REPORT / FIRST EPISODE PSYCHOSIS

AUGUST 31, 2018

Initial Evaluation of First Episode Psychosis Early Intervention Programs in Texas, 2018

Submitted to Texas Health and Human Services Commission

The University of Texas at Austin
Texas Institute for Excellence in Mental Health
Steve Hicks School of Social Work

CONTACT

Texas Institute for Excellence in Mental Health
Steve Hicks School of Social Work
The University of Texas at Austin
1717 West 6th Street, Suite 335
Austin, Texas 78703

Phone: (512) 232-0616 | Fax: (512) 232-0617
Email: txinstitute4mh@austin.utexas.edu
sites.utexas.edu/mental-health-institute

CONTRIBUTORS/PROJECT LEADS

Mark D. Kramer, Ph.D.

Molly A. Lopez, Ph.D.

ACKNOWLEDGEMENT

This work is funded through a contract with the Texas Health and Human Services Commission (HHSC). The contents are solely the responsibility of the authors and do not necessarily represent the official views of Texas HHSC.

Recommended Citation: Kramer, M. D., & Lopez, M. A. (2018). Initial evaluation of first episode psychosis early intervention programs in Texas, 2018. Texas Institute for Excellence in Mental Health, School of Social Work, University of Texas at Austin.

CONTENTS

Initial Evaluation – First Episode Psychosis Program

Background

Schizophrenia and related disorders are some of the most burdensome of mental health and substance use disorders in terms of years lived with disability (Whiteford et al., 2013). The first episode of psychosis (FEP) often remains undetected for months to years (median 74 weeks; Addington et al., 2015). A number of coordinated specialty care services and conventional therapeutic forms of care have been utilized in the treatment of FEP, often introduced as a comprehensive treatment program. Specialized FEP programs appear to help prevent relapse (Alvarez-Jimenez, Parker, Hetrick, McGorry, & Gleeson, 2011), reduce readmission and treatment dropout rates (Craig et al., 2004), and reduce the use of supportive housing and hospitalization (Bertelsen et al., 2008). Additionally, these programs have been shown to reduce positive and negative symptoms of psychosis and increase functional improvement relative to treatment-as-usual (Harvey, Lepage, & Malla, 2007; Norman et al., 2011).

The Texas Health and Human Services Commission (HHSC) has contracted with investigators from the Texas Institute for Excellence in Mental Health (TIEMH) at the University of Texas at Austin to conduct a multi-year, independent evaluation of the effectiveness of FEP programs as implemented in mental health agencies across the state. Each of the FEP programs in Texas has adopted the same model for intervention, the OnTrackNY model, which is a specific instantiation of the Recovery After an Initial Schizophrenia Episode (RA1SE) Connection Program. In fiscal year 2017, TIEMH investigators conducted a preliminary evaluation of the effectiveness of the program through qualitative interviews with providers and an initial empirical view of trajectories of clinical symptomatology in FEP participants vs. treatment-as-usual groups. Provider reports regarding the success of implementation of the program and clinical outcomes of their clients were uniformly positive. Focusing primarily on clinical symptomatology, empirical results were suggestive of more rapid stabilization and less evidence of worsening of symptoms of psychosis and other broadband domains of symptomatology in FEP participants than in the control groups (Kramer & Lopez, 2017). The empirical findings were preliminary, however, in that the analyses relied only on existing measurement tools, the Child and Adolescent Needs and Strengths (CANS) and Adult Needs and Strengths Assessment (ANSA), which did not allow for a more targeted examination of key symptom domains (e.g., positive and negative symptoms).

The current report provides an update on the status of the forthcoming comprehensive evaluation protocol and efforts toward collection of data across sites. Additionally, we provide a preliminary analysis of the effectiveness of FEP programs in Texas through the lens of psychosocial functioning. As noted, prior research has indicated that specialized treatment programs for FEP result in improved functioning in program participants relative to standard care (Harvey et al., 2007; Norman et al., 2011). However, neither of these studies focused exclusively on FEP programs as implemented in state-funded mental health clinics in the United States and both relied upon the Global Assessment of Functioning (GAF; American Psychiatric Association, 1994) scale. The validity of the GAF has been questioned, in part because the scale appears to conflate symptoms of mental illness with functioning (e.g., Aas, 2010). Both the CANS and ANSA include 15 items assessing various areas of functioning, and, while these domains are diverse, do not appear to conflate symptoms with functioning. Thus, the current evaluation additionally serves to attempt a conceptual replication of prior work showing superiority of FEP programs in improving overall psychosocial functioning using an arguably purer functioning measure and more sophisticated analytic techniques (e.g., latent class growth analysis).

Purpose and Methodology

The purposes of the current report are two-fold. First, progress regarding the roll out of a comprehensive, empirical, longitudinal program evaluation study will be described. Second, a preliminary evaluation of the effectiveness of early onset psychosis programs will be conducted with a primary focus on life domain functioning as a pertinent outcome. In this evaluation, the team will utilize administrative data, the CANS and ANSA, to establish preliminary outcome information on those served by the FEP program since its initiation in Texas. On average, the CANS is administered to adolescents once every three months and the ANSA to young adults once per six months. Outcomes will be compared to young people in treatment with similar characteristics, identified through propensity matching, to examine the benefit of the FEP program compared to treatment-as-usual. Since this evaluation will be limited to administrative data, the findings should be considered exploratory in nature.

Program Evaluation Progress. The process of development of the comprehensive evaluation protocol will be described, followed by a brief summary of the protocol itself. The protocol approval process, from both the Texas HHSC and various institutional review boards (IRB), will be articulated. Finally, we will report the status of progress of the evaluation protocol and data collection procedures at each specific state-funded mental health agency in Texas providing FEP services.

Initial Outcome Evaluation. The evaluation team has access to CANS and ANSA data for the fiscal years 2014 through 2018. Approximately 500 adults between the ages of 18 and 30, and 65 young adults between the ages of 15 and 17 participated in FEP programs across the state during this time period. Both the CANS and ANSA include items that pertain to psychosocial functioning in the Life Domain Functioning sections of the assessments. Each domain of functioning is assessed with one item. Given the disparate range in content of these sets of items, factor analyses were used to appropriately weight the contributions of specific items to total functioning domain scores. We utilize the first four waves of CANS data and the first eight waves of ANSA data as sample sizes dwindled beyond these time points in adolescents and young adults, respectively. Through analysis of functioning scores across these waves of assessment using latent class growth analysis, we are interested in clarifying trajectories of psychosocial functioning over time. The modeling techniques of factor analysis and latent class growth analysis are described in more detail below.

Program Evaluation Progress

Development of the Protocol. TIEMH investigators have been granted access to CANS data for adolescents, ANSA data for young adults, and service encounter data spanning the 2013 through 2018 fiscal years. In the initial evaluation of the effectiveness of FEP programs in Texas, Kramer and Lopez (2017) determined through qualitative interviews with providers across the state that there was limited standardization of outcome measurement beyond the CANS and ANSA across sites. It was recommended that a comprehensive assessment of the program's effectiveness utilize additional reliable and valid instruments to supplement the CANS and ANSA data. This past fiscal year, TIEMH investigators identified a list of pertinent outcome domains to be assessed and further identified instruments in the literature that have been relied upon consistently to effectively and efficiently measure these domains. After identification of the measures, TIEMH held two conference calls with representatives of HHSC and Team Leads/Program Managers across FEP sites to elicit feedback regarding the selection of specific tools and the breadth of outcome domains targeted in the protocol. Following these meetings, TIEMH investigators developed a protocol of instruments and a data collection plan regarding how often specific self-report, clinician-report, and investigator-rated tools would be administered. This protocol was

submitted to HHSC for review. HHSC recommended specific changes, which were addressed, and the revised protocol was approved by HHSC. TIEMH then acquired the proposed rating instruments and developed online Qualtrics surveys in English and Spanish, and alternative paper-and-pencil based surveys.

The Protocol. The FEP program evaluation protocol includes a comprehensive battery of intake and outcome assessments to be administered over the course of the next 36 months. The battery includes self-report scales administered to participants in the FEP program and nominated family members/caregivers, clinician-rated scales, and one investigator-rated scale. Self-report scales are to be administered to FEP participants at study enrollment, 6 months, 12 months, and at either 18 months or discharge from the program, whichever occurs first. Clinicians are to complete enrollment and discharge forms, and brief monthly assessments. The vast majority of these data are expected to be collected using Qualtrics online administration on tablets provided by TIEMH. Finally, investigators will visit each participating site across Texas for the purpose of conducting a structured interview assessing fidelity to the OnTrackNY training model.

Participant self-report measures. Administered at enrollment (T1), six (T2) and 12 months (T3), and 18 months/discharge (T4) will be the *World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0)*, *Patient Health Questionnaire 9-Item Depression (PHQ-9) scale*, *Personal Well-being Index (PW-BI)*, *Patient Health Questionnaire 7-Item Generalized Anxiety Disorder (GAD-7) scale*, SCORE-15 Index of Family Functioning and Change* scale**,* and *Mental Health Confidence Scale (MHCS)*.* The full self-report battery is approximately 100 items, whereas exclusion of optional measures (denoted by *) reduces the length by half.

Clinician-rated measures. An investigator-developed *Enrollment Form* will be rated by clinician's at T1, and includes basic client demographic information, history of illness items, and symptomatology questionnaires. Every following month, providers are to complete an investigator-developed *Event Tracking Form* that records pertinent life events such as job acquisition/loss and crisis service utilization. Included also in the monthly ratings are the *Positive Symptom Rating Scale (PSRS) version 5.0, Brief Negative Symptom Assessment Scale (BNSA) version 5.0, Alcohol Use Scale (AUS)*, and Drug Use Scale (DUS)**. Finally, at 18 months or discharge (T4), clinicians are to rate the symptom scales and an investigator-developed *Discharge Form* that includes items such as reason for discharge and school/employment status.

Nominated family member/caregiver-rated measure. FEP participants have the option of nominating a family member, loved one, or caregiver to complete a survey at T1, T2, T3, and T4. The *Burden Assessment Scale (BAS)* measures the impact that the participant's illness has on the family member.

Investigator-rated measure. OnTrackNY developed a scale to assess fidelity to the training model. Items are rated by investigators through conduct of a site visit that includes interviews with all team members, two participants in the FEP program, and one family member. The visit also includes attending a weekly clinical staffing, medical chart review, and review of supplemental data such as quarterly reports of aggregated data to HHSC. Each item is rated on a scale of "Unacceptable (0)," "Acceptable (1)," or "Exceptional (2)." Thus, scores range from 0 to 200.

Institutional Review Boards. Following approval of the program evaluation protocol by HHSC, TIEMH investigators submitted the evaluation plan to the University of Texas at Austin's IRB. The evaluation was deemed by the University of Texas as "non-human subjects research" as participants would not be randomized to treatment vs. control groups and the self-report forms would pose no additional risks to them than standard assessments in clinical care. Four of the agencies additionally have their own internal IRB's and approval processes. These include *The Harris Center for Mental Health and IDD*, *Metrocare Services*, *Integral Care*, and *MHMR Tarrant County*.

Site-Specific Status Updates

The Harris Center for Mental Health and IDD. This center serves the Houston area. The agency has agreed to participate in the program evaluation. This agency has its own IRB, and the program evaluation protocol has been approved. A site visit for fidelity assessment and data collection training was conducted in July 2018. The site's overall score on the fidelity assessment was 112.

Metrocare Services. This agency serves the Dallas area. The agency has agreed to participate in the program evaluation. This agency has its own IRB, and the program evaluation protocol has been recently approved. A site visit for fidelity assessment and data collection training will be conducted in the near future.

Bluebonnet Trails Community Services. This agency serves counties surrounding Austin. The agency has agreed to participate in the program evaluation. At last correspondence, agency administrators were developing a document for TIEMH investigators ensuring HIPAA compliance for their participants.

Burke Center. This agency serves Lufkin and surrounding areas. The agency has not agreed to participate in the program evaluation, but may do so at a later date.

Texas Panhandle Centers. This agency serves Amarillo and surrounding areas. The agency has not agreed to participate in the program evaluation, but wanted to be kept apprised of the participation status of other sites.

Integral Care. This agency serves the Austin area. The agency has agreed to participate in the program evaluation. This agency has its own IRB. The protocol is still under review.

Tropical Texas Behavioral Health. This agency serves the southern-most regions of the state. The agency has agreed to participate in the program evaluation. A site visit for fidelity assessment and data collection training was conducted in August 2018. The site's overall score on the fidelity assessment was 103.

MHMR Tarrant County. This agency serves the Fort Worth area. The agency has agreed to participate in the program evaluation. This agency has its own IRB, and expedited review documents have been submitted. The IRB determined that the protocol should come under full review, and has requested additional documentation from TIEMH investigators. We are currently in the process of developing this documentation.

Emergence Health Network. This agency serves the El Paso area. The agency has agreed to participate in the program evaluation. A site visit for fidelity assessment and data collection training was conducted in July 2018. The site's overall score on the fidelity assessment was 93.

The Center for Health Care Services. This agency serves the San Antonio area. The agency has not agreed to participate in the program evaluation owing to recent staff turnover. This site may participate in the evaluation in the future, however.

Initial Outcome Evaluation

Baseline Characteristics. Propensity matching of participants allows for comparison of two groups that were not randomly selected in an attempt to estimate the effect of an intervention by accounting for covariates predicting receipt of treatment. The covariates that we included were gender, age, and Time 1 scores on the Life Domain Functioning items of either the CANS or ANSA for younger and older participants, respectively. Descriptive

Table 1. *Descriptive Statistics for FEP Participants and Controls in Young Adulthood*

	FEP Participants		*Controls*		*t-test*
Young Adults (N = 509 per)	*M*	*SD*	*M*	*SD*	*p*
Sex	1.32	0.47	1.32	0.47	1.000
Age	22.38	3.46	22.97	3.56	.007*
T1 Life Domain Functioning	10.33	6.07	8.71	6.57	.000*
Physical/Medical	0.33	0.57	0.36	0.61	.531
Family	0.88	0.84	0.76	0.89	.040
Employment	2.00	1.27	1.71	1.34	.003*
Social	1.22	0.98	0.94	0.97	.000*
Recreational	1.13	1.05	0.83	1.00	.000*
Intellectual Development	0.08	0.30	0.14	0.46	.013*
Sexuality	0.05	0.26	0.04	0.24	.676
Living Skills	0.48	0.68	0.35	0.65	.003*
Residential Stability	0.41	0.72	0.41	0.77	.921
Legal	0.41	0.73	0.46	0.76	.257
Sleep	1.22	0.91	1.03	0.98	.001*
Self-Care	0.60	0.84	0.41	0.66	.000*
Decision Making	0.93	0.84	0.78	0.81	.005*
Recovery Involvement	0.61	0.78	0.51	0.77	.045
Transportation	0.44	0.66	0.41	0.73	.459
T2 Life Domain Functioning	10.59	5.42	10.07	6.01	.232
T3 Life Domain Functioning	10.50	5.58	9.93	6.15	.274
T4 Life Domain Functioning	10.09	5.62	9.91	6.11	.767
T5 Life Domain Functioning	9.41	5.48	9.48	6.11	.916
T6 Life Domain Functioning	8.97	5.13	9.85	6.29	.229
T7 Life Domain Functioning	8.92	5.36	8.99	6.26	.942
T8 Life Domain Functioning	8.30	5.37	9.40	6.71	.267

Note. Higher values of Life Domain Functioning items and summed scale scores indicate lower levels of functioning. *Indicates significance at the p<.01 level.

statistics for the samples and comparisons between FEP participants and controls are presented in Tables 1 and 2. For young adults, propensity matching was successful for sex, but less successful for age (i.e., FEP participants were slightly younger than controls) and Time`1 Life Domain Functioning summed scaled scores (i.e., FEP participants had worse functioning at baseline than controls). As can be seen at the item level, where differences were significant, FEP participants showed generally worse functioning for those items than propensity-matched controls. However, Life Domain Functioning summed scaled scores were not significantly different between the two groups at Times 2 through 8, importantly for subsequent latent class growth analyses. For adolescents, propensity matching was successful for sex, age, and baseline Life Domain Functioning summed scaled scores. The only significant differences between FEP participants and controls were at the Time 1 psychosocial functioning item level. Participants were functioning worse at baseline than controls in terms of sleep, and functioning better than controls in terms of legal issues. Adolescent FEP participants did not differ from controls in terms of Time 2, Time 3, or Time 4 Life Domain Functioning summed scaled scores, again important for subsequent latent class growth analyses.

Table 2. *Descriptive Statistics for FEP Participants and Controls in Adolescence*

	FEP Participants		*Controls*		*t-test*
Adolescents (N = 67 per)	*M*	*SD*	*M*	*SD*	*p*
Sex	1.35	0.48	1.36	0.48	.959
Age	16.10	1.26	16.24	0.76	.467
T1 Life Domain Functioning	10.64	5.97	9.94	7.87	.585
Family	1.21	0.84	1.26	1.03	.765
Living Situation	0.84	0.77	0.89	0.99	.761
School	1.34	0.98	1.32	1.08	.887
Social	1.22	0.96	1.08	0.97	.393
Recreational	1.14	1.07	0.80	0.98	.071
Developmental	0.11	0.31	0.24	0.56	.101
Communication	0.21	0.55	0.29	0.63	.450
Judgment	0.90	0.83	1.06	0.99	.323
Job	0.69	1.11	0.41	0.85	.403
Legal	0.26	0.61	0.56	0.88	.030
Medical	0.24	0.57	0.23	0.58	.892
Physical	0.19	0.44	0.14	0.46	.512
Sexual	0.09	0.28	0.11	0.43	.767
Sleep	1.21	0.97	0.80	0.93	.020
Independent Living	1.24	1.14	0.84	1.04	.115
T2 Life Domain Functioning	9.21	4.81	9.28	6.00	.945
T3 Life Domain Functioning	7.70	5.25	8.36	5.38	.607
T4 Life Domain Functioning	7.37	5.38	8.67	5.10	.377

Note. Higher values of Life Domain Functioning items and summed scale scores indicate lower levels of functioning.

Demographics. Adult participants in the FEP program (*N* = 509) ranged in age from 16 to 31, with an average of 22.38 (*SD* = 3.46); 67.9% were male (*n* = 345) and 32.1% were female (*n* = 163). While 16- and 17-year-olds are generally served within the child system, the adult service system included some adolescent participants. Matched adults in the control group (*N* = 509) ranged in age from 18 to 30, with an average age of 22.97 (*SD* = 3.56); 67.9% were male (*n* = 345) and 32.1% were female (*n* = 163). The ethnicity of adult participants in the FEP program included 33.0% Caucasians, 28.3% Hispanic Americans, 33.4% African Americans, 1.4% Asian Americans, 0.2% Native Americans, 2.4% Mixed Race, and 0.2% Pacific Islanders. The ethnicity of matched adult controls included 49.8% Caucasians, 23.0% Hispanic Americans, 21.9% African Americans, 0.8% Asian Americans, 0.3% Native Americans, 2.9% Mixed Race, and 0.1% Pacific Islanders. Adolescents in the FEP program (*N* = 67) ranged in age from 13 to 17, with an average age of 16.10 (*SD* = 1.26); 64.6% were male (*n* = 42) and 35.4% were female (*n* = 23). The ethnicity of adolescent participants in the FEP program included 41.8% Caucasians, 25.4% Hispanic Americans, 22.4% African Americans, 1.5% Native Americans, and 6.0% Mixed Race. Matched adolescents in the control group (*N* = 67) ranged in age from 15 to 17, with an average age of 16.23 (*SD* = 0.76); 64.2% were male (*n* = 43) and 35.8% were female (*n* = 24). The ethnicity of matched adolescent controls included 49.3% Caucasians, 32.8% Hispanic Americans, 13.4% African Americans, and 4.5% Mixed Race.

Factor Structures of Life Domain Functioning Items. The ANSA and CANS each assess a range of domains of psychosocial functioning. Each of these Life Domain Functioning items is scored in the same way, on a zero to three scale assessing the necessity of attention to each of these areas. Higher scores indicate worse functioning.

The internal consistency of the ANSA functioning items (k = 15) ranged from .74 to .78 in the FEP participants and controls across the eight waves of assessment. The internal consistency of the CANS functioning items (k = 15) ranged from .74 to .83 in the adolescent FEP participants and controls across four waves of assessment. These are reasonably good levels of reliability. However, given fairly disparate content of the items and limited evidence for the test-retest reliability of the ANSA and CANS Life Domain Functioning summed scores over time, we utilized factor analyses and formal longitudinal measurement invariance modeling to estimate overall scores for the ANSA and CANS psychosocial functioning domains. The fit of confirmatory factor analysis (CFA) models was evaluated using a number of criteria, including the root mean square error of approximation (RMSEA) and comparative fit index (CFI). Following identification of well-fitting structural models, the factor scores were estimated from the model using regression-based approaches and these factor scores were then utilized in subsequent analyses.

Trajectories of Life Domain Functioning. Latent class growth analysis (LCGA; Jung & Wickrama, 2008) is a statistical technique that identifies latent subpopulations of individuals within the overall trajectory of a variable over time. The technique characterizes the number and nature of latent trajectories that comprise the overall trajectory of the full sample. In other words, the overall trajectory may show improvements in functioning over time, however, within that trajectory are others who may worsen over time, stay the same, or show marked improvement. The method does not assume a specific number or shape of trajectories, but the estimated models are judged in terms of fit to the observed data in much the same way as CFA models. Lower values of the Bayesian information criterion (BIC), sample-size adjusted BIC (ssaBIC), higher values of entropy, which represents the accuracy of classification of individuals to particular trajectories, and class sizes were taken into account in selecting the best-fitting models. Additionally, significant Vuong-Lo-Mendell-Rubin (VLMR) likelihood ratio test and parametric bootstrapped likelihood ratio test (BLRT) statistics indicate that the number of classes in the model being evaluated results in improvement of fit beyond the prior model with lesser classes. In the current evaluation, we compared trajectories between individuals in the FEP program with propensity-matched controls in the treatment-as-usual group.

Factor Analyses. Structural modeling analyses were conducted in *Mplus* (Muthén & Muthén, 1998-2012). The one-factor model of ANSA Life Domain Functioning items measured at the first wave of assessment fit the data well (RMSEA = .057; CFI = .957). The well-fitting ANSA 1-factor functioning model (Figure 1) yielded correlations between summed scale scores and estimated factor scores from the strong measurement invariance model (described below) ranging between r = .94 and r = .97 across the eight waves of assessment. The one-factor model of CANS Life Domain Functioning items (Figure 2) measured at the first wave of assessment fit the data less well, though adequately (RMSEA = .091; CFI = .910). Correlations between summed scale scores and estimated factor scores from the strong measurement invariance model (described below) ranged between r = .91 and r = .96 across the four waves of assessment.

While the one-factor model of functioning fit the data in young adults and in adolescents at the baseline assessment, cross-sectional analyses are limited in evaluating the fit of the model over time. In addition, it is unknown which parameters of the model can be constrained over time without conducting formal confirmatory factorial invariance modeling (Meredith, 1993). There are several types of measurement invariance, but for the current purpose we evaluated the fit of the strong invariance model, which specifies that factor loadings and intercepts are equal over time. This model is interpreted to mean that change in endorsement of specific items over time is accounted for by changes at the level of the latent factors.

The strong invariance model was fit to ANSA data provided by young adults in the FEP program and controls between fiscal years 2014 and 2018, and the first eight ANSA assessments were used. The sample sizes at time points one to eight were N = 1,017, n = 708, n = 548, n = 431, n = 340, n = 264, n = 207, and n = 170, respectively.

The model indicated that there was change in the psychosocial functioning factor over time in the overall sample (i.e., FEP participants and propensity-matched controls). In standard deviation units, the Functioning factor

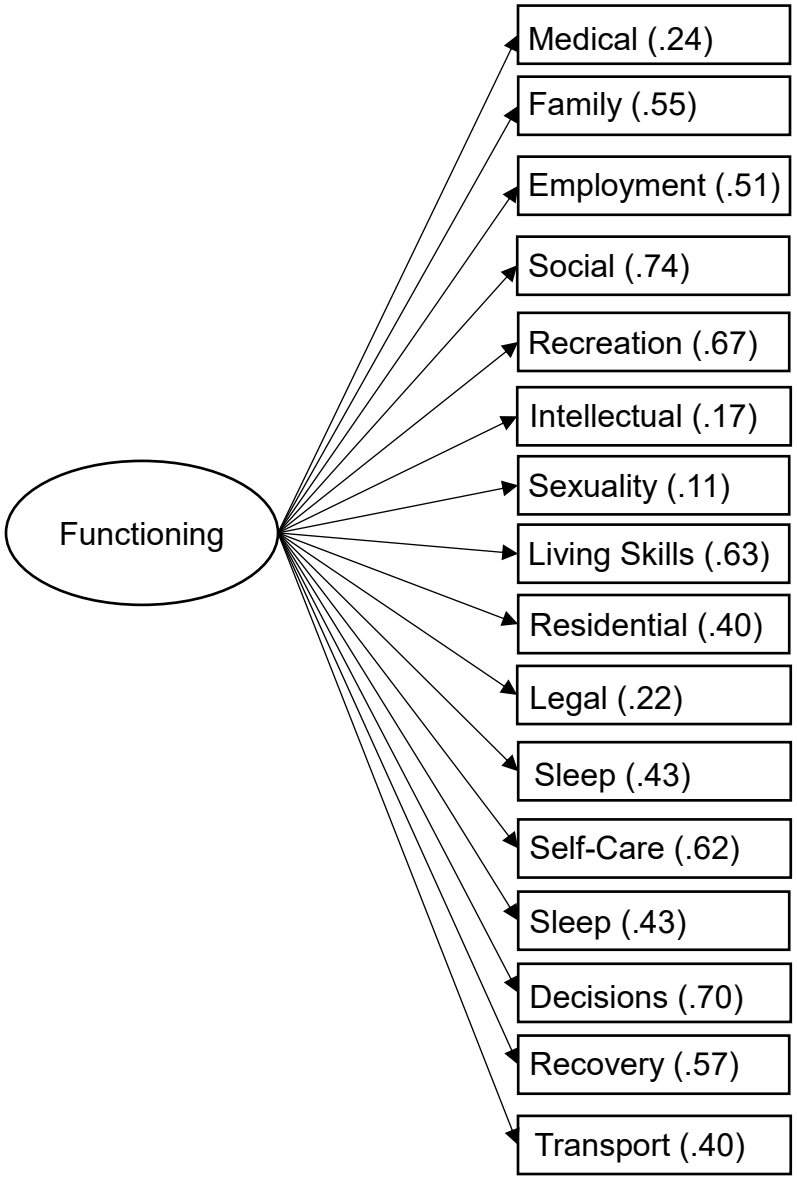

Figure 1. 1-factor ANSA functioning model.

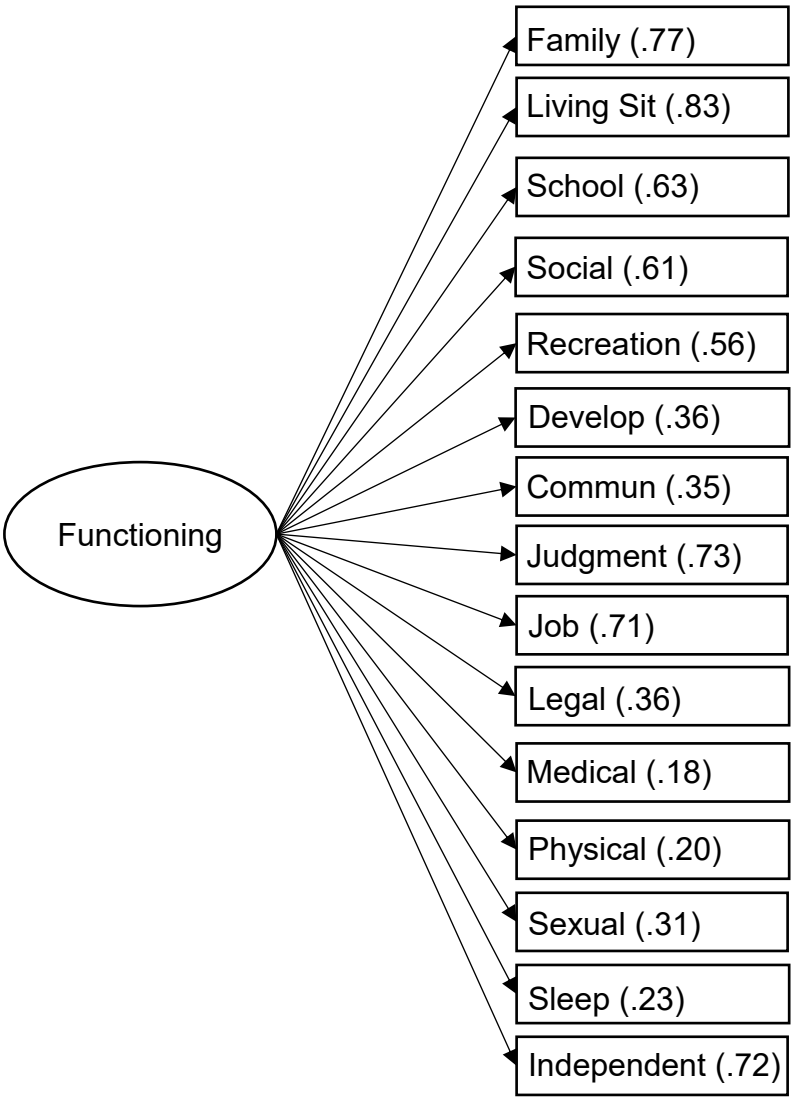

Figure 2. 1-factor CANS functioning model.

increased from Time 1 to Time 2 by 0.15 (p<.001) and increased from baseline to Time 3 by 0.13 (p=.019). Relative to the Time 1 assessment, there were no further significant increases or decreases in the Functioning factor at the following five time points. Notably, in the overall sample, the increases at Times 2 and 3 suggest early worsening of functioning, with stabilization back to baseline levels over the remaining period. This early worsening may also be explained by an increase in the accuracy of the ratings, as rapport is developed between the young adult and provider and additional information on functioning is obtained.

The strong invariance model was fit to CANS data provided by adolescents in the FEP program and controls between fiscal years 2014 and 2018, and the first four ANSA assessments were used. The sample sizes at time points one to four were N = 133, n = 116, n = 70, and n = 50, respectively. The model indicated that there was no significant change in the psychosocial functioning factor over time in the overall sample (i.e., FEP participants and propensity-matched controls). Although the changes were not significant, in standard deviation units, the Functioning factor decreased from Time 1 to Time 2 by 0.09 (p=.395), decreased from baseline to Time 3 by 0.27 (p=.107), and decreased from baseline to Time 4 by 0.34 (p=.070). Notably, in the overall sample, the decreases, though not statistically significant perhaps owing to the small sample sizes, suggest modest improvements in functioning over time in the overall adolescent sample.

Latent Class Growth Analyses. One advantage of LCGA is that it clusters individuals into latent class trajectories that can differ across groups and that do not simply reflect the overall mean-level changes in the factors. In other words, LCGA trajectories account for heterogeneity in change over time. We utilized factor scores from each of the eight time points estimated in the ANSA sample strong invariance model in the estimation of two- to five-trajectory models. The two- to five-trajectory models were tested for the ANSA Functioning factor FEP participant sample (N = 509) and separately in age, gender, and baseline Life Domain Functioning propensity-matched controls (N = 509). We selected each best-fitting model based on the lowest BIC values, balancing high entropy and significant VLMR and BLRT values with adequate sample sizes in each trajectory. Fit statistics for the LCGA in the FEP and control groups are presented in Table 3. As bolded in the table, we selected the three-class models as best fitting for both the FEP and control groups. Trajectories for the FEP participants and controls are presented in Figures 3 and 4.

Table 3. *Fit statistics for ANSA Latent Class Growth Analysis Models in First Episode Psychosis Participants (N = 509) and Controls (N = 509)*

	BIC	*ssaBIC*	*Entropy*	*VLMR p*	*BLRT p*
FEP Participants					
2-class	26423	26366	0.79	.136	0.000
3-class	**26224**	**26152**	**0.89**	**.034**	**0.000**
4-class	26179	26096	0.74	.782	0.000
5-class	26049	25953	0.78	.306	0.000
Controls					
2-class	23643	23586	0.62	.247	0.000
3-class	**23418**	**23348**	**0.92**	**.379**	**0.000**
4-class	23248	23166	0.83	.503	0.000
5-class	23134	23039	0.85	.835	0.000

The 3-class LCGA models of Functioning in both young adult FEP participants and controls yielded recovery groups (i.e., starting low in functioning and improving over time), stable groups comprising the majority of the samples, and worsening groups (i.e., started high in functioning, worsened, and began heading toward the means). In FEP participants, the compositions of the classes were 7.4% Recovery, 79.7% Stable, and 12.9% Worsening. In

controls, the compositions of the classes were 11.3% Recovery, 83.9% Stable, and 4.8% Worsening (Figure 4.). Thus, in terms of proportions of individuals in FEP programs vs. treatment-as-usual comprising the trajectories, controls fared better than FEP participants, showing a higher percentage of individuals in the Recovery class and a lower percentage of individuals in the Worsening class than the FEP group. However, there was evidence that the FEP group recovered more pronouncedly, with the curve dipping below the mean for a period of time, whereas the control Recovery trajectory only approached the mean.

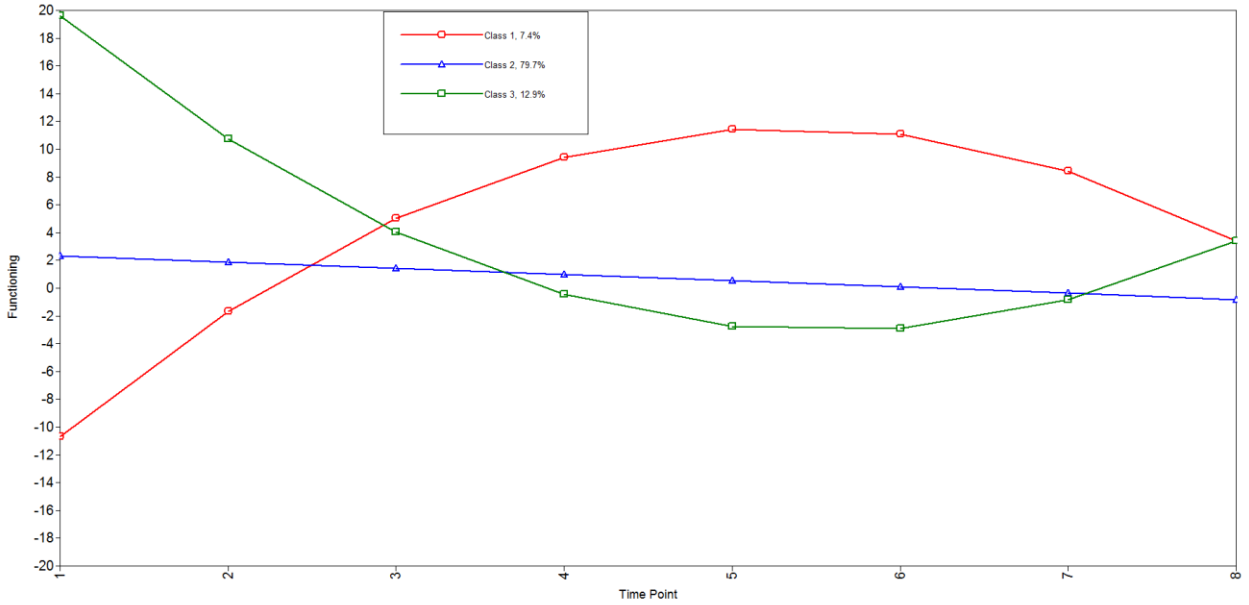

Figure 3. Latent class growth analysis of ANSA Functioning factor over time in FEP participants. Note that higher scores indicate worse functioning. Mean is 0 and standard deviation is 100.

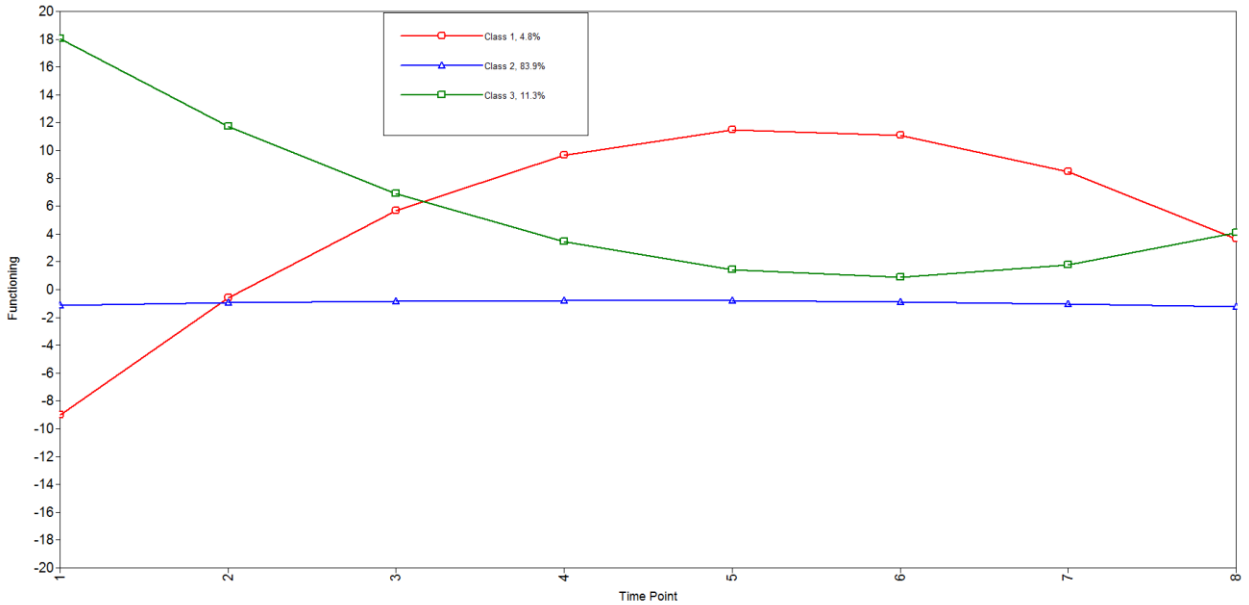

Figure 4. Latent class growth analysis of ANSA Functioning factor over time in propensity-matched controls. Note that higher scores indicate worse functioning. Mean is 0 and standard deviation is 100.

We utilized factor scores from each of the four time points estimated in the CANS sample strong invariance model in the estimation of two- to five-trajectory models. The two- to five-trajectory models were tested for the CANS Functioning factor FEP participant sample (*N* = 67) and separately in age, gender, and baseline Life Domain Functioning propensity-matched controls (*N* = 67). We selected each best-fitting model based on the lowest BIC values, balancing high entropy and significant VLMR and BLRT values with adequate sample sizes in each trajectory. Fit statistics for the LCGA in the FEP and control groups are presented in Table 4. As bolded in the table, we selected the three-class models as best fitting for both the FEP and control groups. Trajectories for the FEP participants and controls are presented in Figures 5 and 6.

Table 4. *Fit statistics for CANS Latent Class Growth Analysis Models in First Episode Psychosis Participants (N = 67) and Controls (N = 67)*

	BIC	*ssaBIC*	*Entropy*	*VLMR p*	*BLRT p*
FEP Participants					
2-class	2382	2338	1.00	.017	0.040
3-class	**2379**	**2322**	**0.83**	**.051**	**0.013**
4-class	2384	2315	0.90	.422	0.013
5-class	2393	2311	0.86	.726	0.667
Controls					
2-class	2570	2526	0.83	.089	0.000
3-class	**2580**	**2524**	**0.86**	**.507**	**1.000**
4-class	2586	2517	0.82	.419	1.000
5-class	2566	2484	0.93	.034	0.000

The 3-class LCGA model of Functioning in the adolescent FEP participants (Figure 5.) yielded a Recovery group (i.e., starting very low in functioning in improving over time; 1.5%), a High Stable group (i.e., starting high in functioning and remaining stable; 31.9%), and a Modest Recovery group (i.e., started somewhat low in functioning and stabilizing toward the mean; 66.7%). In controls (Figure 6.), the 3-class LCGA model of Functioning yielded a Recovery group (i.e., starting very low in functioning and improving over time; 38.7%), a High Stable group (i.e., starting high in functioning and remaining stable; 56.7%), and a Low Stable trajectory (i.e., starting low in functioning and staying low; 4.6%). Thus, the FEP sample included a marked Recovery trajectory, though the proportion was quite small (1.5%), and a Modest Recovery trajectory (66.7%), and no Low Stable trajectory was seen in the control sample. It would appear that the intervention facilitated stabilization in functioning or mild to marked recovery, and importantly may have prevented chronic low functioning seen in the control group. In other words, the FEP sample consisted of individuals who either had poor functioning that improved over time (most modestly) or good functioning that remained stable over time. In contrast, the control sample included individuals with on-going, chronic poor functioning. These findings should be deemed preliminary as sample sizes were quite small.

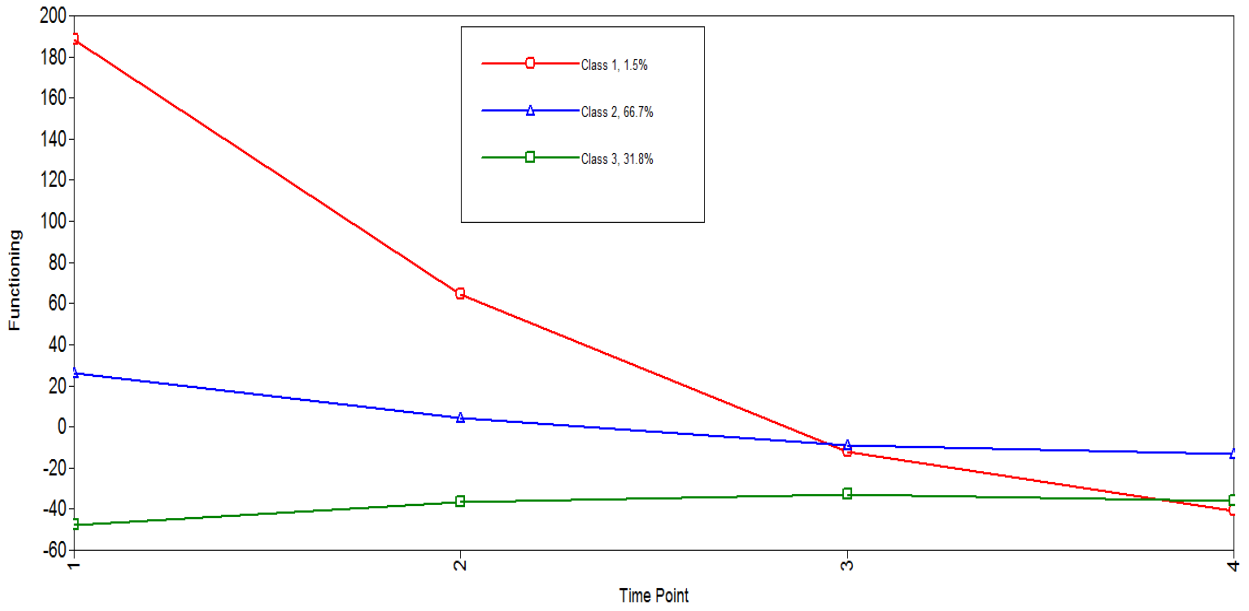

Figure 5. Latent class growth analysis of CANS Functioning factor over time in FEP participants. Note that higher scores indicate worse functioning. Mean is 0 and standard deviation is 100.

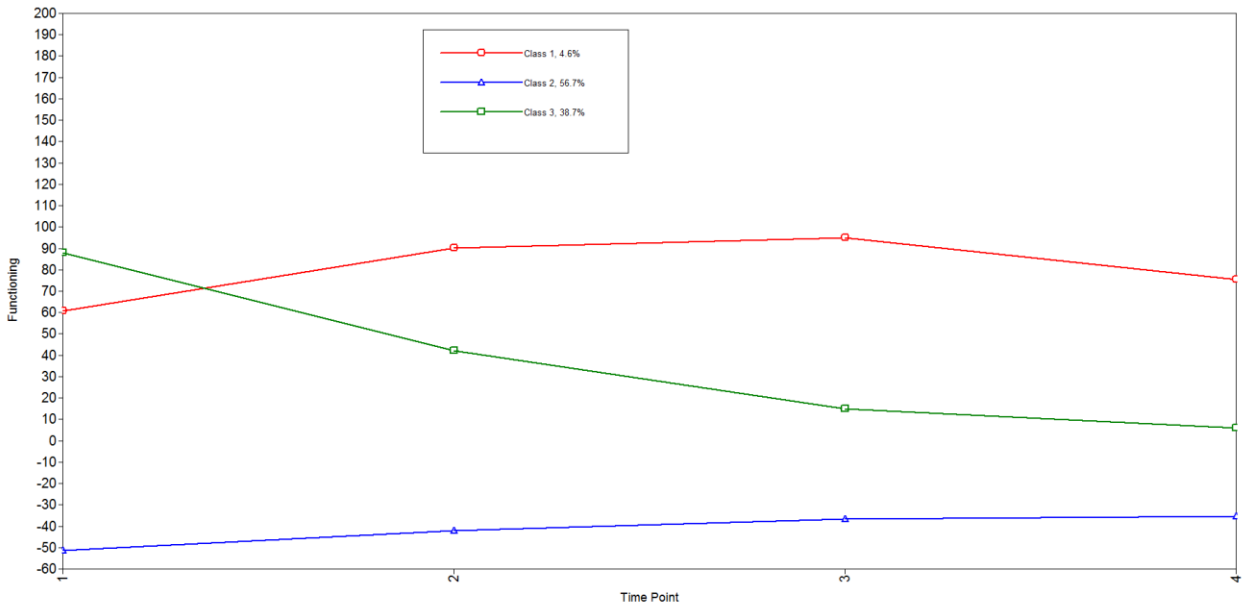

Figure 6. Latent class growth analysis of CANS Functioning factor over time in propensity-matched controls. Note that higher scores indicate worse functioning. Mean is 0 and standard deviation is 100.

Conclusions

Efforts of TIEMH investigators in fiscal year 2017 have largely comprised activities related to the initiation of a formal longitudinal-empirical study of the effectiveness of coordinated specialty care programs for the treatment of FEP in adolescents and young adults as implemented in state-funded mental health agencies across Texas. Toward this end, investigators have developed an evaluation protocol with input from stakeholders (i.e., providers and HHSC), received approval of the protocol from the University of Texas IRB, sought approval from four agency-specific IRB's, developed an online data collection process, and began fidelity assessments through several site visits. The protocol includes administration of self-report, family member-rated, and clinician-rated instruments covering a broad array of domains to supplement administrative data, including duration of untreated psychosis, clinical symptomatology, psychosocial functioning, subjective well-being, perceived family relationships, pertinent life events, school/employment participation, and discharge outcomes. The fidelity of visited programs to the OnTrackNY treatment model is clearly in the acceptable range for each (i.e., scores near 100). All fidelity assessments will need to be conducted prior to identification of relative strengths and weaknesses of specific programs, and evaluation of their potential relationships to psychosocial outcomes. Currently, seven of ten sites have agreed to participate in the program evaluation through collection of data at their agencies, and it is possible one or two others may participate at a later date. The proportion of individuals in FEP programs who agree to participate in the evaluation cannot be estimated as data collection has just begun at the visited sites.

The empirical evaluation of the effectiveness of the FEP program in young adults focused on psychosocial functioning as measured by the ANSA Life Domain Functioning items in comparison to propensity-matched controls. The results of longitudinal trajectory analyses were mixed. First, both FEP program participants and controls demonstrated Recovery, Worsening, and Stable trajectories. The FEP group showed lesser proportions of Recovery (i.e., 7.4% vs. 11.3%) and greater proportions of Worsening (12.9% vs. 4.8%) than the control group. However, those FEP participants in the Recovery class demonstrated better functioning than the Recovery controls across time points four through seven in particular. By time eight, the two groups were similar in functioning levels. It is important to recognize, first, that the reliability of the estimated trajectories is reduced as sample size decreases over time. That may partially account for the similar levels of functioning by time eight. Second, it is not entirely clear what overall factor scores on the Life Domain Functioning items measure. The highest loadings in the one-factor model indicate that the scores are best represented by Social, Recreational, Decision-making, Living Skills, and Self-Care items. It may be that FEP program participants and controls are better differentiated in terms of other specific types of functioning (e.g., academic functioning, employment). Differentiation may also be improved when additional measures of functioning (e.g. self-report, clinician) are available. The full evaluation protocol includes the WHO-DAS 2.0, which measures multiple domains of functioning, each with more than one item (i.e., cognition, mobility, self-care, getting along, life activities, and participation). Comparisons to the current findings with the ANSA will be made to further clarify these mixed results.

The empirical evaluation in adolescents focused on psychosocial functioning as measured by the CANS Life Domain Functioning items in comparison to propensity-matched controls. The sample size was much smaller (N = 67 per group) than that of the young adult sample, so the results are exploratory. Nevertheless, FEP participants showed both Recovery and Modest Recovery groups, and a High Stable group. Controls, however, showed a Low Stable trajectory not evident in the FEP group as well as Recovery and High Stable trajectories. This may be preliminary evidence that the FEP program during adolescence may yield improved recovery in functioning and perhaps prevention of chronically poor functioning. Here, the highest loading items on the CANS Functioning factor were Living Situation, Family, Judgment, Job, and Independent Living. These items provide a somewhat different picture of what the CANS scores represent relative to the ANSA scores that focused a bit more on social/recreational functioning and life skills. It may be postulated that FEP programs better target these former

types of psychosocial functioning (e.g., through Supported Employment and Education Specialists, Family Partners) than those measured by the ANSA. Again, these findings are preliminary, but the full evaluation will include assessment of psychosocial functioning using the same instrument in adolescents and young adults, thus eliminating the confound between age group and functioning measure.

Recommendations

Based on this 2018 evaluation, TIEMH makes the following recommendations:

1. Although seven out of ten sites with FEP programs in Texas have agreed to participate in the formal evaluation of their effectiveness, representatives from HHSC should consider encouragement of the others to take part as well.
2. Given that the most precise estimates of outcomes can be achieved with greater sample sizes, providers should encourage individuals in FEP programs to participate in the formal evaluation.
3. Based on the reported census differences between adolescents and young adults currently enrolled in FEP programs (i.e., $N = 509$ vs. $N = 67$), it may be worth exploring the reasons for this discrepancy to ensure that adolescents in need of such services are being adequately made aware of them and outreach efforts include a focus on adolescent-serving organizations.
4. The somewhat mixed findings with regard to psychosocial functioning in the larger young adult sample should not be deemed definitive in supporting nor undermining the effectiveness of the FEP program.
    - The evidence that young adult FEP participants in the Recovery trajectory showed greater improvements in functioning than controls in the Recovery trajectory is consistent with prior research using the GAF.
    - No prior studies of functioning changes in FEP program evaluation have utilized latent trajectory analyses in comparing treatment and control groups, thus the composition of Recovery and Worsening trajectories across groups is novel and requires replication.
5. The more positive findings with regard to psychosocial functioning in the adolescent sample likewise should not be deemed definitive in supporting the effectiveness of the FEP program.
    - Though suggestive of better prevention of chronic functioning difficulties and perhaps greater proportions of individuals in recovery trajectories in FEP program participants than controls, the sample size was diminutive particularly at later time points.
    - Again, no prior studies have used latent growth modeling and the findings await replication.
6. These preliminary findings, particularly in the small adolescent sample, should not be deemed evidence that FEP programs are more effective in youth than older adults.
7. The relatively higher ratings of problematic functioning on Social and Recreational items on the ANSA scale at the initial assessment, and their high factor loading on the overall Life Domain factor, along with the limited recovery seen in the young adult FEP sample, could suggest a need for additional interventions targeting social inclusion goals and engagement in community.
8. Clearly, more empirical research is needed to follow up on preliminary findings from last fiscal year's report and the current report, and to evaluate other psychosocial constructs using well-validated measures and larger sample sizes.

References

Aas, I. H. M. (2010). Global Assessment of Functioning (GAF): Properties and frontier of current knowledge. *Annals of General Psychiatry*, 9-20.

Addington, J., Heinssen, R. K., Robinson, D. G., Schooler, N. R., Marcy, P., et. al. (2015). Duration of untreated psychosis in community treatment settings in the United States. *Psychiatric Services, 66*(7), 753-756.

Alvarez-Jimenez, M., Parker, A. G., Hetrick, S. E., McGorry, P. D., & Gleeson, J. F. (2011). Preventing the second episode: A systematic review and meta-analysis of psychosocial and pharmacological trials in first-episode psychosis. *Schizophrenia Bulletin, 37*, 619-630.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders (4th ed)*. American Psychiatric Association: Washington, D.C.

Bertelsen, M., Jeppesen, P., Petersen, L., Thorup, A., Ohlenschlaeger, J, Quach, P., …Nordentoft, M. (2008). Five-year follow-up of a randomized multicenter trial of intensive early intervention vs. standard treatment for patients with a first episode of psychotic illness: The OPUS Trial. *Archives of General Psychiatry, 65*(7), 762-771.

Craig, T. K., Garety, P., Power, P., Rahaman, N., Colbert, S., Fornells-Ambrojo, M., & Dunn, G. (2004). The Lambeth Early Onset (LEO) Team: Randomized controlled trial of the effectiveness of specialized care for early psychosis. *BMJ, 329*(7474), 1067.

Harvey, P. O., Lepage, M., & Malla, A. (2007). Benefits of enriched intervention compared with standard care for patients with recent-onset psychosis: A meta-analytic approach. *Canadian Journal of Psychiatry, 52*(7), 464-472.

Kramer, M. D., & Lopez, M. A. (2017). *Initial evaluation of first episode psychosis early intervention programs in Texas*. Texas Institute for Excellence in Mental Health, School of Social Work, University of Texas at Austin.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*(4), 525-543.

Muthén, B., & Muthén, L. K. (1998-2012). *Mplus User's Guide*. Muthén & Muthén: Los Angeles.

Norman, R. M. G., Manchanda, R., Malla, A. K., Windell, D., Harricharan, R., & Northcott, S. (2011). Symptom and functional outcomes for a 5-year early intervention program for psychoses. *Schizophrenia Research, 129*, 111-115.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., …Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet, 382*, 1575-1586.