Implementing and Analyzing the Advantages of Voice AI as Measurement-Based Care to Address Behavioral Health Treatment Disparities Among Youth in Economically Disadvantaged Communities

Yared Alemu, PhD[1] ⓘ, Elizabeth Cárdenas Bautista, PhD[2] ⓘ, Sarah Vinson, MD[3] ⓘ, Patrick Ohiomoba, MS[4] ⓘ, Nakia Melecio, PhD[5] ⓘ, Hermon Amare, MD[6] ⓘ, Jaya Vivia[7] ⓘ and Harshitha Karippara[8] ⓘ

[1]TQIntelligence, Inc., Atlanta, Georgia, USA; [2]Morehouse Healthcare, Atlanta, Georgia, USA; [3]Department of Psychiatry & Behavioral Sciences, Child and Adolescent Psychiatry Fellowship Program, Morehouse School of Medicine, Atlanta, Georgia, USA; [4]Marvin, Harvard John A. Paulson School of Engineering and Applied Sciences, Cambridge, Massachusetts, USA; [5]Mental Health Advocate, Addis Ababa, Ethiopia; [6]Quarry Lane School, San Ramon, California, USA

Corresponding Author: Catrell Harris, Email: charris@reingold.com

Abstract

Objectives: This study evaluates one aspect of the TQIntelligence (TQI) Measurement-Based Care (MBC). Once fully developed and validated, the platform is intended to quantify the level of emotional distress using voice artificial intelligence and track treatment outcomes in children and adolescents from low-income communities. The focus here is on non-voice data collected and evaluated as part of developing a voice-based machine learning model. In addition, this article explores the benefits and challenges of implementing such MBC systems in community behavioral health settings that serve low-income communities.

Methods: Researchers collected data from a large community sample of children and adolescent patients receiving mental health and family preservation services. The TQI worked with three community behavioral health agencies in the Southeastern United States. While patients were in treatment, researchers collected data from electronic medical records, including diagnosis and treatment history. Multiple therapists in this pilot study collected psychiatric screening data using the Symptoms and Functioning Severity Scale, Patient Health Questionnaire, and The Adverse Childhood Experiences survey to screen for depressive disorders and trauma exposure. The therapist also collected a 15–45-second voice sample from pediatric patients at the beginning of the session.

Results: The results show that TQI's MBC, like other MBC systems, demonstrated significant benefits, including detecting improvement in psychiatric symptom severity for children and adolescents who have received consistent services. The data indicate that providing consistent data-driven services contributes to care, which results in a substantial decrease in psychiatric symptoms, particularly between sessions 15 and 20, and remission of psychiatric symptoms with no indication of relapse between sessions 20 and 40 in a significant proportion of patients.

Conclusions: The application of long-term treatment and TQI's data-driven system shows benefits, such as the reduction and remission of psychiatric symptoms in children and adolescent populations. This study contributes to the literature as it clearly demonstrates support for investment in and the application of digital mental health systems to detect and quantify emotional distress. It demonstrates that these systems effectively inform a larger strategy to address a long history of inequity in treatment outcomes for youth from marginalized communities.

*The Appendix after the References defines the acronyms used in the article.

Only 20% of behavioral health providers use measurement-based care (MBC).[1] Two reasons for MBC's low-uptake outcomes include a need for stronger consensus regarding optimal use (in frequency and consistency) and the absence of a widely utilized data analytics infrastructure. TQIntelligence (TQI) has built and implemented a measurement-based system for community behavioral health providers, which includes the use of a novel artificial intelligence (AI)-enabled voice algorithm designed to provide psychiatric decision and triaging support to pediatric populations. The success of the implementation and related outcomes varies depending on the organization and the therapist's involvement in the pilot.[2]

This article contributes to measurement care and its effectiveness in the literature. In addition, it challenges the dominant narrative that such systems are too complicated and ineffective in community behavioral health that serves children and adolescents from low-income communities. In fact, MBC is effective in enhancing treatment decisions, improvement in patient engagement in care, detecting changes in symptoms, and reducing symptom severity.[3] Despite its effectiveness, most community behavioral health agencies require greater capacity to implement MBC. In these agencies, there are often training gaps in MBC, time constraints of sessions, and sizeable caseload-to-provider ratios. Most payers and providers need more service-quality incentives that translate to provider investment in the utilization of MBC as it is commonly implemented currently.

Background

Before the onset of the COVID-19 pandemic, children and adolescents in marginalized communities experienced a mental health crisis, which was partly due to a lack of quality mental health services. This becomes apparent when measured in the rate of suicide for Black youth, which increased by 80% from 2000 to 2020.[4]

By 2025, the healthcare system will be lacking 225,000 therapists—further exacerbating the current crisis.[2] Therefore, it is critical to have upstream interventions and highly efficient use of the available mental healthcare resources. Clinicians who care for the most vulnerable youth might benefit from the integration of innovative digital mental health interventions to identify problems.

While the COVID-19 pandemic was, predictably, a tremendous stressor that adversely impacted the population's mental health, an increase in the rates of youth mental health problems predates the pandemic. According to the Substance Abuse and Mental Health Services Administration (SAMHSA) and US Department of Health and Human Services (HHS) Centers for Disease Control and Prevention, between 2016 and 2020, the number of children aged 3–17 years diagnosed with anxiety grew by 29%, and the percentage with depression by 27%.[5] In 2020, suicide was the second leading cause of death for young people aged 10–14 years and 25–34 years, and among the top nine leading causes of death for people aged 10–64 years.[6] Additionally, there has been an increase in suicide rates among children and adolescents,[7] especially in low-income communities.[8]

A meta-analysis of 29 studies with more than 80,000 total participants reported that an increase in emergency department visits in 2021 alone was 22.3%, with one out of every four adolescents experiencing clinically relevant depressive and anxiety symptoms—double the rate of pre-pandemic levels.[9]

Our rationale for focusing on young people is partly related to two factors: the age of onset of mental disorders, which typically start during early developmental stages and manifest in almost half of all individuals before they reach 14 years of age, and the role trauma plays in more than three-quarters of youth suffering from mental health disorders.[10]

To address the increased rate of mental health needs along with the projected shortage of mental health providers, novel, scalable screening, diagnostic tools, and methods are urgently needed. These novel tools must be developed to embrace the emerging promises of AI in the context of existing gold standards, including diagnostic interviews and empirically validated surveys. In combination, innovative digital health and scientifically validated psychiatric surveys and treatment plans are likely to address the gaps in mental health disparities that directly impact the most marginalized populations, including those from low-income communities.

The MBC is defined as the systematic administration of symptom rating scales and the use of the results to drive clinical decision-making at the level of the individual patient.[11] It provides "enhanced precision and consistency in disease assessment, tracking, and treatment to achieve optimal outcomes."[12] Symptom rating scales are structured instruments patients use to report their perceptions of psychiatric symptoms.

Potentially, MBC can enhance the therapeutic relationship between the patient and the provider. It can lead to a more informed and engaged patient who participates meaningfully in treatment and shares in decision-making and the development of treatment goals. An early meta-analysis of six studies with nearly 300 therapists and more than 6,000 patients reported that those randomly assigned to MBC had significantly and substantially better outcomes than patients randomly assigned to usual care.[13] Since the increased use of telehealth during the COVID-19 pandemic, the integration of MBC and telehealth is considered an innovative approach to improve mental

health outcomes by enhancing access to care, quality, and collaboration with patients and providers.[14] Novel studies on the implementation of telehealth and MBC indicate that there is enhanced communication with patients and providers and improvement of traditional treatment and depressive symptoms.[15]

By observing the clinical effectiveness of their treatments systematically, providers might find it easier to hone their clinical skills over time. The lack of structure for such observations might contribute to the poor outcomes often observed in routine care. Furthermore, the routine use of symptom rating scales is a tool for clinical practices to quickly evaluate the effectiveness of their quality improvement initiatives and demonstrate to payers that their treatments are effective. The application of MBC supports the recognition of stagnation in treatment response. As reiterated by researchers Fortney and colleagues,[13] "based on clinical judgment alone, mental health providers detect deterioration for only 21.4% of their patients who experience increased symptom severity." The lack of timely identification of patients whose treatment is not indicating improvement is what researchers Henke and colleagues[16] called "clinical inertia, the act of not changing the patient's treatment plan even with a lack of substantial improvement in symptom severity."

While MBC is promising, there are some caveats to its use and effectiveness. Mental health symptom rating scales or measures should be psychometrically valid, culturally receptive, and used for the normed population. The use of measures that are unreliable, sensitive to change, or have poor concurrent validity could misinform clinical decision-making. To be effective, MBC programs should be implemented into existing clinical workflows and collect symptom severity data from patients frequently and shortly before or during the clinical encounter. The low utility of MBC includes administrative burden, lack of clinical implementation knowledge, and concern of data-biased clinicians.[17] Specifically, clinical settings may also play a role in the low utility of MBCs. Community-based clinicians often face the challenges of maintaining a heavy caseload while being constrained in time. The most commonly reported reasons for not implementing MBC paperwork are the required time and lack of personnel resources.

The ongoing mental health crisis continues to be addressed by mental health providers by utilizing diagnostic guidelines and tools, including MBC, and the implementation of AI and machine learning (ML) in digital mental health tools. Researchers Koutsouleris and colleagues[18] discuss the promise and challenges of augmenting MBC and other clinical tools with the integration of AI in psychiatric research and clinical practice to address gaps in mental health treatment.

Administering brief measures can streamline treatment by creating a quick method for assessing symptoms and treatment progress, or lack thereof. The exploration of real-world AI implementation within MBC can also provide promise to address key issues in mental health delivery. The implementation of AI and ML in healthcare tools is still in nascent development within mental health. However, the progression of AI shows great promise in mental health research and clinical practice.

Researchers Graham and colleagues[19] provided an overview of 28 AI and mental health studies that showed early evidence of accuracies of AI's potential in mental healthcare and emphasized the need for continuous development as the applicability of mental health tools improves. These studies reviewed using electronic health records, mood rating scales, and the implantation of novel monitoring systems and platforms such as smartphones and social media to identify mental health concerns. The potential for AI in mental healthcare is vast, as the implementation of AI as a supportive tool in early diagnosis and individualized interventions continues to be refined and improved.

Researchers Tutun and colleagues[20] developed innovative AI-based tools to detect and diagnose mental health disorders efficiently. These researchers created innovative tools to improve clinical diagnostic decisions by replacing time and cost-consuming paper-based mental health measures for an AI intervention. Our study is in line with the ongoing efforts to incorporate AI in mental healthcare, from research to clinical application, in a culturally responsive and ethical approach. Advances in AI and voice recognition technology could contribute to a transformative opportunity to develop voice biomarker algorithms to support healthcare providers in assessing mental disorder severity and tracking treatment outcomes using objective measurements; objective measures have significant clinical and practical benefits. The potential for AI in mental healthcare is impressive and will require continuous development and incorporation of feedback from providers and community stakeholders.

The TQI, recipient of the National Science Foundation Small Business Innovation Research (SBIR) Phase I, Phase II, and Supplemental funding, built and continually improved ClarityConnect, TQI's Voice AI product suite. ClarityConnect is driven by a pediatric voice psychiatric distress algorithm (patent serial No.: 17/550544) that, through a voice sample from a patient in a mental health setting, can assess the severity of a mental health problem within minutes and advise providers on how to route individuals to the appropriate resources and care. The product is designed to provide psychiatric decision support and triaging

solutions focused on the unique needs of children and adolescents between the ages of 5 and 18 years. The solution is intended to serve as an augmented clinical intelligence to support clinicians.

The scientific foundation for our pediatric voice psychiatric distress algorithm is based on training the ML model in emotional recognition and clinically severe distress states. The algorithm is based on the scientific relationship between trauma, stress, and the human voice. Our Voice AI technology is developed with the complexity of the human voice that involves the coordination of 100 muscles in the chest, neck, and throat. The platform uses speech's prosodic and acoustic features to predict each person's level of emotional distress (negative emotions).

Similar to how a thermometer measures temperature, TQI uses voice/speech to measure the severity of emotional distress. Our approach to Speech Emotion Recognition classification is unique. Instead of focusing on the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnosis, the focus is on a broader clinical issue of emotional distress or negative emotion, which is present across all mental health diagnoses. This model has helped track treatment outcomes related to the reduction of distress and progression and its impact on the person's level of functioning, as described in our preliminary data.[2] To this end, we have focused on detecting and classifying primary or archetypal emotions of anger, fear, sadness, neutral, and happiness. In detecting, quantifying, and tracking emotional distress using these archetypal, TQI's ML model recognizes and connects trauma-based psychological acuity to vocal emotion recognition. Along with TQI's voice AI technology (Figures 1 and 2), TQI's MBC system includes the use of validated clinical instruments to support the assessment of patient symptoms. The clinical instruments selected for integration into TQI's MBC system consider the practicality and adaptability in diverse clinical settings.

The prevalence of early-onset trauma-based psychiatric disorders for Black and Hispanic youth is related to the high rates of traumatic exposures, which are intergenerational and partly due to a history of discrimination, systemic racism, structural inequity, and poverty that shapes the social determinants of health, and poor access to quality mental health care.[21] In this pilot, we used the original 10-question Adverse Childhood Experiences (ACEs) questionnaire to capture some forms of traumatic exposure. The ACEs have not been normed specifically on marginalized groups. However, a literature review on the ACEs questionnaire has assessed the appropriateness of practical and ethical implementation of the questionnaire in a variety of settings with pediatric and adult populations.[21] In addition, we used the Symptoms and Functioning Severity Scale (SFSS) and the Patient Health Questionnaire (PHQ-9) to measure certain behavioral and emotional consequences of trauma. Our voice algorithm is also a tool to measure the severity of trauma-based emotional and behavioral disturbances.[2]

The ACEs encompass adversity that children face in their environment, including traumatic events, neglect, maltreatment, physical and emotional abuse, and dysfunctional family experiences that occur before they turn 18 years old. The ACEs are associated with a greater risk of developing anxiety, depression, disruptive behavior disorders, and substance use disorders, which could have a negative impact later in their life.[22] The ACEs are common.[23] However, families from lower socio-economic status bear a more significant burden of exposure and are the primary target of our research.

The ACEs tend to co-occur and accumulate and, in turn, increase the likelihood of internalizing symptoms (e.g. depressed mood, anxious thoughts, and inattention) and externalizing behaviors (e.g. aggression, self-harm behavior, and posttraumatic hyperarousal). Additionally, they can adversely impact youth neurobiologically, developmentally, and socially. Consequences across the lifespan may include disrupted neurodevelopment, socio-emotional dysfunction, and cognitive impairment. In addition, ACEs are associated with an increased risk of future substance use, poor physical health, and early
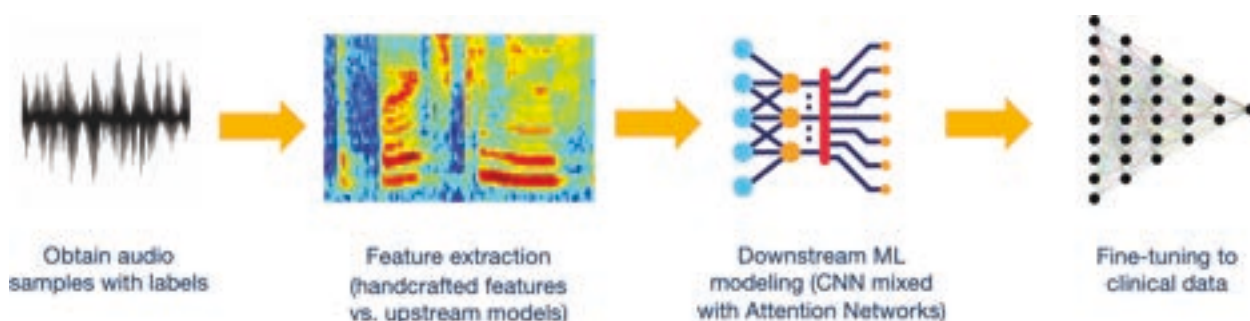

Obtain audio samples with labels

Feature extraction (handcrafted features vs. upstream models)

Downstream ML modeling (CNN mixed with Attention Networks)

Fine-tuning to clinical data

Fig. 1. TQI's (TQIntelligence) Voice AI (artificial intelligence) analytics platform.
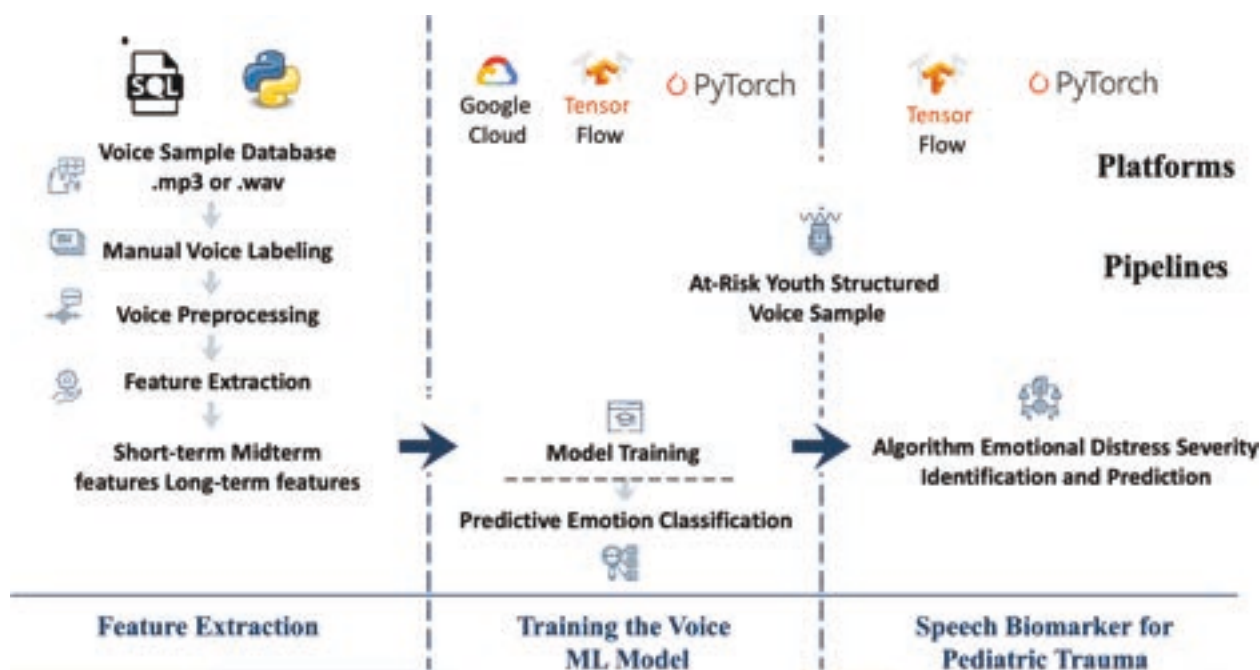
Fig. 2. TQI's (TQIntelligence) AI (artificial intelligence) technology flow chart.

chronic disease, disability, and mortality.[24] The ACEs scores tend to be higher for families from lower socioeconomic status.[25]

The SFSS is a scientifically validated survey normed for the mental health treatment of youth aged 11–18 years. The SFSS is a valuable component of TQI's patient-centered outcomes management platform because it includes patients and families in the data collection process. The SFSS has three versions: one for therapists, one for caregivers, and one for patients over 11 years old. Comparative rating scale scores attained at the point of care help to address therapist-family member gaps in the perception of symptoms and treatment response. Respondents rate the frequency of symptoms and behaviors over the preceding 2–3 weeks. The SFSS is trauma-informed and has been validated with clinical samples like the one in this project and studies with the SFSS. Clients of therapists who received feedback on the SFSS improved faster than those without such feedback. Effects were more substantial when clinicians had multiple sources, i.e. from the patient, caregiver, and clinician.[26]

Methods

Participants

Investigational review board guidelines regarding consent, data perimeters, risks, and privacy guided this study. The TQI established comprehensive policies to ensure robust data security and address potential privacy risks through a thorough encryption process. The data collected by TQI, including voice samples and other protected health information (PHI), were stored in The Health Insurance Portability and Accountability Act of 1996 (HIPPA)-compliant cloud infrastructure. The TQI's policies included the collection and storage of data in accordance with the data privacy regulations from the community health agencies that participated in the study. Additionally, the TQI data system also employs encryption both in transit and at rest, meaning data are encrypted as it is transferred between systems and when it is stored on data servers. This dual approach to data encryption ensures the data collected for this study remained secured through all stages of data handling. Participants in the study were also provided with assigned identification numbers to ensure their privacy; all this information was provided to participants in a detailed informed consent, which was required before participation in the study.

Participants were recruited from three community behavioral health agencies in the Southeastern United States, where participants received outpatient therapeutic services. The therapist at these sites included master-level unlicensed therapists supervised by licensed mental health providers with clinical training in the areas of anxiety, depression, and trauma-informed psychotherapy. These settings primarily focused on children and adolescent populations. Inclusion criteria for participation in this study included 5–18-year-olds with no long-standing history of suicidality or violence, psychotic disorder, or unmanaged bipolar disorder.

Given the vulnerability of the pediatric population and potential distrust of the utilization of novel technologies, the research team ensured thorough training for

counselors who used the TQI's MBC system and sufficient time with potential participants and caregivers to provide detailed information about our study and our AI technology. Additionally, potential participants and caregivers were provided with detailed information in written consent forms, which they had the opportunity to review and ask for clarification if needed. Participants were also reminded their participation was voluntary, and that their decision to participate or not would have no impact on their access to mental health services. The decision to participate in this study was also supported by the clinical judgment of the onsite therapist and their clinical supervisor. The recruited participants were already in the therapist's caseload and eligible to participate in this study based on meeting the study's parameters. Patients disclosed their ability and willingness to participate, with both parent/caregiver consent and patient assent. Clinical directors in prospective agencies serve as the clinical oversight of all patients-related matters and participate in established clinical meetings with TQI clinical leadership staff for support with the implementation of TQI's MBC data-driven system.

Before enrolling in this study, the patients had a detailed psychosocial assessment by another clinician, including a specific DSM diagnosis and treatment plan. The therapist used this intake assessment to consider each patient's appropriateness for the study. The therapist also consulted their clinical supervisor for additional support. The researchers outlined the exclusion and inclusion criteria to guide appropriate decisions regarding extending the opportunity to participate.

In adherence to best practices, all prospective participants were informed and reminded of the voluntary nature of their participation in the study through written and verbal information. The parents of existing patients could decline to be part of this study. There were no changes to the community standard treatment process if the parent/the patient declined participation. The therapist could also consult with their supervisor and the study's principal investigator, Dr Yared Alemu, regarding patients' eligibility for this study.

The initial stages of rapport building between the therapist and the patient/family were integral to the engagement and recruitment strategy. Before discussing the study, the therapist invested time and energy in the first four to five sessions, building rapport with the patient and family. Participants and parents were provided the opportunity to discuss with clinical staff any concerns associated with participating in the study including privacy and confidentiality implications with the use of novel AI technology. The contract signed by the therapist for compensation outlined the specific parameters, including SFSS or the PHQ-9 and the voice sample. Therapists were compensated $10 per data collected based on the previously mentioned parameters.

After the initial training, which included data safety protocols, ThinkQuality installation, data collection, and real-time feedback on psychiatric severity and outcome measures, the principal investigator and the study site directors met with the therapist every other week to monitor the progress and troubleshoot issues the therapist encountered. In this study, the therapist's main objective was to optimize the patient's treatment rather than enroll as many participants as possible. All therapists in the study were under clinical supervision by the agency-designated supervisors for licensure. The supervisor addressed the therapist's caseload, attrition patterns, and patient progress. This process was independent of this study and provided yet another checkpoint to address any potential concerns with patient/family coercion related to study participation.

Patients were already receiving behavioral health services from the three community behavioral agency sites in the Southeastern United States before being asked to participate in this project. Patients were provided with comprehensive informed consent for treatment and study participation. All appropriate treatment consent forms were signed during the intake and study enrollment process. The intake counselor was tasked with discussing the study with the participants. An informational flyer was provided to the family digitally and manually as part of the participant recruitment process. They were also offered a recruitment brochure, a copy of which can be found in the appendix.

Data Collection

The TQI's MBC data collection workflow and TQI's MBC survey frequencies are illustrated in Figure 3 and Table 1, respectively.

The data collection process included concurrently administering scientifically validated screening surveys ACE, PHQ-9, and SFSS.[27] Trained therapists collected the data ($n = 15$). Voice samples and SFSS, PHQ-9, and ACEs scale responses were collected using the TQI app ThinkQuality, which was available in IOS and Android. Each therapist received a 3–5-hour in-person training about ThinkQuality installation, data collection, and real-time feedback on psychiatric severity and outcome measures.

Data were collected from patients every week at the beginning of their treatment session as part of their mental health or other family-based intervention visits. Visits occurred in a suitable (private) location at home (including foster and group homes) or school as part of a state's commitment to access services. Data were collected virtually during the service shift secondary to the COVID-19 pandemic.
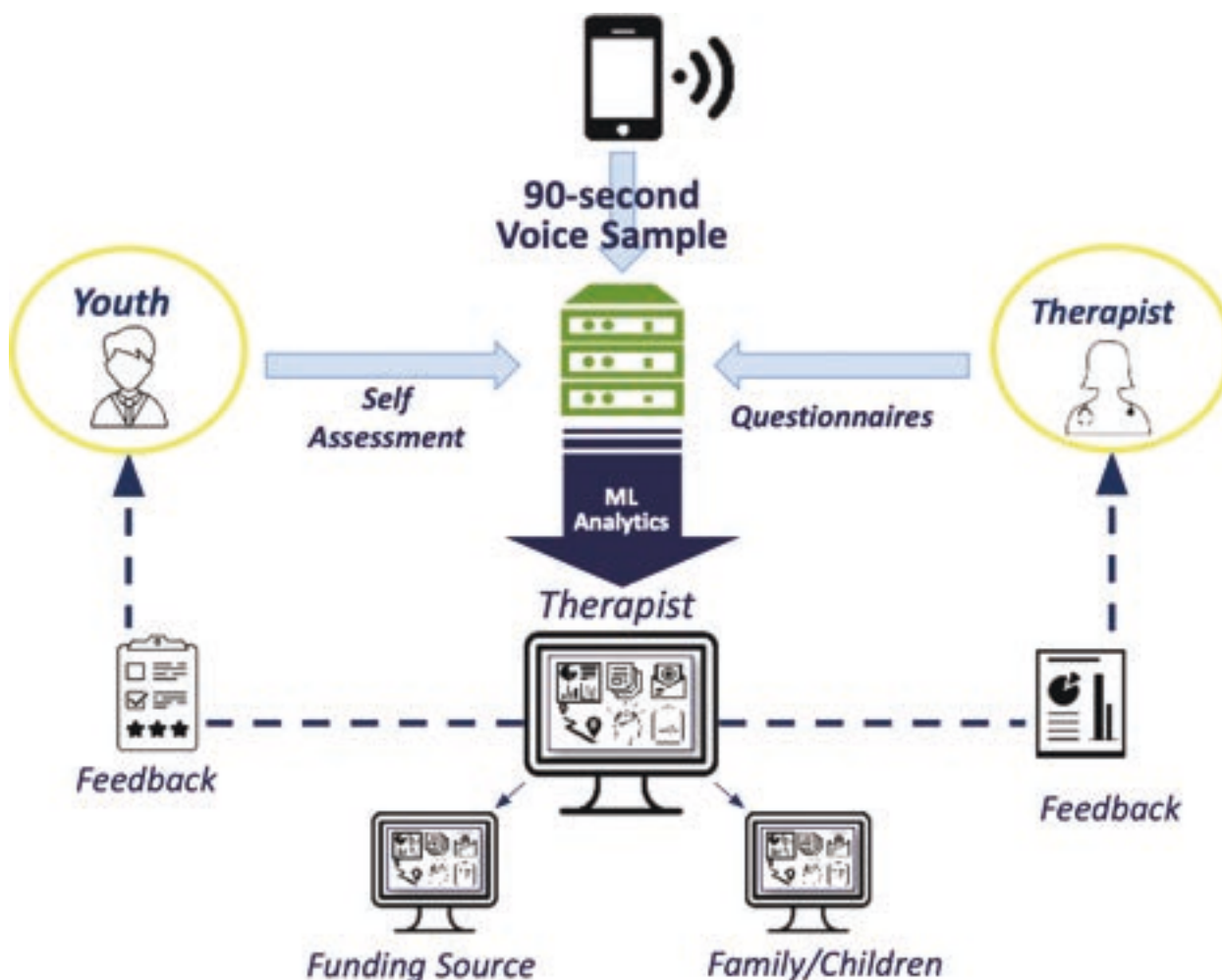
Fig. 3. TQI's MBC Data collection workflow. MBC: Measurement-Based Care; TQI: TQIntelligence.

Table 1. TQI's MBC survey frequencies.

Survey	Frequency	APP/Portal	Risk scores
PHQ-9	Required the first time logging in collect until the score is 8 or below	Phone APP only	16 or higher
SFSS: Clinician	1 time weekly	Phone APP only	65 or higher
SFSS: Client (11 years or older)	1 time weekly	Phone APP only	65 or higher
SFSS: Cargiver	1 time monthly	Phone APP only	65 or higher
Structured voice sample (reading sample provided)	1 time weekly	Phone APP only	N/A
Unstructured voice sample	1 time weekly	Phone APP only	N/A
ACEs	1 time and can be updated	Web portal only	4 or higher

ACEs: Adverse Childhood Experiences; APP: application; MBC: Measurement-Based Care; SFSS: Symptoms and Functioning Severity Scale; TQI: TQIntelligence; PHQ: Patient Health Questionnaire.

The app was used to score the surveys immediately and made the results available to the therapist to share with their patients if clinically appropriate; real-time data availability was intended to close the gap in transparency, accountability, and family engagement. The data collected by TQI, validated screening surveys, voice samples, PHI-related information, and consent forms were digitally stored on a HIPPA-compliant cloud-based technology on the Google platform. Additionally, all three community behavioral sites held accredited agencies with strict requirements of collecting and storing all patient-related information
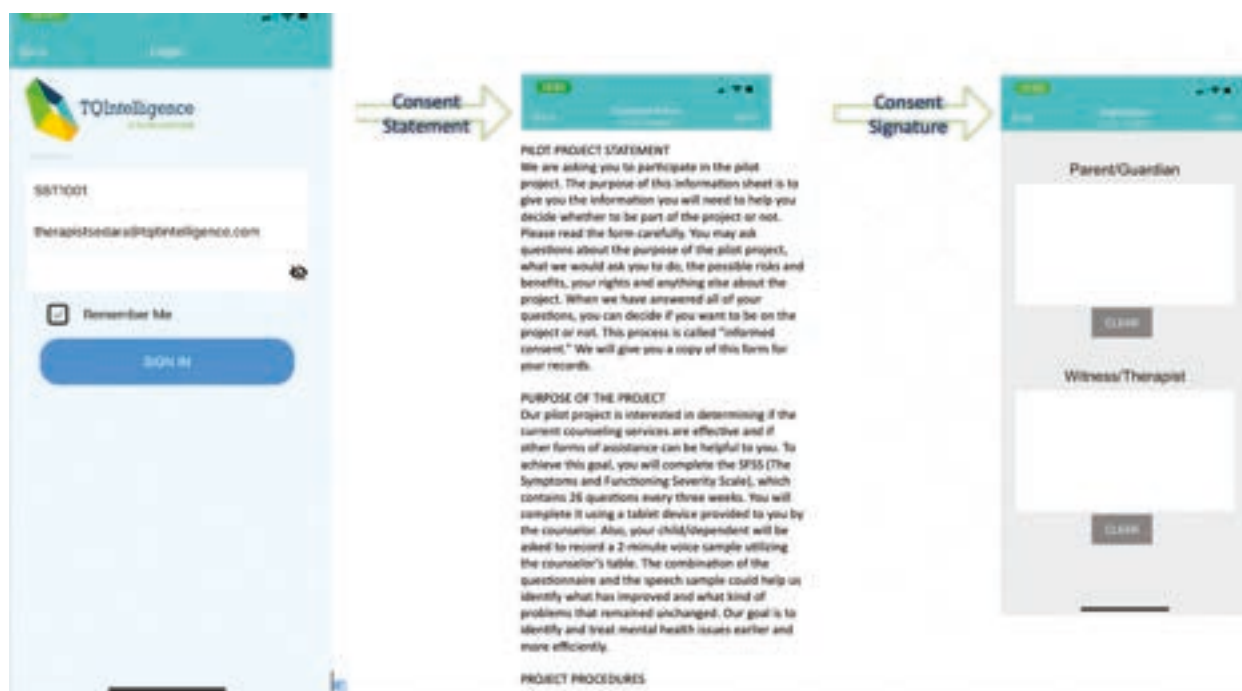
Fig. 4. TQI's App login page (left) and TQI's Consent (right). TQI: TQIntelligence.

Patient Participation

Once patients are logged into their account, TQI's application directs participants to a detailed consent form page, which requires the signature of the caregiver to continue to the TQI MBC system activities (Figure 4).

The activities for patients enrolled in the study included data collected using multiple surveys and a voice sample. The surveys included the SFSS, the PHQ-9 (Figure 5), and ACEs. The SFSS has three versions: SFSS-Clinician (therapist), SFSS-Youth for patients 11 years old and older, and the SFSS-Caregiver; the SFSS version can be found in the appendix. Participants started with the PHQ-9 and continued until their score was below 8. Participants then continued with the SFSS and ACE.

Once completed with the surveys, participants were directed to provide a voice sample (Figure 6), which was completed with the assistance of the assigned therapist. Masters-level unlicensed therapists collected voice samples on readings pre-selected by the research team or unstructured prompts. For the pre-selected recordings, patients were allowed to choose one of the four-sentence groups and read one of the two brief paragraphs from the Alice in Wonderland children's book (See Appendix 2). Previous work regarding the voice ML model, published recently, indicated 80% accuracy in detecting psychiatry severity.[28]

During the voice recording process, participants were informed they could pause the recording if needed, as the TQI MBC system allowed for a brief pause. The TQI technology also has a playback feature where participants can hear their previous recording. In addition, participants were informed they could decide to cancel and/or start over the recording process. If participants were satisfied with their recording, they were instructed to close and submit their recording.

Finally, when patients and clinicians submit all data, the results of surveys are available to participants and clinicians via tables or graphs on the TQI's application (Figure 7).

The intention of the redundant measurement of psychiatric symptoms, for example, the PHQ-9, ACEs survey, and SFSS, was to corroborate the reliability of scores, including the inconsistency of scores due to the therapist's perception of psychiatric severity. The ACEs questionnaire was selected because studies have shown an ACEs score of 4 or more has been associated with an increased risk of trauma symptoms, including aggression, self-harm behavior, posttraumatic arousal, and inattention, as well as neurobiological, developmental, and social delays.[29] ACEs also exacerbate the chances of future drug abuse; poor physical health; and early chronic disease, disability, and mortality.[30] The 9-item PHQ-9 is a validated self-report measure that identifies depressive symptoms.[31] PHQ-9 scores of 0–4, 5–9, 10–14, 15–19, and 20–27 indicate minimal, mild, moderate, severe, and very severe depression symptom severity, respectively.[32]

Data were collected and summarized systematically from screening surveys ACEs, PHQ-9, and SFSS. For each screening survey, participants' specific data (age, gender, diagnosis, Current Procedural Terminology [CPT] codes for treatment types, voice samples, and survey scores)
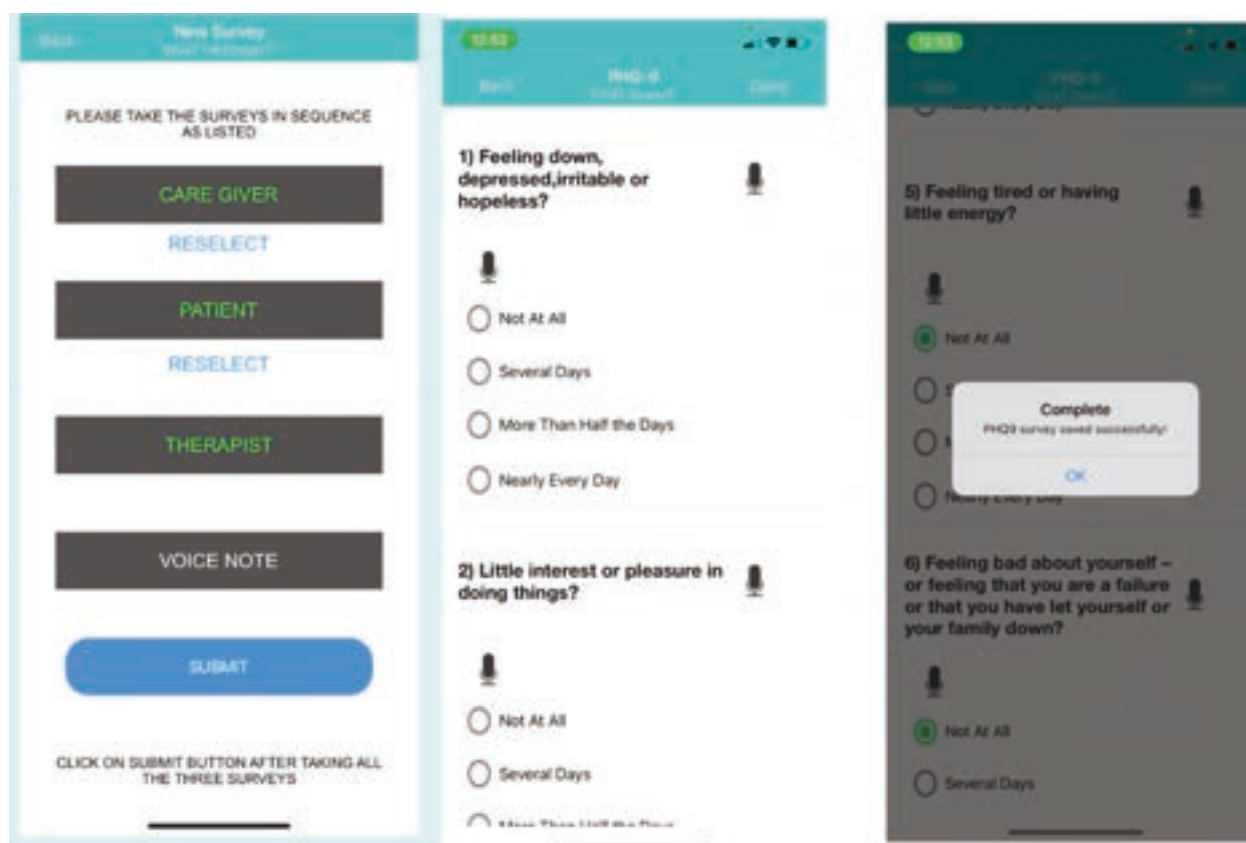
Fig. 5. TQI's MBC system, PHQ-9 (left), and SFSS (right). MBC: Measurement-Based Care; SFSS: Symptoms and Functioning Severity Scale; TQI: TQIntelligence; PHQ: Patient Health Questionnaire.

were collected and organized categorically. First, we used descriptive statistics, such as mean, median, and standard deviation, for each survey to understand central tendencies and variability. To analyze the data, we identified survey trends and considered how they relate to our objective to evaluate the TQI MBC system to quantify and track treatment outcomes.

To compute the best-fit lines of the data, we individually input images of the graphs into a Python graph data analyzer. Python is a widely used programming language for software development for scientific applications, statistics, and web applications.[33] The Python program identified data points on each graph, which expressed the relationship between the data and the line of best fit. Along with outputting the graph with the line of best fit, the Python data analyzer also outputs r, r^2 and n values. In a statistical analysis, r represents the correlation coefficient, indicating the strength and direction of a linear relationship between two variables r^2 or the coefficient of determination. This quantifies how well the independent variable(s) explains the variation in the dependent variable. It ranges from 0 to 1, where higher values suggest a better fit of the regression model. Finally, "n" denotes the sample size, representing
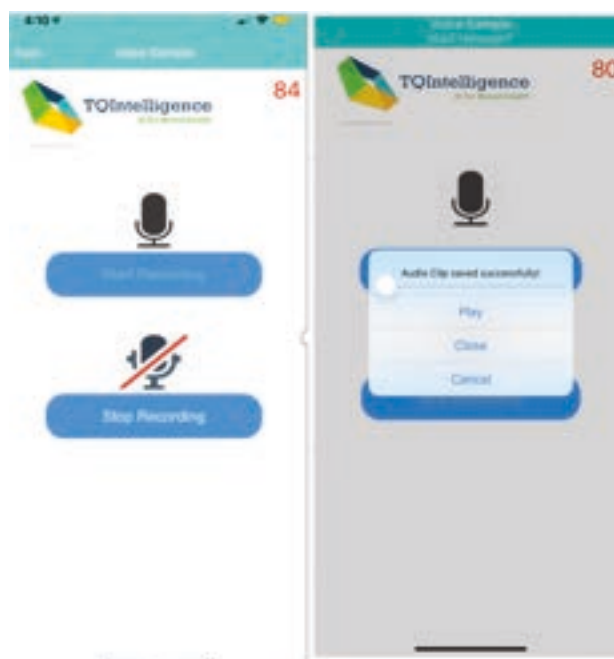

Fig. 6. TQI's MBC system voice sample page. MBC: Measurement-Based Care; TQI: TQIntelligence.

Fig. 7. TQI's MBC system, survey results. MBC: Measurement-Based Care; TQI: TQIntelligence.

the number of data points or observations in a dataset. Larger "*n*" values generally result in more robust and reliable statistical conclusions. These measures are crucial for assessing regression models' goodness of fit and reliability, providing valuable insights into the relationships and data quality. Figure 8 shows a detailed workflow figure.

Results

Our analysis reveals that youth with elevated SFSS Internalizing scores indicate negative emotional experiences that are inwardly focused, including depression, anxiety, and social withdrawal. These internalizing symptoms often manifest as a significant emotional burden that impedes daily functioning and overall well-being. The results show a strong negative correlation between SFSS symptom severity and the number of counseling sessions. Specifically, we found that as participants engaged in more counseling sessions, their internalizing scores significantly decreased. This relationship was quantified with a correlation coefficient (*r*) of –0.86, indicating a very strong inverse relationship, meaning that as the number of counseling sessions increases, the severity of internalizing symptoms decreases substantially. The negative correlation indicates an inverse relationship, where higher levels of one variable are associated with lower levels of the other. In practical terms, this means that the more counseling sessions the youth

attended, the more their symptoms of depression, anxiety, and social withdrawal were alleviated.

Additionally, the coefficient of determination (*r*²) is 0.74. This value indicates that approximately 74% of the variance in internalizing scores can be explained by the number of counseling sessions. The *r*² value, which ranges from 0 to 1, provides insight into the proportion of the variability in the dependent variable (internalizing scores) that is predictable from the independent variable (number of counseling sessions). An *r*² of 0.74 is considered very high, suggesting a substantial impact of the counseling sessions on the reduction of internalizing symptoms. This high *r*² value underscores the practical significance of the relationship, illustrating the effectiveness of consistent counseling sessions in reducing internalizing symptoms.

The primary aim of this pilot study was to evaluate the TQI MBC system to quantify the level of emotional distress and track treatment outcomes in children and adolescents from low-income communities. Findings show that when services are delivered consistently and outcome data are collected systematically, there are significant reductions in psychiatric severity. This supports the use of TQI's MBC system, as it is beneficial in contributing to a treatment plan that results in the remission of psychiatric symptoms in youth from low-income communities.

For the analysis, we used a sample size of 80 participants, and the correlation was statistically significant at
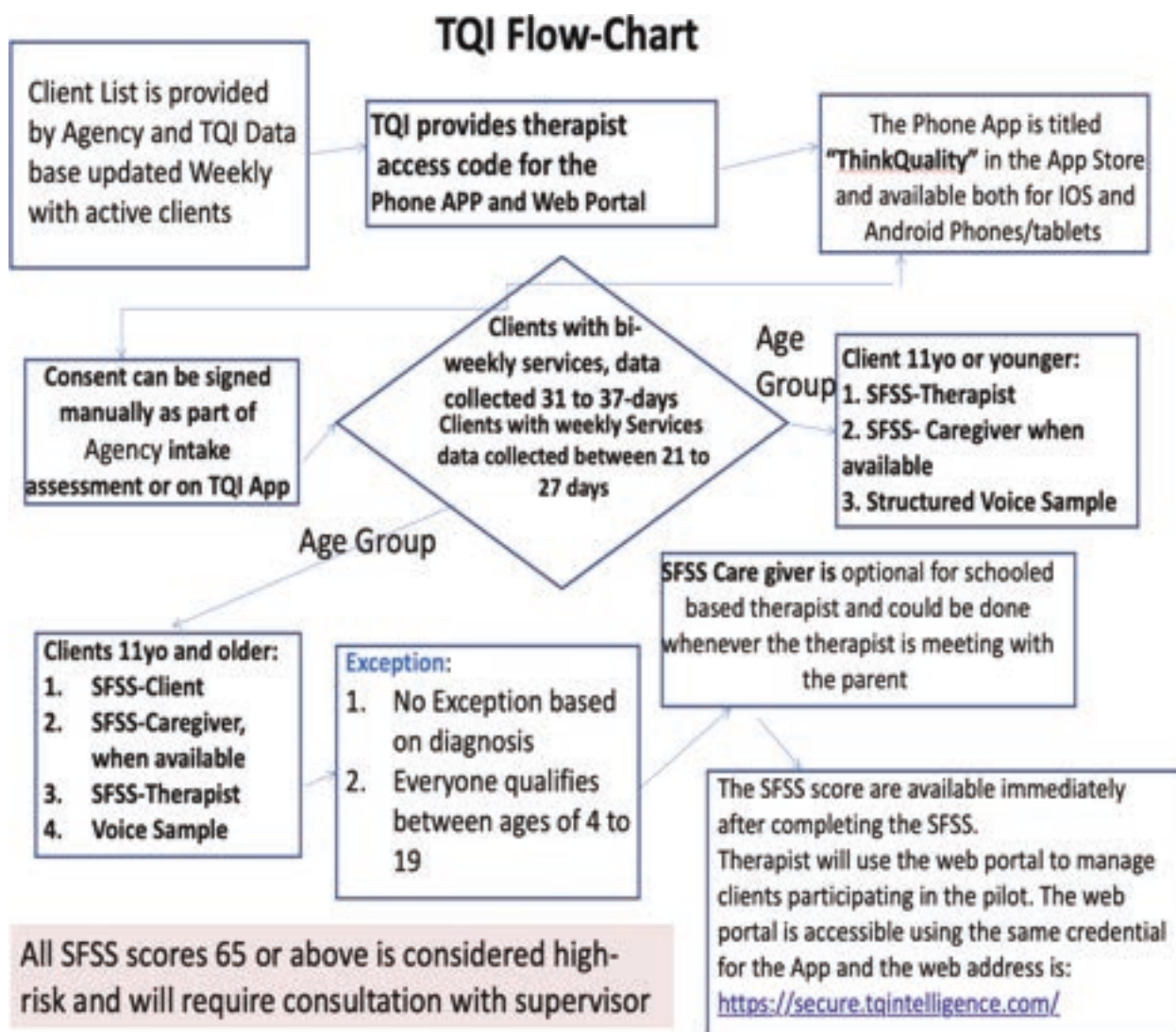
Fig. 8. TQI flow chart of the MBC system. MBC: Measurement-Based Care; SFSS: Symptoms and Functioning Severity Scale; TQI: TQIntelligence.

the 0.001 alpha level, providing robust evidence for the observed effect. A statistically significant result at this level means there is less than a 0.1% probability that the observed correlation occurred by chance, further reinforcing the reliability of the findings. The significant reduction in the participants' internalizing scores as they remained engaged in a treatment plan, which included using the TQI measurement-based system, suggests that this systematic intervention was highly beneficial. This is particularly important for youth from low-income communities, who often have limited access to high-quality, data-driven services and are frequently labeled as treatment-resistant.

The strong negative correlation and the high percentage of explained variance indicate that a consistent intervention, combined with systematic outcome tracking, can significantly benefit youth experiencing internalizing symptoms. The use of the TQI measurement-based

system to track outcomes and guide treatment decisions appears to play a crucial role in these improvements. By providing real-time feedback and data-driven insights, the TQI system helps tailor interventions to the specific needs of each participant, thereby enhancing the overall effectiveness of the counseling sessions.

In summary, the findings highlight the importance of sustained engagement in counseling and the use of advanced measurement tools in achieving significant reductions in internalizing symptoms among youth. This evidence supports the continued use and expansion of such interventions, particularly in underserved communities where access to mental health services is often limited. The results can be found in Figure 9, which details the statistical outcomes of the analysis.

Similarly, the youth with elevated SFSS Externalizing scores also benefited from consistent treatment.
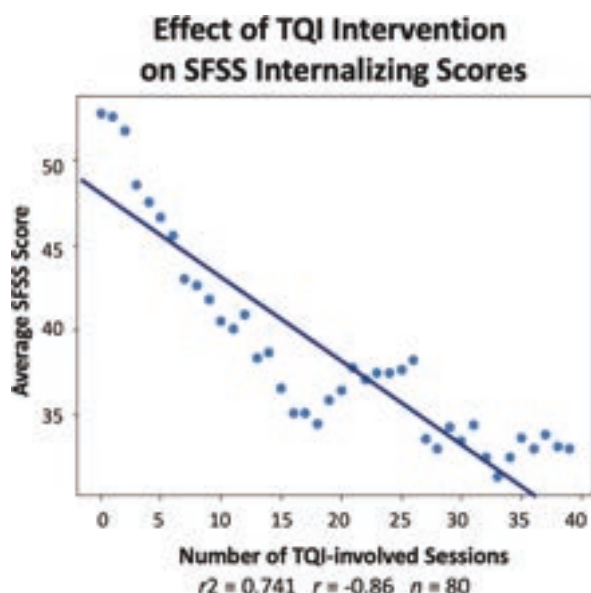
Fig. 9. Effect of TQI system-informed intervention and SFSS internalizing scores. SFSS: Symptoms and Functioning Severity Scale; TQI: TQIntelligence.
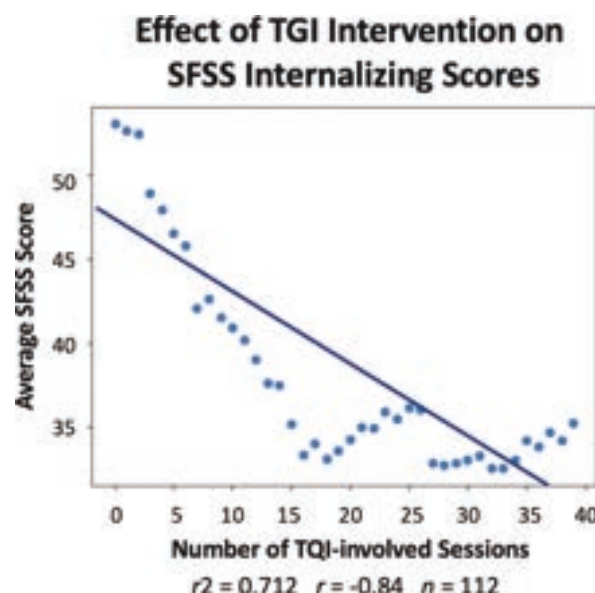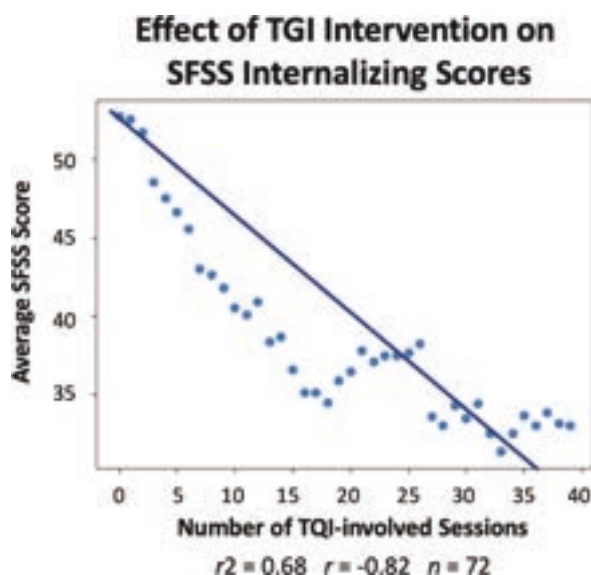

Fig. 10. Effect of TQI intervention and SFSS Externalizing scores. SFSS: Symptoms and Functioning Severity Scale; TQI: TQIntelligence.

Externalizing symptoms are mental health symptoms and behaviors that are outwardly expressed, including disruptive and aggressive behaviors. Results show a negative correlation between SFSS externalizing symptoms and the number of mental health sessions. In other words, there are significant reductions in the youth SFSS Externalizing scores as the youth remained engaged in care that incorporates the TQI measurement-based system. This result is based on the analysis of a sample size of 72, an $r = -0.82$, and $r^2 = 0.68$. Data-driven long-term treatment is

Fig. 11. Results of SFSS total score and TQI Measurement-Based system. SFSS: Symptoms and Functioning Severity Scale; TQI: TQIntelligence.

effective for this population, with significant initial elevations in their SFSS Externalizing scores. The results can be found in Figure 10.

Consistent with psychiatric symptom reduction on the SFSS Externalizing and SFSS Internalizing subscales, the SFSS total score demonstrated a significant decrease in psychiatric severity in youth over time in treatment engagement. There was a substantial negative correlation between the SFSS total scores and the number of therapeutic sessions. This result is based on the analysis of a sample size of 112 with $r = -0.84$ and $r^2 = 0.71$. The results can be found in Figure 11.

Discussion
Principal Findings
Our pilot study revealed patients who participated in mental health treatment that incorporated the TQI MBC system while receiving counseling between 15 and 40 sessions, primarily by master-level unlicensed therapists, demonstrated significant improvement in psychiatric symptom severity as measured by SFSS scores. Patients with two types of symptom expression styles, internalized or externalized presentation, received benefits over 4–10 months. The pilot study is consistent with other MBC studies that use data-driven services, which are more effective partly due to systemic data collection and real-time feedback that allows the therapist to recognize when interventions are failing.

The current behavioral health system of care for racial and ethnic minorities is falling short in addressing the crisis in the mental health needs of children and adolescents.[34] In addition, the shortage of qualified mental healthcare

providers and the financial and time constraints experienced in community clinics shape the dominant service delivery model, one without systematic MBC, that is slow to incorporate new technology and is often delivered by master's level clinicians who are only provisionally licensed. This study adds to a growing body of evidence that MBC should be a routine part of mental health treatment.

The low uptake of MBC and technology-supported interventions in mental health treatment settings, particularly those in low-resourced communities, are indicative of sub-optimal mental health service quality. Given the undue burden of adverse social determinants of mental health borne by members of these communities, the resultant increased risk of mental illness, and the limited resources available to support therapeutic treatment interventions, public sector mental health needs innovative MBC. This study provides promising results for a novel tool that supports the implementation of data-driven services associated with robust treatment outcomes over time and could potentially improve the disparity in mental health treatment outcomes.

Limitations

This pilot study has several limitations. A notable limitation is the small sample size, which impacted the variability in the data for all patient types, including those who only briefly attended or attended services but attended inconsistently. The limited sample size for this pilot decreases the power of analytics and limits the ability to generalize the results. In addition, patient demographics are not sufficiently heterogeneous regarding diagnoses, gender, type of treatment received, and socioeconomic status. Part of the limitations is also driven by the difficulty of comparing two groups, one receiving service and the other not receiving services, which accounts for the lack of control group in this study. Such comparisons may be contrary to the ethical guidelines that will not allow for a control group comparison to determine if the remission of symptoms directly results from using MBC.

Additionally, there was limited information regarding other factors that may affect treatment, including if patients continue to see the same therapist for the entirety of the treatment. Deploying TQI's voice AI technology in community mental health settings presents several challenges. Technological challenges include inadequate infrastructure and unreliable internet connectivity, which are common in low-income or rural areas, as well as ensuring data quality and security and integrating the new technology with existing systems. Logistically, community mental health settings often face resource constraints, including limited staff, funding, and technological tools, allocating resources for implementing and maintaining the AI system difficult. Training mental health professionals to use the technology and providing ongoing technical support are additional logistical hurdles.

Culturally, gaining acceptance and trust from both professionals and patients is crucial. Yet, there may be skepticism and resistance, compounded by the need for the AI system to recognize and interpret a variety of languages and dialects accurately. The stigma surrounding mental health and concerns about privacy can further hinder patient participation. Ethically and legally, obtaining informed consent, ensuring the system is unbiased and fair, and complying with data privacy regulations is a significant challenge that must be navigated carefully to protect patient rights and maintain trust.

Implications and Future Work

The integration of consistent and accessible MBC is crucial in addressing the ongoing mental health crisis among children and adolescents, particularly in low-income communities. The TQI's data-driven MBC system shows benefits in reducing and remitting psychiatric symptoms for children and adolescents in marginalized populations as part of their mental health treatment plan. Future directions for TQI's MBC system include the ongoing improvement of the TQI voice analysis algorithm to improve usability by embracing human-centered design principles so the technology can be seamlessly integrated into complex clinical and nonclinical settings. Valued feedback from participants, providers, clinical staff, and stakeholders will be implemented to ensure the ongoing improvement of TQI's data-driven system. To continue developing and training TQI's voice algorithm, future direction includes developing larger pilot sites and longitudinal studies to track the long-term effectiveness of the intervention of implementing the TQI voice algorithm with multiple diverse pediatric populations, including school, community, and clinical settings, expanding to international, multilingual, and multicultural populations.

Significant consequences accompany the lack of an accurate and timely diagnosis and intervention of psychiatric disorders. The disparities in treatment outcomes in these communities are partly related to inaccurate mental health diagnoses and inadequate treatment planning.[35] There is an urgent call and a significant public interest to improve mental health treatment outcomes by developing an alternative measure of psychiatric severity for clinical decision support. These technologies must be scalable, flexible to diverse clinical workflows, and practical to support communities with a high prevalence of psychiatric disorders. The TQI's data-driven MBC system is a novel approach contributing to mental healthcare practice more broadly, helping to address existing mental health disparities, including patient benefit from the digital health transformation, and pave the way for proactive, collaborative, and efficient approaches to lessen inequity in high-quality, evidence-based mental health treatment intervention.

Conflicts of Interest

Dr Alemu is the Founder and CEO of TQIntelligence and the principal investigator for the NSF SBIR Phase I grant. The company, TQIntelligence.

Data Availability Statement (Das), Data Sharing, Reproducibility, and Data Repositories

Data sharing does not apply to this study as no data sets were generated or analyzed during this study. Preliminary work was completed using a public dataset, and the dataset is cited in the paper.

Application of AI-Generated Text or Related Technology

None was used.

References

1. Boswell JF, Hepner KA, Lysell K, Rothrock NE, Bott N, Childs AW, et al. The need for a measurement-based care professional practice guidelines. Psychotherapy. 2022;60(1):1–16. https://doi.org/10.1037/pst0000439

2. Alemu Y, Chen H, Duan C, Caulley D, Arriaga R, Sezgin E. Detecting clinically relevant emotional distress and functional impairment in children and adolescents: protocol for an automated speech analysis algorithm development study. JMIR Res Protoc. 2023;12:e46970. https://doi.org/10.2196/46970

3. De Jong K, Conijn JM, Gallagher RA, Reshetnikova AS, Heij M, Lutz MC. Using progress feedback to improve outcome and reduce drop-out, treatment duration and deterioration: a multilevel meta-analysis. Clin Psychol Rev. 2021;85:102002. https://doi.org/10.1016/j.cpr.2021.102002

4. Sheftall AH. The tragedy of black youth suicide [Internet]. Washington, DC: Association of American Medical Colleges; 2023 [cited 2024 Jan 05]. Available from: https://www.aamc.org/news/tragedy-black-youth-suicide

5. Substance Abuse and Mental Health Services Administration. HHS releases new national guidelines for improving youth mental health crisis care [Internet]. [place unknown]: SAMHSA; 2022 [cited 2024 Jan 23]. Available from: https://www.samhsa.gov/newsroom/press-announcements/20221110/hhs-releases-new-national-guidelines-improving-youth-mental-health-crisis-care

6. Ehlman DC, Yard E, Stone DM, Jones CM, Mack KA. Changes in the suicide rates-United States 2019–2020. MMWR Morb Mortal Wkly Rep. 2022;71(8):306–12. https://doi.org/10.15585/mmwr.mm7108a5

7. Bridge JA, Ruch DA, Sheftall AH, Hahm HC, O'Keefe VM, Fontanella CA, et al. Youth suicide during the first year of the COVID-19 pandemic. Pediatrics. 2023;151(3):1–15. doi: 10.1542/peds.2022-058375

8. Hoffmann JA, Farrell CA, Monuteaux MC, Fleegler EW, Lee LK. Association of pediatric suicide with county-level poverty in the United States, 2007–2016. JAMA Pediatr. 2020;174(3):287–94. https://doi.org/10.1001/jamapediatrics.2019.5678

9. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr. 2021;175(11):1142–50. https://doi.org/10.1001/jamapediatrics.20212482

10. Substance Abuse and Mental Health Services Administration. Understanding child trauma [Internet]. [place unknown]: SAMHSA; 2023 [cited 2024 Jan 05]. Available from: https://www.samhsa.gov/child-trauma/understanding-child-trauma

11. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. Cogn Behav Pract. 2015;22(1):49–59. https://doi.org/10.1016/j.cbpra.2014.01.010

12. Harding KJK, Rush AJ, Arbuckle M, Trivedi MH, Pincus HA. Measurement-based care in psychiatric practice: a policy framework for implementation. J Clin Psychiatry. 2011;72(8):1136–43. https://doi.org/10.4088/JCP.10r06282whi

13. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith RG, Schoenbaum M, et al. A tipping point for measurement-based care. Psychiatr Serv. 2016;68(2):179–88. https://doi.org/10.1176/appi.ps.201500439

14. Kemp K. Breaking the stigma of mental health: improving the quality of mental healthcare through telehealth and measurement-based care. Texas State Undergrad Res J. 2021;9(i):8–17.

15. Chokshi S, Senathirajah Y, Yadav V, Winsberg M, O'Callaghan E, Sullivan S, et al. A comparative evaluation of measurement-based psychiatric care delivered via specialized telemental health platform versus treatment as usual: a retrospective analysis. Cures. 2022;14(1):2–9. https://doi.org/10.7759/cureus.21219

16. Keepers BC, Easterly CW, Dennis N, Domino ME, Bhalla IP. A survey of behavioral health care providers on use and

barriers to use of measurement based care. Psychiatr Serv. 2023;74(4):331–443. https://doi.org/10.1176/appi.ps.202100735

17. Koutsouleris N, Hauser TU, Skvortsova V, De Choudhury M. From promise to practice: toward the realization of AI-informed mental health care. Lancet Digit Health. 2022;4:e829–40. https://doi.org/10.1016/s2589-7500(22)00153-4

18. Henke RM, Zaslavsky AM, McGuire TG, Ayanian JZ, Rubenstein LV. Clinical inertia in depression treatment. Med Care. 2014;47(9):959–67. https://doi.org/10.1097/MLR.0b013e31819a5da0

19. Graham A, Depp C, Lee EE, Nebeker C, Tu X, Ho-Cheol K, et al. Artificial intelligence for mental health and mental illness: an overview. Curr Psychiatry Rep. 2019;21(11):116. https://doi.org/10.1007/s11920-019-1094-0

20. Tutun S, Johnson ME, Ahmed A, Albizri A, Irgil S, Yesilkaya I, et al. An AI-based decision support system for predicting mental health disorders. Inform Syst Front. 2022;25:1261–76. https://doi.org/10.1007/s10796-022-10282-5

21. Roberts AL, Gilman SE, Breslau J, Breslau N, Koenen KC. Race/ethnic differences in exposure to traumatic events, development of post-traumatic stress disorder, and treatment-seeking for post-traumatic stress disorder in the United States. Psychol Med. 2011;41(1):71–83. https://doi.org/10.1017/S0033291710000401

22. Rariden C, SmithBattle L, Yoo JH, Cibulka N, Loman D. Screening for adverse childhood experiences: literature review and practice implications. J Nurse Pract. 2020;17(1):98–104. https://doi.org/10.1016/j.nurpra.2020.08.002

23. Hughes K, Bellis MA, Hardcastle KA, Sethi D, Butchart A, Mikton C. The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. Lancet Public Health. 2017;2(8):356–66. https://doi.org/10.1016/s2468-2667(17)30118-4

24. Anderson AS, Siciliano RE, Henry LM, Watson KH, Gruhn MA. Adverse childhood experiences, parenting, and socio-economic status: associations with internalizing and externalizing symptoms in adolescence. Child Abuse Neglect. 2022;125:105493. https://doi.org/10.1016/j.chiabu.2022.105493

25. Herzog JI, Schmahl C. Adverse childhood experiences and the consequences on neurobiological, psychosocial and somatic conditions across the lifespan. Front Psychiatry. 2018;9:420. https://doi.org/10.3389/fpsyt.2018.00420

26. Kerker BD, Zhang J, Nadeem E, Stein REK, Hurlburt MS, Heneghan A, et al. Adverse childhood experiences and mental health, chronic medical conditions, and development in young children. Acad Pediatr. 2015;15(5):510–17. https://doi.org/10.1016/j.acap.2015.05.005

27. Bickman L, Athay MM, Riemer M, Lambert EW, Kelley SD, Breda C, et al. Manual of the peabody treatment progress battery [Electronic version]. Nashville, TN: Vanderbilt University; 2010.

28. Riemer M, Athay MM, Bickman L, Breda C, Douglas Kelley S, Vides de Andrade AR. The peabody treatment progress battery: history and methods for developing a comprehensive measurement battery for youth mental health. Adm Policy Ment Health. 2012;39:3–12. https://doi.org/10.1007/s10488-012-0404-1

29. Caulley D, Alemu Y, Burson S, Cárdenas Bautista E, Tadesse GA, Kottmyer C, et al. Objectively quantifying pediatric psychiatric severity using artificial intelligence, voice recognition technology, and universal emotions: a pilot study for artificial intelligence-enabled innovation to address youth mental health crisis. JMIR Res Protoc. 2023;12:e51913. https://doi.org/10.2196/51912

30. Finkelhor D, Shattuck A, Turner H, Hamby S. A revised inventory of adverse childhood experiences. Child Abuse Neglect. 2015;48:13–21. https://doi.org/10.1016/j.chiabu.2015.07.011

31. Monnat SM, Chandler RF. Long-term physical health consequences of adverse childhood experiences. Sociol Q. 2015;56(4):723–52. https://doi.org/10.1111/tsq.12107

32. Jha MK, Grannemann BD, Trombello JM, Clark EW, Eidelman SL, Lawson T, et al. A structured approach to detecting and treating depression in primary care: VitalSign6 project. Ann Fam Med. 2019;17(4):326–35. https://doi.org/10.1370/afm.2418

33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13. doi: 10.1046/j.1525-1497.2001.016009606.x

34. Logroño Naranjo SI, Estrada Brito NA, Núñez Vásconez VA, Rosero Ordóñez EM. Analysis of the use of the python programming language for statistical calculations. Espirales Rev Multidiscip Investig Cientif. 2021;6(2):1–10. https://doi.org/10.31876/er.v6i41.813

35. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the quality chasm: a new health systems for the 21st century. Washington, DC: National Academies Press (US); 2001. https://doi.org/10.17226/10027

36. Mongelli F, Georgakopoulos P, Pato Mt. Challenges and opportunities to meet the mental health needs of underserved and disenfranchised population in the United States. Focus. 2020;18(1):16–24. https://doi.org/10.1176/appi.focus.20190028

Appendix 1
Acronyms defined
ACEs: Adverse Childhood Experiences
AI: artificial intelligence
CPT: Current Procedural Terminology
DSM: Diagnostic and Statistical Manual of Mental Disorders
HHS: US Department of Health and Human Services
HIPPA: The Health Insurance Portability and Accountability Act of 1996
MBC: Measurement-Based Care
ML: machine learning
PHQ-9: Patient Health Questionnaire
SAMHSA: Substance Abuse and Mental Health Services Administration
SFSS: Symptoms and Functioning Severity Scale
TQI: TQIntelligence

Appendix 2
"Alice in Wonderland" Reading Passage
Choice 1
"It's no use going back to yesterday because I was a different person then."
"But I don't want to go among mad people," Alice remarked.
"Oh, you can't help that," said the Cat: "we're all mad here. I'm mad. You're mad."
"How do you know I'm mad?" said Alice.
"You must be," said the Cat, "or you wouldn't have come here."
"She generally gave herself very good advice (though she very seldom followed it)."

Choice 2
"Why, sometimes I've believed as many as six impossible things before breakfast."
"Would you tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to."
"I don't much care where –"
"Then it doesn't matter which way you go."
"If I had a world of my own, everything would be nonsense. Nothing would be what it is, because everything would be what it isn't. And contrary wise, what is, it wouldn't be. And what it wouldn't be, it would. You see?"