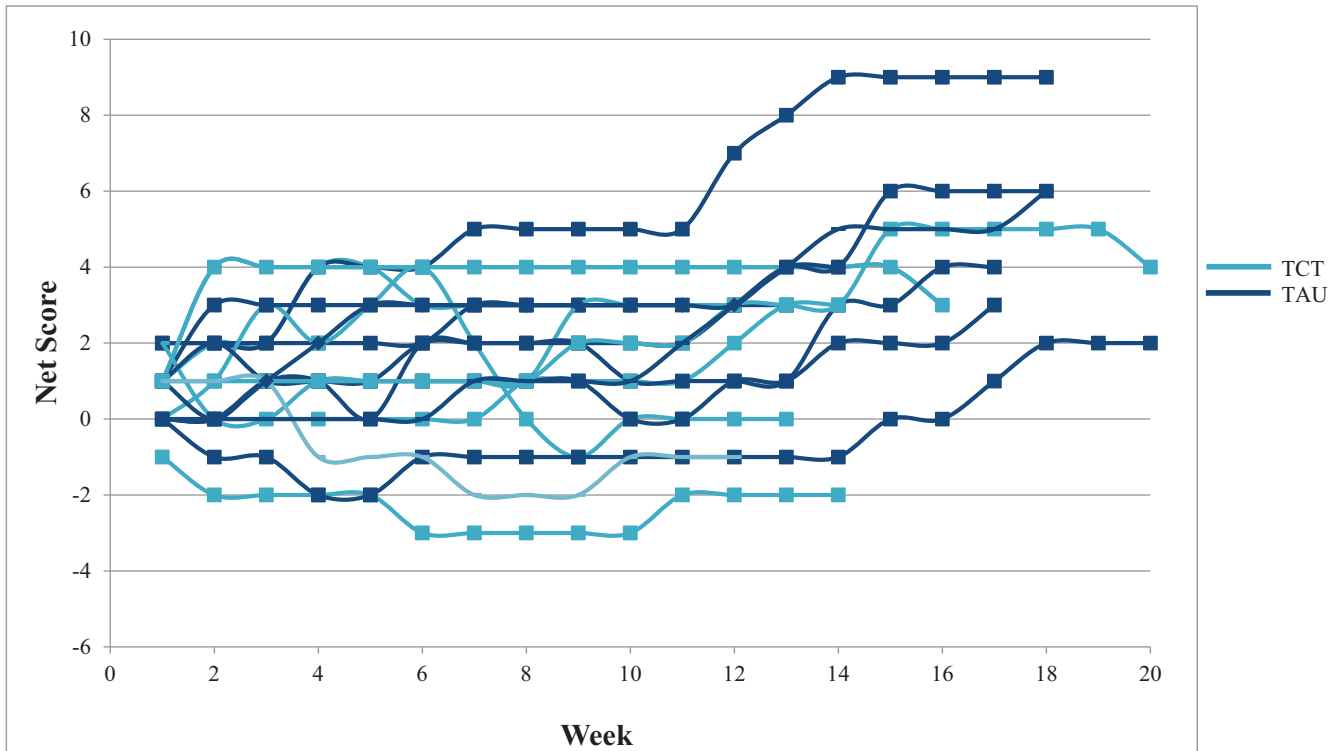
Figure 3. Social functioning trajectories. See the online article for the color version of this figure.

increased over time despite staff attempts to engage them and programming structure meant to encourage interaction with staff, peers, and family.

> Participant 8 continues to attend several groups each week, however, remains a passive participant. He is only social with staff when he is attempting to get his needs met. [. . .] Participant 8 reports these as his coping skills: time alone in his room and pacing the hallways (Participant 8, Week 13).

Progress Notes in Context With Primary TCT Outcomes

Based on narrative analysis, five TAU participants (50%) and five TCT participants (56%) improved at least marginally (see Table 3). The overall trajectories were compared with change scores for the parent study's two primary outcomes, SAPS Hallucinations and MCCB Verbal Learning and Memory to identify how many individual participants had congruent qualitative trajec-

tories and change scores on each measure (e.g., both improving). Scores were congruent for 32% of participants between the narrative analysis and SAPS Hallucinations, and 37% of participants between the narrative analysis and MCCB Verbal Learning and Memory. When comparing the symptom severity content analysis with SAPS Hallucinations scores, 26% of participants had congruent scores.

Participation in TCT did not appear to impact the specific trajectories of the three themes (i.e., program participation, social functioning, and symptom severity). Seven participants in each treatment group (78%) improved in program participation (see Table 3), although examination of individual trajectories (see Figure 1) indicated that TCT participants had broader variability in their trajectories, with some participants displaying the strongest, longest-term improvements and others having the most trouble with program participation, with the TAU having relatively lesser within-group variability. The only participant to improve in symptom severity was a TCT participant; Nearly all participants in both

Table 3
Trajectories by Group and Theme

Variables	Program participation			Social functioning			Symptom severity		
	Improvement	No change	Deterioration	Improvement	No change	Deterioration	Improvement	No change	Deterioration
TAU	7	1	1	9	0	0	0	2	7
TCT	7	0	2	5	2	2	1	0	8
Total	14	1	3	14	2	2	1	2	15

Note. TCT = targeted cognitive training; TAU = treatment as usual.

TAU (78%) and TCT (89%) deteriorated. Assessing the symptom severity individual trajectories further (see Figure 2), most deterioration across both groups was relatively minor (scores of −1 or −2), but one TAU and one TCT participant experienced moderate deterioration (scores of −5 and −8, respectively), and another TAU participant experienced severe deterioration (score of −16). For social functioning, all nine TAU participants (100%) improved, whereas five TCT participants (56%) improved and two (22%) showed no change. Individual trajectories for social functioning (see Figure 3) show most participants across both groups progressing closely together, suggesting no differences between the groups, although there was one TAU participant who had large gains beginning in week 12 unlike any other participant in the study.

Discussion

Qualitative analysis of clinical progress notes demonstrated utility in this study as a method to provide context and increase understanding of recovery trajectories in clinical trials. The progress notes describe personalized trajectories that capture changes in multiple domains including program participation, social functioning, and symptom severity. Individual outcome trajectories varied between the participants overall and by subdomain, facilitating multidimensional and individualized understanding of outcomes. Approximately half of the participants (53%) improved overall over the course of the study. Most participants had improvements in program participation and social functioning (78% for both categories), whereas 83% of participants experienced worsening in symptom severity. There was also consistent evidence that these notes were not congruent with the quantitative outcome measures used in the parent study. Comparison of our content analysis of symptom severity with SAPS Hallucination symptom severity scores did not find congruence with the two measurement approaches for most participants (26% agreement). Comparison of narrative analysis trajectories of each participant with the SAPS Hallucination scores and MCCB Verbal Learning and Memory was also not congruent for most participants (32% and 37% agreement, respectively).

One way to interpret these findings is that the progress notes are not accurate depictions of patient status or trajectories, limiting their utility. As mentioned previously, Cradock et al. (2001) found discrepancies because of bias in progress note reporting. However, in their case, the progress notes depicted patients as experiencing fewer symptoms and side effects compared with quantitative measures of symptom severity. Our findings indicate the opposite: that clinical progress notes depicted most patients experiencing worsening symptoms, whereas the SAPS Hallucinations found improvements. Rather than considering this evidence of measurement flaws in either the progress notes or the quantitative measures, we argue that this is evidence of measurement difference. Symptom reports in the progress notes captured all types of symptoms, whereas the SAPS Hallucinations measures hallucinations alone. Although some participants did continue to experience hallucinatory symptoms, clinicians reported a range of symptoms including paranoia, anxiety, depression, sleep disturbance, and aggressive behavior. The intervention studied, TCT, specifically targets auditory functioning and symptoms that result because of auditory dysfunction. Discrepancies between these measurement approaches therefore do not undercut either approach but rather indicate that participants experienced targeted symptom reduction, supporting TCT's effectiveness in the auditory pathway.

That the overall qualitative trajectories and two primary outcomes are incongruent for about two thirds of the participants are perhaps not surprising, but it is informative. Specifically, one of the primary outcomes, cognition, is not an area that the clinical progress notes commonly captured, whereas program participation and social functioning were not captured in the primary outcomes of the parent study. If our interest is to better align clinical research and clinical work to improve scientific impact and increase clinician use of scientific results, then becoming better aligned with clinicians in terms of key outcome domains is needed.

Similarly, to achieve recovery-oriented care, outcome measurement must prioritize outcomes most important to the target population. In a qualitative study, outpatients with psychotic disorders described their primary treatment goals as getting by (having fundamental needs met), getting back (living with mental illness), and getting on (living without prominent mental illness; Yarborough, Yarborough, Janoff, & Green, 2016). In one rank-based study, improved cognition was ranked in the top three most prioritized treatment goals among outpatients with schizophrenia, but reduced symptoms was not (Kuhnigk, Slawik, Meyer, Naber, & Reimer, 2012). Desired outcomes vary by patient, so using personalized outcome measures (e.g., Lindhiem, Bennett, Orimoto, & Kolko, 2016) can measure change in individually identified outcomes. Within the context of a TCT trial, connection to tangible, meaningful improvements for patients may improve patient buy-in and facilitate identification of pertinent outcomes (see Thomas et al., 2018, for further discussion).

Most study participants, regardless of group, had improvements in both program participation and social functioning (78% of all participants for both categories). There was no indication that TCT participants improved more than TAU; if anything, the findings suggest that more TAU participants experienced improvements in social functioning specifically. Validated tests of functioning were not included in the parent study's assessment battery, so it is not possible to compare measurement approaches within this sample. However, Fisher and colleagues included functional outcomes in their studies of TCT with outpatient populations and have not yet found significant improvements, although functional capacity does improve (cf. Fisher et al., 2009, 2015, 2017). This provides evidence that our finding is congruent with the broad literature and suggests that the clinical progress notes may accurately depict no improvement in functioning because of engagement in TCT.

Beyond comparison with quantitative outcome measures, the clinical progress notes revealed meaningful detail, context, and variability in outcome trajectories. For example, although there was no change in Participant 3's symptom severity as measured by SAPS Hallucination scores, qualitative analysis revealed that he experienced an acute onset of anxiety symptoms that abated over time. Participant 3's response to these new symptoms included asking for help and attempting new coping skills, which bodes well for his ability to manage his mental health in the future. Therefore, despite his maintenance status in the quantitative measures of symptom severity, it is clear that Participant 3 experienced positive growth in his recovery during the study period. Recovery is rarely linear, and inclusion of clinician progress notes captures that complexity. These analyses highlight how clinical outcomes (e.g.,

the primary outcomes assessed by the TCT trial) and recovery outcomes (e.g., these more nuanced changes in functioning and coping) both overlap and diverge. Assessing both in the same clinical trial can provide important context regarding effectiveness of an intervention and inform programmatic decisions by highlighting how interventions may impact these complex trajectories. In that sense, we may be able to bring the usual-care clinicians' expertise into clinical trials via these notes, strengthening our understanding of our data and increasing our ability to benefit usual care.

These findings supported the essential tenets of recovery-oriented care (Glynn, 2014; Spaulding, Sullivan, & Poland, 2003): that a person with serious mental illness can achieve functional improvements and meet treatment goals, even if their symptoms do not improve or remit. Although some past work has found improvement in symptoms alongside improvement in functioning among similar populations in high-intensity settings (e.g., Liu et al., 2011; Morrens et al., 2011), other related work aligns with the discrepancy in outcomes found in this study (e.g., Treichler & Spaulding, 2015). Broadly, studies of recovery find that symptom and functional recovery often occur independent of each other (Verma et al., 2012). Although individual outcomes varied in this study, clinicians found that a large majority of participants improved in program participation and social functioning, indicating that this high-intensity setting does lead to functional improvements.

We argue that these data offer a novel, nuanced, and meaningful perspective that would be impossible to attain through traditional quantitative measurement. In this study, our evaluation of these progress notes supports previous findings that TCT leads to targeted and not general symptom reduction and aligns with other researchers' findings that TCT does not lead to overall functional improvement. However, these data cannot replace quantitative measurement and may be best suited to provide context, identifying underlying processes not captured in quantitative data, and inform ways to better align research goals with clinical goals.

Based on these conclusions, we offer two notes on study implications. The first note, for clinical researchers: This study highlights the value of qualitative data generally in clinical trial research and specifically the value of progress notes. These data, and other similar clinical documentation, already exist in clinical environments in which effectiveness studies like this one are conducted, and they offer rich data unlike anything typically collected in a clinical trial. Incorporating these data or data like it provides a way to contextualize whether and how clinical trial participants improve. It can, for example, highlight the disparity between symptom improvement and functional improvement or reveal complicated or heterogeneous trajectories difficult to measure with our usual quantitative methods, as in this study.

The second note, for administrators of mental health programs and similar decision makers: This study similarly highlights how data already being collected in these programs have value to help inform program-level decisions. Clinical progress notes have not, to our knowledge, been typically thought of as useful for purposes like program evaluation or quality improvement, but this study indicates that evaluation of these notes reveals meaningful information that could be used for that purpose, for example, to understand which services are most impactful in recovery trajectories or what processes tend to be most successful in preventing or managing acute symptom deterioration.

We understand that the procedure undertaken here is time consuming and may not be feasible for every researcher or administrator to utilize. Although we were satisfied with our analytic decisions because of its rigor and depth despite its associated time cost, we believe that other analytic strategies are likely to be helpful, depending on project aim. For example, researchers may focus on conducting a narrative or content analysis rather than both or focus on just one theme (e.g., symptom severity). Additionally, with advances in machine learning, some elements of this analysis might be done automatically, decreasing time cost significantly. Ultimately, efforts to include qualitative analysis of naturally occurring clinical data have the potential to improve clinical impact.

Limitations

Assessment of clinician progress notes yielded information-dense and personalized data that allowed us to characterize overall trajectories of patients effectively. Other types of notes may yield additional information about patient progress. The clinicians who wrote these notes were not associated with the study, which offers several strengths (e.g., no allegiance to the intervention, no knowledge of group assignment) but also creates a weakness; specifically, that these clinicians were not prompted to discuss cognition in their notes. Because of the overall number of clinicians who contributed to these notes, particularly given that most only contributed to one participant's note set, it is difficult to identify whether a given clinician might have bias impacting these results. Future studies could consider selecting samples based on clinician authors to identify or control for potential bias. Because this study was conducted in a long-stay inpatient facility, results may not generalize to other less intensive treatment settings. Future studies may expand upon these findings through exploration of progress notes from other patient populations and treatment settings. Utilization of nonresearch data offers an important perspective. This first look at progress notes can inform and encourage further qualitative exploration of outcomes. Triangulation of qualitative data with quantitative data facilitates better understanding of the strengths and limitations of both, increasing ability to make useful clinical decisions based on findings.

Conclusion

This study examined a novel methodology and qualitative analysis of clinical progress notes to understand and contextualize recovery trajectories during a clinical trial of TCT among inpatients with schizophrenia and schizoaffective disorder. This analysis added important context and complexity to understanding participants' trajectories and provided support for quantitative findings of TCT. Approximately half of the participants improved over the course of this study, regardless of TCT or TAU group assignment. The majority of participants saw improvements in program participation and social functioning among the majority of participants, whereas symptoms increased for most participants. These notes did not agree with the primary outcome measures used in the parent study for most participants. Better alignment with clinician- and patient-targeted outcomes may strengthen clinical

trials. Qualitative analysis of routinely collected data has the potential to benefit quality of research and programmatic decision making in usual-care settings.

References

Andreasen, N. C. (1984). *Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City, IA: University of Iowa.

Berring, L. L., Pedersen, L., & Buus, N. (2015). Discourses of aggression in forensic mental health: A critical discourse analysis of mental health nursing staff records. *Nursing Inquiry, 22,* 296–305. http://dx.doi.org/10.1111/nin.12113

Brennan, P., & Silman, A. (1992). Statistical methods for assessing observer variability in clinical measures. *British Medical Journal (Clinical Research Ed.), 304,* 1491–1494. http://dx.doi.org/10.1136/bmj.304.6840.1491

Cradock, J., Young, A. S., & Sullivan, G. (2001). The accuracy of medical record documentation in schizophrenia. *Journal of Behavioral Health Services & Research, 28,* 456–465. http://dx.doi.org/10.1007/BF02287775

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62,* 107–115. http://dx.doi.org/10.1111/j.1365-2648.2007.04569.x

Erdmann, T. P., De Mast, J., & Warrens, M. J. (2015). Some common errors of experimental design, interpretation and inference in agreement studies. *Statistical Methods in Medical Research, 24,* 920–935. http://dx.doi.org/10.1177/0962280211433597

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). *Structured clinical interview for DSM–IV–TR axis I disorders, research version, patient ed.* New York, NY: New York Psychiatric Institute.

Fisher, M., Holland, C., Merzenich, M. M., & Vinogradov, S. (2009). Using neuroplasticity-based auditory training to improve verbal memory in schizophrenia. *American Journal of Psychiatry, 166,* 805–811. http://dx.doi.org/10.1176/appi.ajp.2009.08050757

Fisher, M., Loewy, R., Carter, C., Lee, A., Ragland, J. D., Niendam, T., . . . Vinogradov, S. (2015). Neuroplasticity-based auditory training via laptop computer improves cognition in young individuals with recent onset schizophrenia. *Schizophrenia Bulletin, 41,* 250–258. http://dx.doi.org/10.1093/schbul/sbt232

Fisher, M., Mellon, S. H., Wolkowitz, O., & Vinogradov, S. (2016). Neuroscience-informed auditory training in schizophrenia: A final report of the effects on cognition and serum brain-derived neurotrophic factor. *Schizophrenia Research Cognition, 3,* 1–7. http://dx.doi.org/10.1016/j.scog.2015.10.006

Fisher, M., Nahum, M., Howard, E., Rowlands, A., Brandrett, B., Kermott, A., . . . Vinogradov, S. (2017). Supplementing intensive targeted computerized cognitive training with social cognitive exercises for people with schizophrenia: An interim report. *Psychiatric Rehabilitation Journal, 40,* 21–32. http://dx.doi.org/10.1037/prj0000244

Fusch, P. I., & Ness, L. (2015). Are we there yet? Data saturation in qualitative research. *Qualitative Report, 20,* 1408–1416.

Glynn, S. M. (2014). Bridging psychiatric rehabilitation and recovery in schizophrenia: A life's work. *American Journal of Psychiatric Rehabilitation, 17,* 214–224. http://dx.doi.org/10.1080/15487768.2014.935668

Haselden, M., Dixon, L. B., Overley, A., Cohen, A. N., Glynn, S. M., Drapalski, A., . . . Thorning, H. (2018). Giving back to families: Evidence and predictors of persons with serious mental illness contributing help and support to families. *Community Mental Health Journal, 54,* 383–394. http://dx.doi.org/10.1007/s10597-017-0172-1

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15,* 1277–1288. http://dx.doi.org/10.1177/1049732305276687

Humble, A. M. (2009). Technique triangulation for validation in directed content analysis. *International Journal of Qualitative Methods, 8,* 34–51. http://dx.doi.org/10.1177/160940690900800305

Kuhnigk, O., Slawik, L., Meyer, J., Naber, D., & Reimer, J. (2012). Valuation and attainment of treatment goals in schizophrenia: Perspectives of patients, relatives, physicians, and payers. *Journal of Psychiatric Practice, 18,* 321–328. http://dx.doi.org/10.1097/01.pra.0000419816.75752.65

Lindhiem, O., Bennett, C. B., Orimoto, T. E., & Kolko, D. J. (2016). A meta-analysis of personalized treatment goals in psychotherapy: A preliminary report and call for more studies. *Clinical Psychology: Science and Practice, 23,* 165–176. http://dx.doi.org/10.1111/cpsp.12153

Liu, N. H., Choi, K.-H., Reddy, F., & Spaulding, W. D. (2011). Heterogeneity and the longitudinal recovery of functioning during inpatient psychiatric rehabilitation for treatment-refractory severe mental illness. *American Journal of Psychiatric Rehabilitation, 14,* 55–75. http://dx.doi.org/10.1080/15487768.2011.546293

March, C. A., Scholl, G., Dversdal, R. K., Richards, M., Wilson, L. M., Mohan, V., & Gold, J. A. (2016). Use of electronic health record simulation to understand the accuracy of intern progress notes. *Journal of Graduate Medical Education, 8,* 237–240. http://dx.doi.org/10.4300/JGME-D-15-00201.1

Miller, J. M., & Velanovich, V. (2010). The natural language of the surgeon's clinical note in outcomes assessment: A qualitative analysis of the medical record. *American Journal of Surgery, 199,* 817–822. http://dx.doi.org/10.1016/j.amjsurg.2009.06.037

Morrens, M., Dewilde, B., Sabbe, B., Dom, G., De Cuyper, R., & Moggi, F. (2011). Treatment outcomes of an integrated residential programme for patients with schizophrenia and substance use disorder. *European Addiction Research, 17,* 154–163. http://dx.doi.org/10.1159/000324480

Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., . . . Marder, S. R. (2008). The MATRICS Consensus Cognitive Battery, part 1: Test selection, reliability, and validity. *American Journal of Psychiatry, 165,* 203–213. http://dx.doi.org/10.1176/appi.ajp.2007.07010042

Sabbag, S., Twamley, E. M., Vella, L., Heaton, R. K., Patterson, T. L., & Harvey, P. D. (2011). Assessing everyday functioning in schizophrenia: Not all informants seem equally informative. *Schizophrenia Research, 131,* 250–255. http://dx.doi.org/10.1016/j.schres.2011.05.003

Silberstein, J., & Harvey, P. D. (2019). Cognition, social cognition, and self-assessment in schizophrenia: Prediction of different elements of everyday functional outcomes. *CNS Spectrums, 24,* 88–93. http://dx.doi.org/10.1017/S1092852918001414

Sosulski, M. R., Buchanan, N. T., & Donnell, C. M. (2010). Life history and narrative analysis: Feminist methodologies contextualizing Black women's experiences with severe mental illness. *Journal of Sociology and Social Welfare, 37,* 29.

Spaulding, W. D., Sullivan, M. E., & Poland, J. S. (2003). *Treatment and rehabilitation of severe mental illness.* New York: Guilford Press.

Stern, S., Doolan, M., Staples, E., Szmukler, G. L., & Eisler, I. (1999). Disruption and reconstruction: Narrative insights into the experience of family members caring for a relative diagnosed with serious mental illness. *Family Process, 38,* 353–369. http://dx.doi.org/10.1111/j.1545-5300.1999.00353.x

Thomas, M. L., Bismark, A. W., Joshi, Y. B., Tarasenko, M., Treichler, E. B. H., Hochberger, W. C., . . . Light, G. A. (2018). Targeted cognitive training improves auditory and verbal outcomes among treatment refractory schizophrenia patients mandated to residential care. *Schizophrenia Research, 202,* 378–384.

Treichler, E., & Spaulding, W. (2015). Symptom improvement and maintenance of psychiatric stability in a residential psychiatric rehabilitation setting. *Schizophrenia Bulletin, 41,* S193.

Treichler, E. B. H., Thomas, M., Bismark, A., Hochberger, W. C., Tarasenko, M., Nungaray, J., . . . Light, G. A. (2019). Divergence of

self-reported and performance-based cognitive gains following cognitive training in schizophrenia. *Schizophrenia Research, 210,* 215–220.

Verma, S., Subramaniam, M., Abdin, E., Poon, L. Y., & Chong, S. A. (2012). Symptomatic and functional remission in patients with first-episode psychosis. *Acta Psychiatrica Scandinavica, 126,* 282–289. http://dx.doi.org/10.1111/j.1600-0447.2012.01883.x

Yarborough, B. J. H., Yarborough, M. T., Janoff, S. L., & Green, C. A. (2016). Getting by, getting back, and getting on: Matching mental health services to consumers' recovery goals. *Psychiatric Rehabilitation Journal, 39,* 97–104. http://dx.doi.org/10.1037/prj0000160

Ye, C., & Fabbri, D. (2018). Extracting similar terms from multiple EMR-based semantic embeddings to support chart reviews. *Journal of Biomedical Informatics, 83,* 63–72. http://dx.doi.org/10.1016/j.jbi.2018.05.014