Cureus
Part of SPRINGER NATURE

Open Access Original Article

Access, Engagement, and Depression and Anxiety Outcomes in an Employer-Sponsored Virtual Behavioral Health Program: A Retrospective Cohort Study

Review began 10/10/2026
Review ended 10/28/2026
Published 11/03/2026

© Copyright 2026
Krueger et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.119068

Alexandra P. Krueger [1], Jaynice N. Williams [2], Elizabeth Blair [3], Mary M. Cain [4], Heather K. West [1], Lee C. Barfield [5]

1. Analytics, First Stop Health, LLC, Chicago, USA 2. Psychology, First Stop Health, LLC, Chicago, USA 3. Nutrition, First Stop Health, LLC, Chicago, USA 4. Population Health, First Stop Health, LLC, Chicago, USA 5. Internal Medicine, First Stop Health, LLC, Chicago, USA

Corresponding author: Jaynice N. Williams, cwilliams@fshealth.com

Abstract

Background

Mental health (MH) conditions impose a growing burden on US employers through lost productivity, absenteeism, and downstream medical use. Timely access to MH care remains an unmet need, particularly in underserved and rural communities. Employer-sponsored models that integrate behavioral health with primary care may improve access and outcomes, but evidence is limited.

Objective

The primary objective was to describe access to care and engagement among members who completed at least one visit with the virtual MH program at First Stop Health, an integrated employer-sponsored health platform, during 2025. The secondary objective was to describe change in depression and anxiety symptoms among members with clinically elevated baseline symptoms and follow-up assessments.

Methods

This retrospective cohort study included 1,480 members who completed at least one MH visit between January 1 and December 31, 2025 (study cohort), used for all access and engagement analyses. Symptom outcomes were analyzed in a nested outcome cohort (N = 298) with a baseline Patient Health Questionnaire-9 (PHQ-9) or Generalized Anxiety Disorder-7 (GAD-7) score ⩾ 10 and at least two assessments on each instrument. Outcomes included time to first visit, visit completion, remission (any post-baseline score < 5), response (⩾ 50% decrease and score < 10, or remission), and clinical improvement (⩾ five-point decrease, or response or remission).

Results

The median time from scheduling to first visit was seven days; 54.5% (N = 807) of members were seen within seven days and 82.8% (N = 1,226) within 14 days. Access was similar for the 27.3% (N = 404) of members living in mental health professional shortage areas. Two-thirds (66.0%, N = 977) completed two or more visits. Members using two or more service types had lower dropout after the first visit than those using one (25.7% vs. 39.5%; $p < 0.001$).

In the outcome cohort, remission was achieved by 53.7% (N = 160) on the PHQ-9, 44.0% (N = 131) on the GAD-7, and 60.1% (N = 179) on either instrument, and clinical improvement by 62.1% (N = 185) and 59.1% (N = 176), respectively. Members seen within seven days had numerically higher remission (61.9% vs. 57.4%; $p = 0.47$) and clinical improvement (76.7% vs. 67.2%; $p = 0.09$). At the latest assessment, 28.9% (N = 86) were in remission on both instruments.

Conclusions

A virtual, integrated employer-sponsored MH program delivered timely access across diverse geographies, including shortage areas, and most members with elevated symptoms improved. Earlier access to care was associated with better outcomes, and use of multiple service types was associated with greater engagement. These findings support embedding virtual behavioral health within integrated employer-sponsored health platforms.

Categories: Psychiatry, Psychology, Healthcare Technology
Keywords: anxiety, behavioral health, depression, employer-sponsored health care, gad-7, integrated care, mental health access, phq-9, virtual care

How to cite this article
Krueger A P, Williams J N, Blair E, et al. (November 03, 2026) Access, Engagement, and Depression and Anxiety Outcomes in an Employer-Sponsored Virtual Behavioral Health Program: A Retrospective Cohort Study. Cureus 18(11): e119068. DOI 10.7759/cureus.119068

Introduction

Mental health (MH) conditions impose a significant burden on the United States workforce. Anxiety disorders and major depressive disorder account for an estimated $210.5 billion in annual expenditure, of which approximately half is attributable to lost workplace productivity, including absenteeism and presenteeism [1]. Behavioral health complaints were the fifth largest driver of short-term disability claims, and the fourth largest driver of long-term disability claims in the United States as recently as 2017 [2]. Depression alone is estimated to account for 200 million lost workdays annually, costing employers between $17 and $44 billion per year [3]. Access to timely behavioral healthcare remains an issue. An estimated 48% of adults with a behavioral health condition in the United States are not receiving treatment, with stigma, cost, and limited provider availability cited as the primary barriers [4]; more recent national estimates indicate that approximately half of US adults with any mental illness received no MH treatment in 2022 [5]. Even among commercially insured populations, roughly 41% of persons who screened as depressed did not have a diagnosis, and nearly half did not seek out MH services [6]. Of those who do pursue treatment, a substantial share must wait longer than one week to receive care, and many wait a month or more to initiate treatment [7]. Research has demonstrated that for each additional day of wait time, a greater number of patients disengage from care before receiving treatment [8].

Geographic barriers compound the access challenge. Approximately 137 million Americans live in federally designated mental health professional shortage areas (MHPSAs) [9]. Rural and non-metropolitan communities are disproportionately affected, facing both higher rates of untreated mental illness and lower density of behavioral health specialists. Virtual care models have demonstrated potential to narrow this gap, but evidence for employer-sponsored health benefits programs remains limited. Employers are increasingly sponsoring health services, such as primary care and MH, as a strategy to improve workforce health and reduce total cost of care. Approximately 33% of employers with 5,000 or more employees sponsor primary care services. However, fewer than 30% of these also provide behavioral health services [10]. Integration of behavioral health within primary care settings has been shown to improve care coordination, reduce stigma, increase access, and improve clinical outcomes [11].

Evidence supports both integrated and virtual behavioral healthcare. In a Cochrane review of 79 randomized trials, collaborative care improved depression outcomes relative to usual care over zero to six months (standardized mean difference = -0.34) and also improved anxiety outcomes [11]. Delivered remotely, telemedicine-based collaborative care for patients in rural primary care settings produced higher response (odds ratio = 7.74) and remission (odds ratio = 12.69) than practice-based collaborative care in a randomized trial [12]. Psychotherapy delivered by telehealth produces large pre-post improvements (Hedges' g = 0.99; 0.54) with negligible differences from in-person treatment [13]. Among employer-sponsored programs, integrated behavioral health at employer-sponsored health clinics was associated with mean reductions of 24% in Patient Health Questionnaire-9 (PHQ-9) and 31% in Generalized Anxiety Disorder-7 (GAD-7) scores [14]. In a multi-employer workplace MH program, 69.3% and 68.7% of participants showed reliable improvement in depression and anxiety, with large pre-post effect sizes (Cohen's d = 1.11 and 1.21) [15]. Additionally, a video-based blended care cognitive behavioral therapy program offered as an employer benefit reported large pre-post effect sizes (Cohen's d = 1.08 for depression and 1.33 for anxiety) [16]. However, these employer programs delivered behavioral health either in person at worksite clinics [14] or as standalone services separate from primary care [15,16] and did not report access or outcomes by MH professional shortage area or rurality status. Fully virtual platforms that embed behavioral health alongside primary care, health coaching, and urgent care within a single shared record are not as explored in the academic literature. This study addresses these gaps by describing time to a completed first visit, engagement, and symptom change in a fully virtual, integrated employer-sponsored model, stratified by shortage area and rurality status.

First Stop Health operates an integrated, virtual employer-sponsored health platform that includes an MH program offering therapy, psychiatric medication management, and behavioral health coaching. The First Stop Health MH program is designed to provide timely access to evidence-based care, support care coordination with primary care and health coaching services, and support diverse employer populations across geographies, industries, and demographic groups. This study had two objectives among members newly engaged with the First Stop Health MH program during the 2025 calendar year. The primary objective was to describe, among all members who completed at least one visit (the study cohort), (1) access to care (time from scheduling to a completed first visit), overall and by shortage area and rurality status, and (2) engagement (visit completion, dropout after the first visit, and episodes of care), including by care team composition. The secondary objective was to describe change in depression and anxiety symptoms (PHQ-9 and GAD-7 [17-20]) among the subset of the study cohort with clinically elevated baseline symptoms and follow-up assessments (the outcome cohort).

Materials And Methods

First Stop Health's MH program is an integrated, virtual model designed to connect members with the appropriate level of behavioral healthcare. At intake, members are screened for depression with the PHQ-9 and for anxiety with the GAD-7 [17-20] and can self-elect or be directed to individual therapy, psychiatric medication management, or behavioral health coaching, with coordination across primary care and other

First Stop Health services. This integrated design is intended to widen access to appropriate care and support timely engagement across diverse employer populations.

Study design and data source

The behavioral healthcare model provided by First Stop Health includes four key dimensions: provider mix, technology, integrated processes, and the specific behavioral health interventions provided. These dimensions are designed to remove barriers to timely access to behavioral health services and improve communication and coordination among all providers.

Provider Mix

Behavioral health staff includes psychologists, master's-level licensed MH clinicians (i.e., licensed clinical social workers, licensed marriage and family therapists, and licensed professional clinical counselors), and national board-certified health and wellness coaches. Primary care staff that are available to support behavioral health prescribing include physicians and nurse practitioners with specialization in internal medicine, family medicine, and emergency medicine. Additional clinical staff may include registered dietitians, certified diabetes care and education specialists, board-certified health coaches, and registered nurses.

Technology

All MH providers utilize a proprietary electronic health record that’s enabled with tools to facilitate delivery and coordination of care. All providers working with a patient have access to notes from the rest of the care team. Patients receive telehealth care using a combination of phone, video conferencing, and a patient portal that contains the record of their care. Patients use the portal to self-refer to the provider of their choice, communicate with providers, and schedule appointments.

Integrated Processes and Care Teams

At First Stop Health, MH is embedded alongside other healthcare services, specifically primary care, coaching, nutrition support, diabetes education, and urgent care services. The multidisciplinary care team shares a single record and can collaborate via care coordination notes on the platform. With multiple services under one umbrella, First Stop Health providers educate patients about the variety of services that are available.

The provider mix at First Stop Health also enables internal and external referrals. For example, patients can be internally referred by a behavioral health provider to another First Stop Health service, such as diabetes education or nutrition support, and primary care providers can refer to a wide assortment of clinical specialists, including psychiatrists, via external referrals.

Patients can directly schedule an appointment with any clinician that is licensed to practice in the patient’s state of residence. Patients’ preferences for a specific type of behavioral health practitioner and their clinical needs determine with whom they initiate treatment. If a patient initiates treatment with a health coach but is assessed as meeting medical necessity for MH treatment, they are triaged to therapy under the care of a licensed MH clinician. The MH providers have direct support from a small team of clinical leaders who provide direct consultation and support during patient crises and discuss strategies for complex patients.

Behavioral Health Interventions

All patients opting into MH services can self-elect the type of care with which they choose to initiate services. All patients are screened prior to their first appointment via standardized screening tools (PHQ-9 and GAD-7 [17-20]). In the first few visits, patients are comprehensively assessed, and expectations are set for the care plan via shared decision-making. The types of care available are discussed below.

Individual therapy: Psychotherapy delivered virtually as a 40-minute telehealth appointment. The specific treatment delivered may vary slightly because of the individual provider’s specialty and training; however, the employer-sponsored clinic leadership and quality assurance stress the importance of facilitating evidence-based care. The most common treatment approaches are cognitive-behavioral therapy (CBT), acceptance and commitment therapy (ACT), dialectical behavior therapy (DBT), or mindfulness-based interventions.

Mental health coaching: A 30-minute telehealth session that is goal-driven with a scope focused on behavioral change.

Medication appointment: A telehealth appointment with a primary care doctor to review current medications, discuss and get prescriptions for possible new medications, and/or get a referral to a specialist.

There are no predetermined limits on the number of visits a patient could receive.

Study population and analysis samples

The eligible population consisted of all members (employees and dependents >18 years old) across the United States and territories who had access to the MH program in 2025 through an employer-paid benefit and had claimed their First Stop Health account. Members who had completed an MH visit prior to January 1, 2025, were excluded to ensure the analysis reflected new program engagement only. All members were screened for symptoms of depression with the PHQ-9, a nine-item self-report measure of depressive symptom severity, and for anxiety with the GAD-7, a seven-item self-report measure of anxiety symptom severity [17-20]. The PHQ-2 and GAD-2, consisting of the first two items of the PHQ-9 and GAD-7, were used as brief initial screens. Members who screened negative on the PHQ-2 or GAD-2 were not administered the remaining items of that instrument. Negative screens were treated as scores below the clinical threshold. Only PHQ-9 and GAD-7 scores are reported. Member demographics and geographics are collected from employer eligibility file data and member-reported data.

All analyses were conducted within a single study cohort, from which one outcome cohort was nested. The study cohort comprised all members who completed at least one MH visit between January 1, 2025, and December 31, 2025 (N = 1,480) and was used for all access and engagement analyses. The outcome cohort comprised study cohort members with a baseline PHQ-9 ⩾ 10 and/or GAD-7 ⩾ 10 and at least two PHQ-9 and two GAD-7 assessments on file (N = 298) and was used for all symptom outcome analyses. Three groups of study cohort members were not eligible for the outcome cohort. First, members with both baseline scores below 10 (N = 751) did not have clinically elevated symptoms. Because the response threshold (score below 10) was already met at baseline, change in this group could not be assessed against the outcome definitions. Second, change could not be measured without a follow-up assessment, so clinically elevated members with fewer than two assessments on either instrument (N = 359) were excluded. Their baseline characteristics are compared with those of the outcome cohort. Third, members without a baseline score on file for one or both instruments (N = 72) could not be classified by baseline severity. All three groups were retained in the access and engagement analyses.

Care team composition (one versus two or more distinct service types) was analyzed as a member characteristic rather than as a separate cohort: within the study cohort for engagement and within the outcome cohort for symptom outcomes. Service types were psychotherapy, MH coaching, psychiatric medication management, primary care, health coaching, and urgent care.

Data extraction and derivation

The eligible population was identified from members with access to First Stop Health services during the study period who had claimed their account. Members with a completed MH visit before January 1, 2025, with First Stop Health or an external partner were excluded. From the remaining members, we identified those who started an MH intake session (including abandoned intake sessions), then those who scheduled an MH visit (including visits later cancelled or missed), and finally those with one or more completed MH visits, who formed the study cohort. Visits delivered through external partners, which lack scheduling timestamps, were removed upstream, so days to first visit was calculated only from First Stop Health visits with a recorded scheduling date. The outcome cohort was then derived using the criteria described above. Data were drawn from curated analytic data tables, and a unique hashed member identifier was used for all joins so that clinical data remained linked to the correct member throughout the analysis.

Care team composition was defined by the distinct service types each member used, classified by line of business and service type: MH medication appointments, primary care visits, internal therapy, MH coaching, health coaching, and urgent care visits. MH medication appointments and primary care visits were both delivered by primary care providers. MHPSA designation and rural-urban continuum code (RUCC) were assigned using the ZIP code of the address in each member's profile (RUCC 2023 data). MHPSA designations reflect Health Resources and Services Administration (HRSA) designation data as of May 12, 2026, the date the study data snapshot was captured.

Variables

Access variables included the date of visit scheduling (defined as the earliest recorded scheduling timestamp) and the date of the first completed MH visit. Appointment times were selected based on member preference. Days to first visit were calculated as the interval between these two dates. Members without a valid scheduling timestamp were excluded from access metric calculations.

Clinical variables included PHQ-9 and GAD-7 scores collected routinely at MH visits. Baseline scores were defined as the first recorded score of each instrument on file. Follow-up scores were all subsequent assessments recorded during the study period. Medication initiation, number of visits, episode of care status (open or closed), and service type utilization were also captured.

Demographic and geographic variables included age at the time of initial appointment, sex assigned at birth,

gender identity, geographic region (US Census divisions: South, Midwest, West, Northeast), rurality (rural-urban continuum codes (RUCC)) [21], MHPSA designation status [9], member type (employee or dependent), and industry (North American Industry Classification System (NAICS) category) [22].

Outcome definitions

Access outcomes included median and mean days to first visit, and the proportion of members seen within three, seven, and 14 days of scheduling, overall and stratified by demographic and clinical subgroups. The three-, seven-, and 14-day thresholds correspond to conventional benchmarks for timely access to behavioral healthcare. An episode of care began with the first completed MH visit in the study period and continued while completed visits occurred no more than 60 days apart (no-show and cancelled visits were not counted); it ended at the last completed visit once 60 days passed without another completed visit.

Engagement outcomes included the proportion of members who discontinued MH care after the first visit (dropout, defined as completing only one MH visit), and episode of care characteristics (average and median visits per episode, episode duration).

Clinical outcomes for the outcome cohort were defined using all available post-baseline assessments (best observed score). The PHQ-9 [18] and GAD-7 [20] are validated measures of depression and anxiety symptom severity. Remission was defined as any post-baseline PHQ-9 or GAD-7 score below five (indicating minimal or no symptoms). Response was defined as any post-baseline score reflecting both a 50% or greater decrease from baseline and a score below 10, or meeting the remission criterion. Clinical improvement was considered any post-baseline score at least five points lower than baseline or meeting the response or remission criteria. These outcomes are nested. Members meeting remission also meet response, and members meeting response also meet clinical improvement. Outcomes were assessed separately for each instrument and composite rates credit members who met the threshold on either instrument. Instrument-specific outcomes were calculated for all outcome cohort members, including those whose baseline score on that instrument was below the clinical threshold. Analyses used available data, and missing assessments were not imputed.

The remission threshold (score < five) is consistent with clinical guidance from the instrument developers [18,20]. Because remission, response, and clinical improvement were credited at any post-baseline assessment rather than requiring sustained or end-of-treatment status, these definitions yield higher rates than endpoint-based definitions (see Limitations).

Statistical analysis

Access and clinical outcome metrics were stratified by pre-specified subgroups, including geographic region, baseline symptom severity, MHPSA status, rurality, sex, member type, and wait time band (zero-seven days versus eight or more days). Outcome rates for the outcome cohort are reported cumulatively (ever achieving threshold across all follow-up assessments), which reflects a best-observed (cumulative) framing appropriate for this descriptive analysis. Data were extracted and analyzed in Snowflake using Snowflake SQL (2026 releases). Python 3.12 (Python Software Foundation, Wilmington, DE) (snowflake.snowpark, pandas, seaborn, and matplotlib) was used to create the joint score distribution heat maps, and Microsoft Excel (Microsoft 365, version 2609, Microsoft Corporation, Redmond, WA) was used to create the remaining figures from exported data. As this was a descriptive evaluation, p-values for selected subgroup comparisons were calculated in Microsoft Excel using two-sided Fisher's exact tests on 2 × 2 tables and are reported without adjustment for multiple comparisons. Tested comparisons were the proportion seen within three and seven days for severe versus minimal baseline PHQ-9, the proportion seen within seven days for MHPSA versus non-MHPSA and metro versus non-metro members, remission and clinical improvement by wait-time band, and dropout after the first visit, remission, and clinical improvement by number of service types. Differences across subgroups should be interpreted as descriptive findings and not as inferential claims. To assess potential selection bias, baseline characteristics of the outcome cohort were compared with those of clinically elevated members excluded for insufficient follow-up.

Ethics approval

This study was reviewed by the WCG IRB and determined to be exempt human subjects research under 45 CFR 46.104(d)(4)(iii) (WCG protocol: 20263158, 08/19/2026). All analyzed data were collected during the routine provision of care; members consented to treatment at the time of enrollment, and the analytic dataset was de-identified prior to analysis.

Results

Study population and cohort flow

We examined 1,976 patients who scheduled at least one visit, and 1,480 (74.9%) patients completed at least one MH visit in 2025, constituting the study cohort.

Among the 1,480 members of the study cohort, 657 (44.4%) had a baseline PHQ-9 ⩾ 10 and/or GAD-7 ⩾ 10. Of

these, 298 had at least two assessments on each instrument and formed the outcome cohort (20.1% of the study cohort; 45.4% of clinically elevated members). The remaining 359 (54.6%) clinically elevated members were excluded from the outcome cohort for insufficient follow-up data. A further 751 (50.7%) members had both baseline scores below 10, and 72 (4.9%) lacked a baseline score on one or both instruments; neither group was eligible for the outcome cohort, and both remained in the access and engagement analyses. Figure *1* shows the participant flow, and Table *1* describes the study cohort.

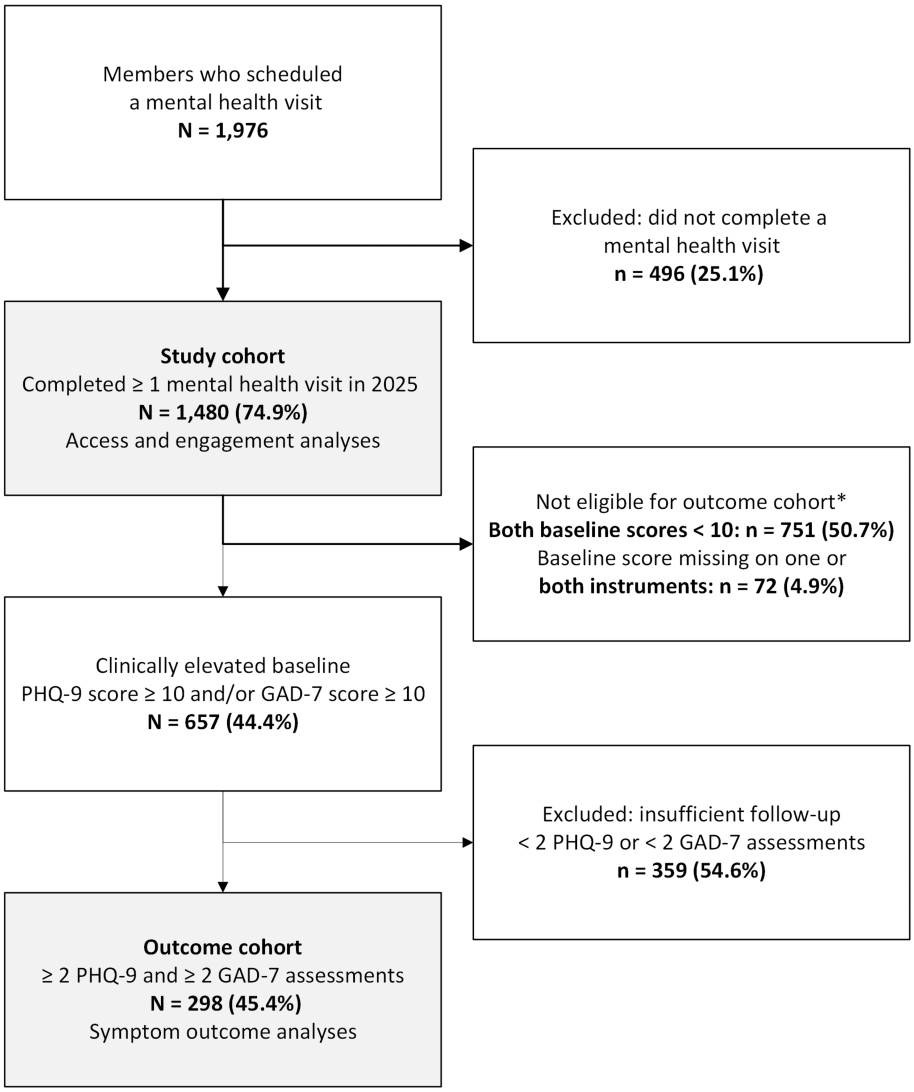


FIGURE 1: Participant flow to the study cohort and nested outcome cohort.

Percentages are of the box immediately above. * Retained in all access and engagement analyses.

Characteristic	Measure	N or Mean (%)	SD/Range
Age (years)	Mean	36.7	SD 10.9; Range 18–72
Sex assigned at birth			
Female	N (%)	1,060 (71.6%)	—
Male	N (%)	412 (27.8%)	—
Unknown/not reported	N (%)	8 (0.5%)	—
Screening tools			
Baseline GAD-7	Mean	7.5	SD 6.7; Range 0–21
Baseline PHQ-9	Mean	6.1	SD 7.2; Range 0–27
Baseline GAD-7 ≥ 10 & Baseline PHQ-9 ≥ 10	N (%)	329 (22.2%)	
Geographic region			
South	N (%)	732 (49.5%)	—
Midwest	N (%)	466 (31.5%)	—
West	N (%)	190 (12.8%)	—
Northeast	N (%)	91 (6.1%)	—
Other	N (%)	1 (0.1%)	—
MHPSA status			
Outside MHPSA	N (%)	1,076 (72.7%)	—
Inside MHPSA	N (%)	404 (27.3%)	—
Rurality (RUCC)			
Metro	N (%)	1,258 (85.0%)	—
Non-metro	N (%)	217 (14.7%)	—
Unknown	N (%)	5 (0.3%)	
Member type			
Employee (primary)	N (%)	1,224 (82.7%)	—
Dependent	N (%)	256 (17.3%)	—

TABLE 1: Baseline demographic and clinical characteristics of the study cohort.

GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9; RUCC = rural-urban continuum code; MHPSA = mental health professional shortage area.

Access to care

Among the 1,480 members in the study cohort, the median time from scheduling to first completed MH visit was seven days (mean = 8.7 days, SD = 7.0). Overall, 24.5% (N = 363) of members were seen within three days, 54.5% (N = 807) within seven days, and 82.8% (N = 1,226) within 14 days of scheduling. Figure *2* shows the proportion of members seen within seven days of scheduling and the median days to first visit, by subgroup. Members in an "Other" region category, without a baseline assessment on file, or without a rurality code are not shown.

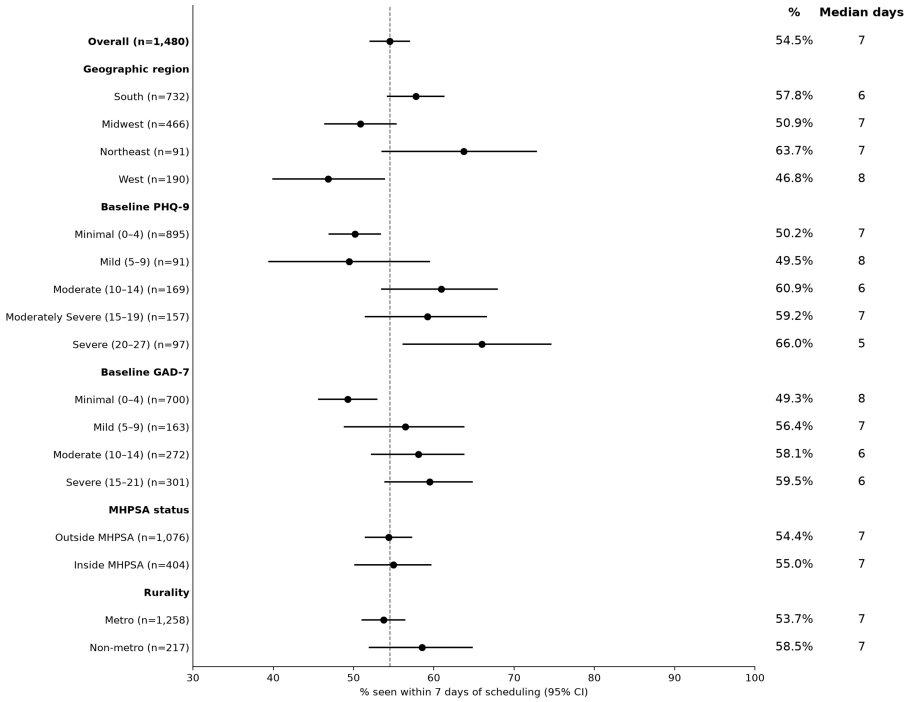


FIGURE 2: Proportion seen within seven days of scheduling by subgroup in the study cohort.

Points show the percentage of members seen within seven days of scheduling; error bars show 95% Wilson confidence intervals. The dashed line marks the overall rate (54.5%). Median days from scheduling to first completed visit are shown on the right.

GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9; MHPSA = mental health professional shortage area.

Members with higher baseline symptom severity accessed care in a timelier manner. Among members with a severe baseline GAD-7 score (15-21), the median time to first visit was six days, compared to eight days for those with minimal GAD-7 scores (0-4). A similar pattern was observed for PHQ-9. Members with severe baseline PHQ-9 scores (20-27) had a median wait time of five days, with 39.2% (N = 38) seen within three days and 66.0% (N = 64) within seven days. The difference between severe and minimal baseline depression was statistically significant at both the three-day (39.2% (N = 38) vs. 21.5% (N = 192), $p < 0.001$) and seven-day (66.0% (N = 64) vs. 50.2% (N = 449), $p = 0.004$) thresholds.

Access to care was similar across all geographic regions. Members in the South had the shortest median wait (six days), while those in the West had the longest (eight days). Members residing in designated MHPSAs demonstrated comparable access to those living outside MHPSAs, with a median of seven days in both groups and 55.0% (N = 222) of MHPSA members seen within seven days compared to 54.4% (N = 585) of non-MHPSA members ($p = 0.86$). Rural (non-metro) members demonstrated access comparable to metro members (median of seven days in both groups; 58.5% (N = 127) vs. 53.7% (N = 676) seen within seven days, a difference that was not statistically significant, $p = 0.21$). Dependents, individuals who are not the primary beneficiary, accessed care slightly faster than employees (median = six vs. seven days).

Engagement

Visit Completion

Among the study cohort (N = 1,480), 66.0% (N = 977) of members completed two or more MH visits in 2025 and 27.6% (N = 409) completed five or more visits. Conversely, 34.0% (N = 503) of members completed only a single MH visit, defined as the dropout rate. Engagement varied somewhat by baseline symptom severity. Members with a moderately severe PHQ-9 baseline (15-19) had the lowest dropout rate after the first visit (27.4%, N = 43), while members with severe PHQ-9 scores (20-27) had a higher dropout rate (42.3%, N = 41). Among GAD-7 severity groups, dropout rates were relatively consistent: 33.0% (N = 231 of 700) for minimal, 37.4% (N = 61 of 163) for mild, 32.4% (N = 88 of 272) for moderate, and 32.9% (N = 99 of 301) for severe baseline anxiety.

Episode of Care

A total of 1,808 episodes of care were identified among 1,480 members, reflecting an average of 1.22 episodes per member over the study period. An episode of care is defined as a contiguous run of completed MH visits where no two consecutive visits are more than 60 days apart. The average episode length was 32.1 days (median = 14 days; range = 0-335 days). The average number of visits per episode was 6.0 (median = 3; range = 1-79). Among all episodes, 1,264 (69.9%) were closed by March 2026, with 544 (30.1%) remaining open, consistent with ongoing care for a subset of members.

Integrated Care Teams

Among the 1,480 members in the study cohort, 888 (60.0%) received services from a single MH service type, while 592 (40.0%) utilized two or more service types, constituting the integrated care population. Members utilizing two or more service types demonstrated higher total visit volume (average of 7.8 total visits vs. 3.7 for single-provider members) and higher MH-specific visit volume (5.3 vs. 3.7 MH visits per member), as well as substantially lower MH dropout rates after the first visit (25.7% (N = 152) vs. 39.5% (N = 351); $p < 0.001$). Baseline clinical severity was not directly compared between service-count groups, so this association should not be interpreted causally. In addition, therapy, MH coaching, and medication management are all MH visit types, so members whose care team included two of these necessarily completed at least two MH visits and could not meet the dropout definition; this contributes mechanically to the lower dropout rate in the multi-provider group.

Among the most common two-provider care teams, therapy combined with MH coaching (N = 89) had a 0% dropout rate after the first MH visit, as did therapy combined with medication management (N = 18). Health coaching combined with therapy (N = 26) demonstrated the highest average total visit volume among two-provider care teams (10.7 visits per member). Average total visit volume rose with the number of service types used, reaching 35.7 visits per member among members using five service types. Care team makeup and visit distribution are described in Figure *3*.

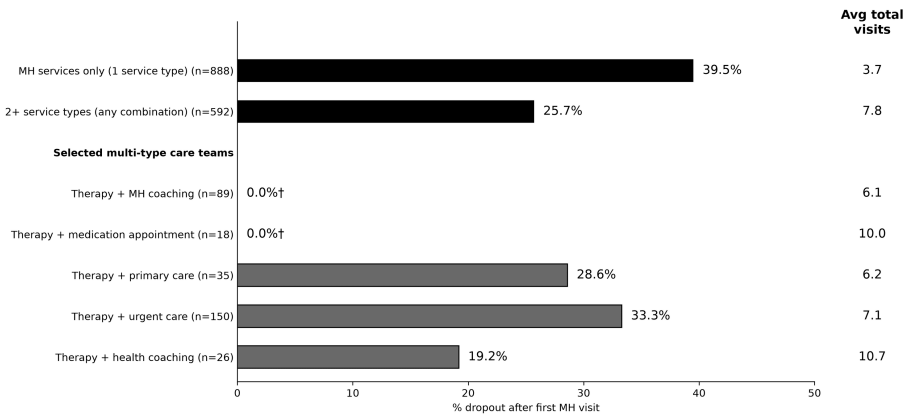


FIGURE 3: Dropout after the first mental health visit by care team type in the study cohort.

Dropout reflects members who completed only one MH visit. † Both service types in this pairing are MH visit types, so members necessarily completed at least two MH visits and could not meet the dropout definition.

MH = mental health (includes therapy, mental health coaching, and medication management visits).

Table *2* presents visit engagement stratified by baseline symptom severity, complementing the access gradient shown in Figure *2*.

Baseline severity	N	≥ 2 visits, N (%)	≥ 5 visits, N (%)	Dropout after 1st visit, N (%)
By baseline PHQ-9 severity				
Minimal (0–4)	895	613 (68.50%)	244 (27.30%)	282 (31.50%)
Mild (5–9)	91	60 (65.90%)	30 (33.00%)	31 (34.10%)
Moderate (10–14)	169	111 (65.70%)	57 (33.70%)	58 (34.30%)
Moderately severe (15–19)	157	114 (72.60%)	39 (24.80%)	43 (27.40%)
Severe (20–27)	97	56 (57.70%)	32 (33.00%)	41 (42.30%)
No baseline assessment	71	—	—	—
By baseline GAD-7 severity				
Minimal (0–4)	700	469 (67.00%)	184 (26.30%)	231 (33.00%)
Mild (5–9)	163	102 (62.60%)	46 (28.20%)	61 (37.40%)
Moderate (10–14)	272	184 (67.60%)	77 (28.30%)	88 (32.40%)
Severe (15–21)	301	202 (67.10%)	95 (31.60%)	99 (32.90%)
No baseline assessment	44	—	—	—

TABLE 2: Engagement by baseline symptom severity in the study cohort.

GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9.

Unlike the access gradient, visit engagement did not increase with baseline symptom severity; members with severe baseline depression (PHQ-9 = 20-27) showed the highest post-intake dropout (42.3%, N = 41). More symptomatic members thus reached care faster but did not accrue more visits, motivating the outcome analysis presented in the next section.

Clinical outcomes

Clinical outcomes were assessed for members with clinically elevated baseline PHQ-9 ⩾ 10 and/or GAD-7 ⩾ 10 who had at least two assessments on file for each instrument (outcome cohort, N = 298). Because the outcome cohort required an elevated baseline score on either instrument rather than both, instrument-specific outcomes include members who were not elevated on that instrument at baseline: 89 members (29.9%) had a baseline PHQ-9 < 5 and 33 (11.1%) had a baseline GAD-7 < 5. These members could meet remission on that instrument without a change in symptoms.

Table *3* presents baseline characteristics of the outcome cohort (N = 298) alongside those of the 359 clinically elevated members excluded for insufficient follow-up. The characteristics of the group with insufficient follow-up did not vary much from the outcome cohort. The majority of outcome cohort members had at least moderate baseline depression or anxiety. A total of 54% (N = 162) of members had exactly two GAD-7 assessments on file, and 58.1% (N = 173) had exactly two PHQ-9 assessments on file. The remainder had three or more assessments (GAD-7: 136 (45.6%); PHQ-9: 125 (41.9%)).

Characteristic	Outcome cohort (N = 298)	Excluded, insufficient follow-up (N = 359)
Age (years), mean (SD)	35.3 (10.4)	35.7 (11.4)
Sex assigned at birth — Female, N (%)	225 (75.5%)	253 (70.5%)
Baseline PHQ-9, mean	11.1	11.4
Baseline PHQ-9 ≥ 10, N (%)	194 (65.1%)	229 (63.8%)
Baseline GAD-7, mean	13.0	13.6
Baseline GAD-7 ≥ 10, N (%)	251 (84.2%)	312 (86.9%)
Both PHQ-9 ≥ 10 and GAD-7 ≥ 10, N (%)	147 (49.3%)	182 (50.7%)
Median days to first visit	7 days	7 days
Median completed visits	3 (range 1–66)	2 (range 1-51)

TABLE 3: Baseline characteristics of the outcome cohort and of clinically elevated members excluded for insufficient follow-up.

Note: PHQ-9 and GAD-7 severity categories are not mutually exclusive; individual members may have both a PHQ-9 ≥ 10 and a GAD-7 ≥ 10.

GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9.

Clinical Improvement

Clinical improvement, defined as any post-baseline score at least five points lower than baseline, or meeting response or remission criteria, was the most inclusive outcome category. At the end of the study period, clinical improvement was observed in 185 of 298 members on the PHQ-9 (62.1%) and 176 of 298 members on the GAD-7 (59.1%). As shown in Figures *4*, *5*, clinical improvement accrual was steeper in the early weeks. By week 10, 45.3% (N = 135) of members had met clinical improvement criteria on the PHQ-9 and 42.6% (N = 127) on the GAD-7.

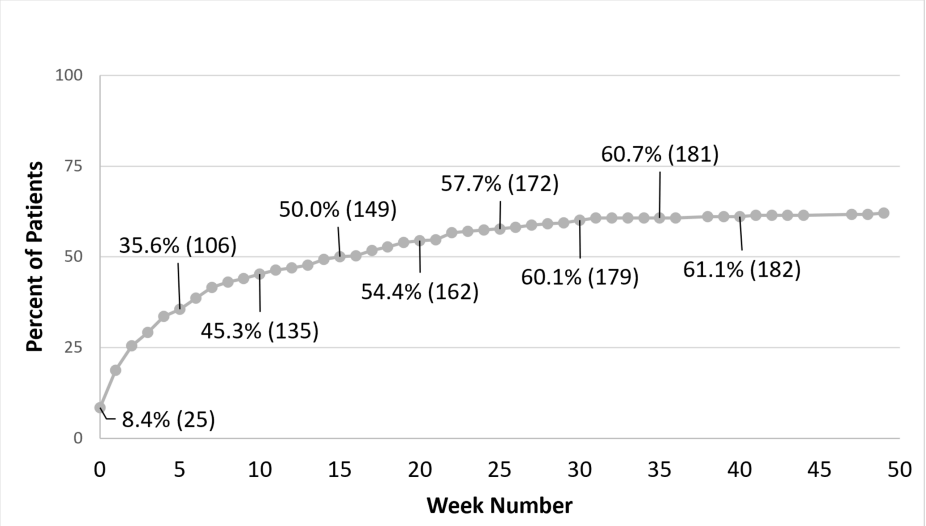


FIGURE 4: Depression clinical improvement in the outcome cohort.

Clinical improvement: Baseline to latest score has ≥ 5 points of improvement OR meets remission or response criteria.

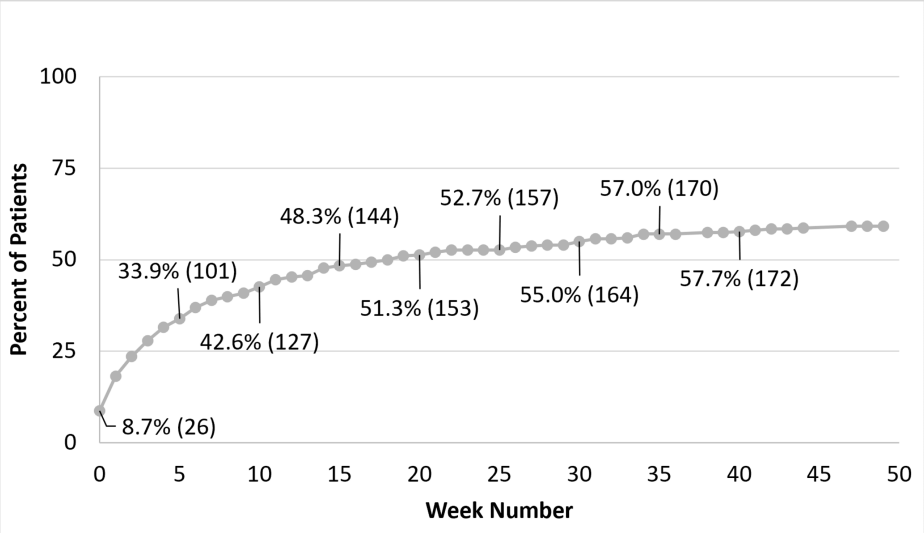


FIGURE 5: Anxiety clinical improvement in the outcome cohort.

Clinical improvement: Any post-baseline score ≥ 5 points lower than baseline, or meeting response or remission criteria. Values are the cumulative percentage (N) of outcome cohort members (N = 298) meeting the criterion by each week.

Response

Response, defined as a ⩾ 50% decrease from baseline and a post-baseline score below 10, or meeting remission criteria, was achieved by 174 of 298 members on the PHQ-9 (58.4%) and 150 of 298 members on the GAD-7 (50.3%) over the study period. As shown in Figures *6*, *7*, response accumulation followed a trajectory broadly similar to remission, with 33.6% (N = 100) of outcome cohort members meeting PHQ-9 response and 29.2% (N = 87) meeting GAD-7 response by week five.

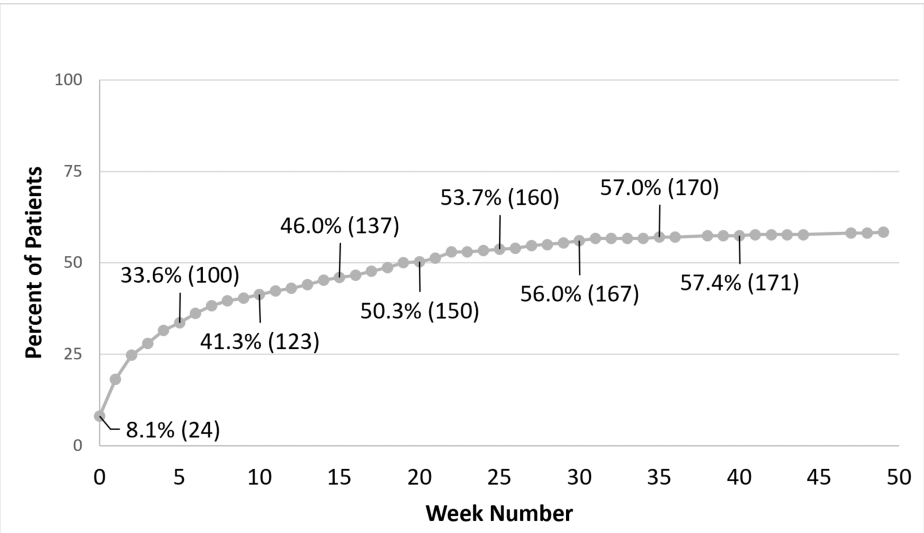


FIGURE 6: Depression response in the outcome cohort.

Response: Any post-baseline score reflecting a ≥ 50% decrease from baseline and a score < 10, or meeting remission criteria. Values are the cumulative percentage (N) of outcome cohort members (N = 298) meeting the criterion by each week.

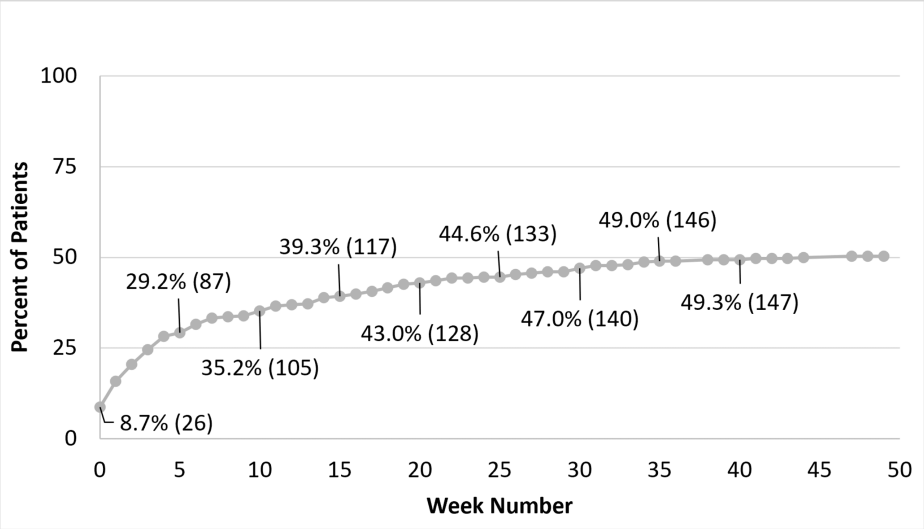


FIGURE 7: Anxiety response in the outcome cohort.

Response: Any post-baseline score reflecting a ≥ 50% decrease from baseline and a score < 10, or meeting remission criteria. Values are the cumulative percentage (N) of outcome cohort members (N = 298) meeting the criterion by each week.

Remission

Using a remission threshold of PHQ-9 or GAD-7 score below five at any follow-up assessment, remission was achieved by 160 of 298 members on the PHQ-9 (53.7%) and by 131 of 298 members on the GAD-7 (44.0%) over the course of the study period. As shown in Figures *8*, *9*, remission accumulation was rapid in the early weeks of treatment. At week five, 32.2% (N = 96) of outcome cohort members had achieved PHQ-9 remission and 28.2% (N = 84) had achieved GAD-7 remission. By week 20, PHQ-9 remission prevalence had reached 47.3% (N = 141), and GAD-7 remission had reached 38.6% (N = 115), with continued gradual improvement. Combined, 179 members (60.1%) met remission on at least one instrument at any follow-up assessment, counted once if they met it on both. Because this count is not restricted to the latest score, it exceeds the dual-remission count at latest assessment.

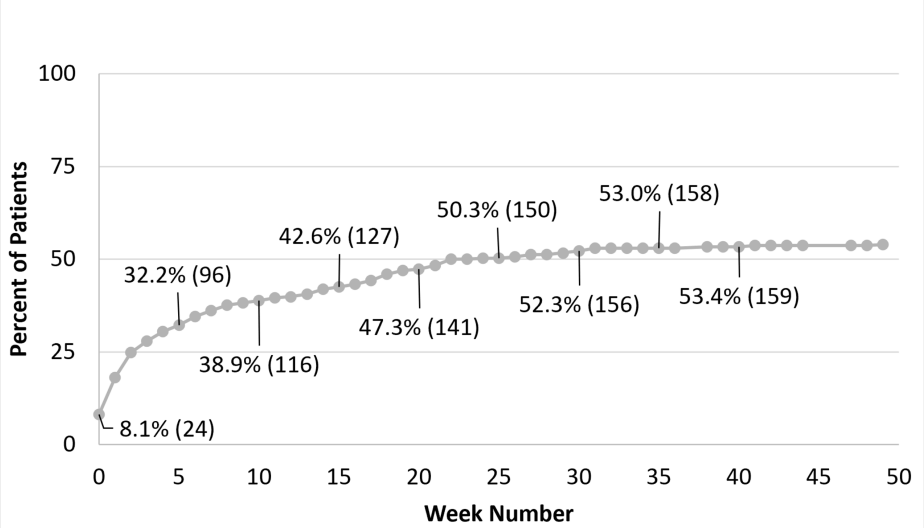


FIGURE 8: Depression remission in the outcome cohort.

Remission: Any post-baseline score < 5. Values are the cumulative percentage (N) of outcome cohort members (N = 298) meeting the criterion by each week.

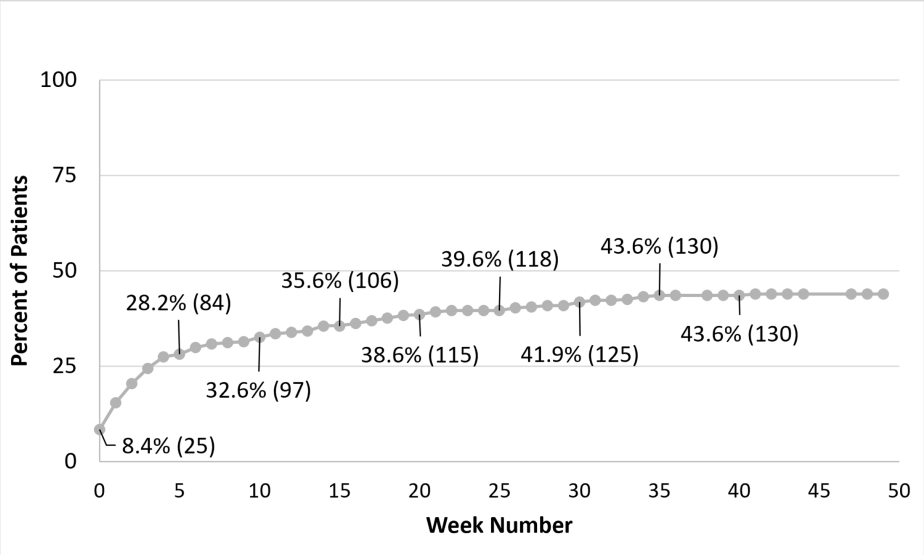


FIGURE 9: Anxiety remission in the outcome cohort.

Remission: Any post-baseline score < 5. Values are the cumulative percentage (N) of outcome cohort members (N = 298) meeting the criterion by each week.

Outcomes by Subgroup

Table *4* presents remission, response, and clinical improvement rates stratified by key subgroups. Members who accessed care within seven days of scheduling demonstrated numerically higher rates across all three outcome categories than those with wait times of eight or more days, although none of the differences was statistically significant (remission: 61.9% (N = 109) vs. 57.4% (N = 70), p = 0.47; response: 69.3% (N = 122) vs. 59.8% (N = 73), p = 0.11; clinical improvement: 76.7% (N = 135) vs. 67.2% (N = 82), p = 0.09). This pattern is consistent with prior evidence that earlier initiation of behavioral health treatment is associated with more durable improvement [14,23].

Subgroup	N	Remission, N (%)	Response, N (%)	Clinical improvement, N (%)
Overall	298	179 (60.10%)	195 (65.40%)	217 (72.80%)
By sex assigned at birth				
Female	225	133 (59.10%)	147 (65.30%)	161 (71.60%)
Male	73	46 (63.00%)	48 (65.80%)	56 (76.70%)
By geographic region				
South	164	89 (54.30%)	99 (60.40%)	114 (69.50%)
Midwest	88	58 (65.90%)	62 (70.50%)	67 (76.10%)
Northeast	20	16 (80.00%)	16 (80.00%)	17 (85.00%)
West	26	16 (61.50%)	18 (69.20%)	19 (73.10%)
Rurality				
Metro	249	152 (61.00%)	165 (66.30%)	186 (74.70%)
Non-metro	47	27 (57.40%)	30 (63.80%)	31 (66.00%)
By wait time band				
0–7 days	176	109 (61.90%)	122 (69.30%)	135 (76.70%)
8+ days	122	70 (57.40%)	73 (59.80%)	82 (67.20%)
By care team type				
1 service type only	247	151 (61.10%)	162 (65.60%)	178 (72.10%)
2+ service types	51	28 (54.90%)	33 (64.70%)	39 (76.50%)
By MHPSA status				
Inside MHPSA	87	54 (62.10%)	59 (67.80%)	64 (73.60%)
Outside MHPSA	211	125 (59.20%)	136 (64.50%)	153 (72.50%)

TABLE 4: Clinical outcomes by subgroup in the outcome cohort.

Outcomes are cumulative (best observed score across all follow-up assessments). Remission: Any post-baseline PHQ-9 or GAD-7 score < five. Response: ≥ 50% decrease from baseline and post-baseline score < 10, or meeting remission criteria. Clinical improvement: Any post-baseline score ≥ five points lower than baseline, or meeting response or remission criteria. Overall outcome rates reported at the final observed time point for the cohort (Week 40+).

GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9; MHPSA = mental health professional shortage area.

Members in MHPSAs demonstrated remission rates comparable to those residing outside MHPSAs (62.1% (N = 54) vs. 59.2% (N = 125)), consistent with the virtual delivery model serving members in areas of traditional provider shortage. Non-metro members had slightly lower remission rates (57.4%, N = 27) compared to metro members (61.0%, N = 152), though small sample sizes in rural subgroups limit interpretability. Members in the Midwest and Northeast demonstrated numerically higher remission rates (65.9% (N = 58) and 80.0% (N = 16), respectively) compared to those in the South and West, though the Northeast subgroup was small (N = 20).

This comparison is limited to the outcome cohort and is separate from the engagement comparison across the full study cohort (N = 592 with two or more service types; Figure *3*). Members utilizing two or more service types (N = 51 in the outcome cohort) demonstrated slightly lower remission rates (54.9%) but higher clinical improvement rates (76.5%) compared to single-provider members (N = 247; remission = 61.1%, clinical improvement = 72.1%); however, none of these differences reached statistical significance given the small integrated-care subgroup (remission, p = 0.44; clinical improvement, p = 0.61), and they may in part reflect higher baseline severity among those routed to integrated care teams.

Joint PHQ-9 and GAD-7 Score Distribution: Baseline vs. Latest Assessment

Figures 10, 11 show the joint distribution of PHQ-9 and GAD-7 severity at baseline and latest available assessment for the outcome cohort. This framing shows the full distribution shift discussed in the clinical outcomes sections across both dimensions.

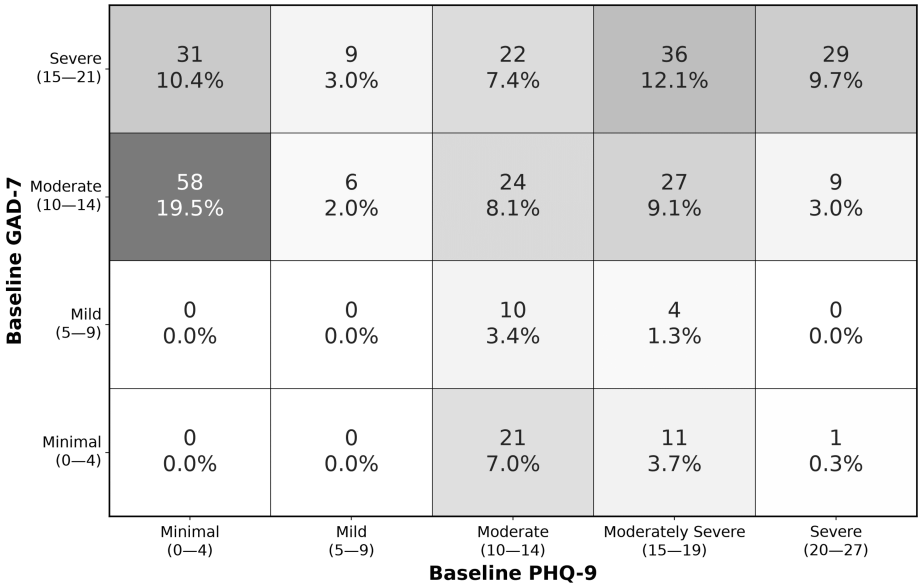


FIGURE 10: Baseline PHQ-9 and GAD-7 scores in the outcome cohort.

Each cell shows the number and percentage of outcome cohort members (N = 298) by baseline PHQ-9 and GAD-7 severity.

GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9.

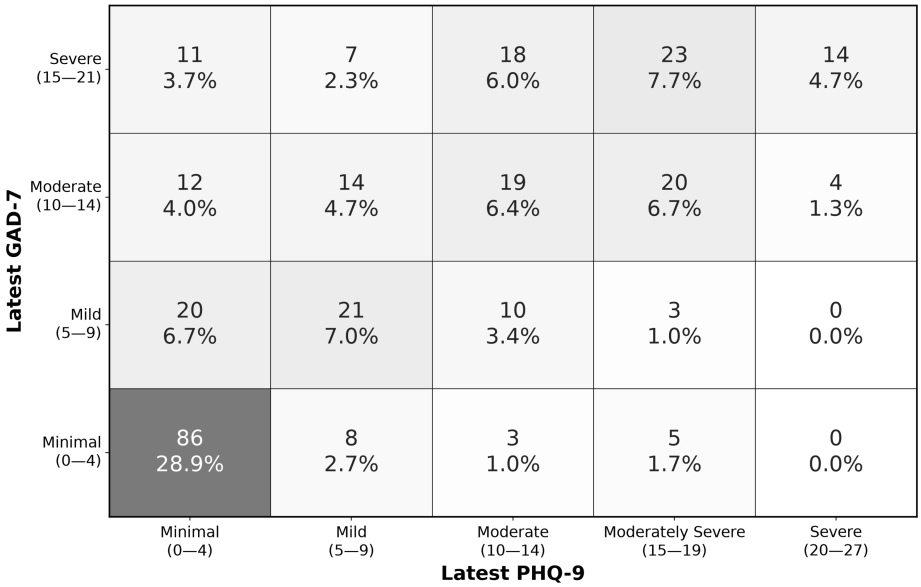


FIGURE 11: Joint latest assessment PHQ-9 and GAD-7 severity distribution in the outcome cohort.

Each cell shows the number and percentage of outcome cohort members (N = 298) by PHQ-9 and GAD-7 severity at the latest available assessment.

GAD-7 = Generalized Anxiety Disorder-7; PHQ-9 = Patient Health Questionnaire-9.

At baseline, symptom burden was concentrated in the moderate to severe range on both axes. The largest single cell was GAD-7 moderate/PHQ-9 minimal (N = 58, 19.5%). This reflects a subgroup whose primary burden was anxiety. A total of 147 members (49.3%) presented with both GAD-7 ⩾ 10 and PHQ-9 ⩾ 10, indicating substantial comorbidity. GAD-7 baseline severity alone accounts for 104 members (34.9%) at baseline.

At the latest assessment, the distribution shifted substantially toward lower severity. Nearly half of the clinically significant members (N = 135, 45.3%) have now shifted into the four bottom-left cells, which display members whose latest scores on both instruments were <10. The most notable change was the lower-left cell (PHQ-9 minimal and GAD-7 minimal), which was empty at baseline. This cell contained 86 members (28.9%) at follow-up, representing dual remission on both instruments.

Discussion

This analysis describes the access, engagement, and clinical outcomes of a virtual, employer-sponsored MH program across a large, geographically diverse employer population during 2025. Four primary findings emerge from this work.

First, the First Stop Health MH program achieved rapid access to behavioral healthcare. The median time from scheduling to first visit was seven days, with more than half of members seen within one week and more than four-fifths seen within two weeks. This compares favorably to community settings where average wait times for an initial behavioral health appointment routinely exceed 30 days [7] and is consistent with findings from prior employer-sponsored integrated care studies [14]. Access was also comparable across geographies. Members in designated MHPSAs demonstrated virtually identical wait times to those outside shortage areas, and non-metro members accessed care at least as quickly as metro members. This pattern suggests that the virtual delivery model may be an effective mechanism for extending behavioral health reach to populations that are traditionally underserved by brick-and-mortar MH systems.

Comparisons across programs should be made cautiously, as access is often measured differently across studies. Although Bondar et al. studied a multi-employer workplace MH program, they reported median times to first available appointment of 1.2 days for psychotherapy and 1.0 day for medication management [15]. However, this access metric reflected appointment availability. Wright et al. used an interval closer to ours, from first visiting the scheduling portal to the first psychotherapy visit, and reported a median of 10 days in employer-sponsored onsite clinics [14]. As defined in this paper, access represents the interval from scheduling to a completed first visit, including member choice of appointment time. Additionally, the finding of equivalent access and outcomes observed between MHPSA and non-MHPSA members is consistent with evidence that virtual psychotherapy is comparable in efficacy to in-person psychotherapy for depressive symptoms [24], supporting the use of virtual delivery to extend MH services into shortage areas without sacrificing effectiveness.

Second, engagement in the program was substantial and improved in integrated care teams. Two-thirds of members who initiated care completed at least two visits, and more than a quarter completed five or more. These engagement levels are consistent with those reported in comparable employer-sponsored behavioral health programs [14]. This return rate after the first visit aligns with community practice. For example, in a large health plan, approximately one-third of adults attending an initial therapy session did not return for a second visit, with evidence suggesting that single-visit episodes may reflect satisfactory treatment rather than disengagement [25]. Dropout estimates from psychotherapy research are lower compared to the findings in this population, averaging 19.7% in a meta-analysis of adult studies [26]. However, these studies utilize different definitions and more selected samples. Visit intensity also aligned with other employer-sponsored programs. Members in this cohort averaged 6.0 visits per episode (median = 3.0), compared with a mean of 5.3 behavioral health visits in the first three months at employer-sponsored clinics [14] and a mean of 5.6 psychotherapy sessions in a multi-employer program [15]. Prior employer-sponsored evaluation has linked greater psychotherapy dose to greater improvement, with diminishing returns beyond a certain level [27], supporting the need for early identification of members at risk of disengagement.

Additionally, across the study cohort, members receiving services from two or more service types had an MH dropout rate of 25.7%, 14 percentage points lower than members with a single service type (39.5%) and accumulated substantially more total and MH-specific visits. The care teams with the lowest dropout rates (0%) included therapy combined with MH coaching and therapy combined with medication management; however, because both pairings consist of two MH visit types, these 0% rates follow from the dropout definition and should not be read as evidence of a protective effect. This pattern is consistent with the collaborative care literature, in which coordinated, team-based models improve depression and anxiety outcomes relative to usual care [11]. In the IMPACT trial, 45% of patients receiving collaborative care achieved a 50% or greater reduction in depressive symptoms at 12 months, compared with 19% in usual care [28].

Third, clinically meaningful symptom improvement was achieved by a majority of members with elevated baseline symptoms. Among the outcome cohort (N = 298), clinical improvement on the PHQ-9 was observed in 62.1% of members and remission (PHQ-9 < 5) in 53.7%. Corresponding rates on the GAD-7 were 59.1%

and 44.0%. These outcomes are broadly comparable to those reported by other employer-sponsored behavioral health programs [14,15]. The most directly comparable study is Bondar et al., which applied nearly identical inclusion criteria (baseline PHQ-9 or GAD-7 ⩾ 10, at least one therapy appointment, and at least two assessments) to 1,132 employees across 66 employers and found that 58.0% achieved reliable improvement with a final PHQ-9 score below 10, and 60.5% achieved the equivalent on the GAD-7 [15]. These comparisons should be interpreted with caution: our instrument-specific rates include all outcome cohort members, including 89 (29.9%) with a baseline PHQ-9 < 5 and 33 (11.1%) with a baseline GAD-7 < 5, who could meet remission on that instrument without symptom change. Our PHQ-9 and GAD-7 rates are therefore likely higher than rates restricted to members elevated on each instrument. Bondar et al. reported depression and anxiety outcomes for subsets of 903 and 837 of their 1,132 participants, respectively, whereas Wright et al., like this study, used their full outcome sample (N = 601) as the denominator for both instruments [14,15]. Wright et al. reported mean reductions of 24% in PHQ-9 and 31% in GAD-7 scores among patients treated in employer-sponsored clinics [14]. Rates in this cohort fall within the range reported by these programs and above pooled estimates from psychotherapy trials for major depression, in which fewer than half of patients met response or remission criteria [29]. At a population level, the Improving Access to Psychological Therapies program reports recovery, defined as moving below clinical thresholds on both depression and anxiety measures, in roughly half of treated patients [30].

The observed association between faster access and better outcomes was reflected in higher remission (61.9%) and clinical improvement (76.7%) rates among members seen within seven days than among those seen after seven days. This is consistent with prior evidence that early initiation of behavioral health treatment is associated with more durable remission [23].

Fourth, members with clinically elevated baseline symptoms demonstrated improvement across both depression and anxiety dimensions simultaneously. Nearly one-third (28.9%) achieved remission on both PHQ-9 and GAD-7 at their latest assessment. This shift, shown in the joint score distribution heat maps, suggests that treatment effects in this population were not confined to a single symptom domain. This finding may be uniquely relevant to employers given that comorbid depression and anxiety carry a disproportionate share of workplace productivity loss and disability burden [31].

Limitations

This analysis has several limitations. First, it was a retrospective, descriptive, single-organization study without a control or comparison group. Causal inference from observational data requires an explicit comparison strategy, such as emulation of a target trial, which this design does not provide [32]. As a result, the associations reported between earlier access and better outcomes, and between integrated care teams and engagement are not causal and may be confounded by unmeasured factors, including baseline severity differences among members routed to multi-provider care teams. In addition, because therapy, MH coaching, and medication management are all MH visit types, members whose care team combined two of these could not meet the dropout definition, which mechanically lowers dropout in the multi-provider group. Second, most outcome cohort members had only two assessments on file, so reported outcomes reflect a cumulative best-observed score across follow-up rather than a sustained trajectory, and remission defined as any post-baseline score below five is a permissive threshold that yields higher rates than definitions requiring sustained or endpoint remission. Clinical outcomes were limited to members with sufficient baseline and follow-up assessments. A total of 359 clinically elevated members were excluded for insufficient follow-up and 72 for a missing baseline score. Table *3* compares the outcome cohort with excluded clinically elevated members. Estimates are subject to selection bias if members who disengaged or were not reassessed differed systematically from those analyzed [33]. Missing follow-up assessments were not imputed, and complete-case estimates can be biased when data are not missing completely at random [34]. Because assessments were administered at visits rather than on a fixed schedule, more engaged members were measured more often, a source of outcome ascertainment bias that session-by-session outcome monitoring is designed to reduce [35]. Regression to the mean may also contribute to improvement from elevated baseline scores [36]. In addition, instrument-specific outcomes include outcome cohort members who were not elevated on that instrument at baseline, which likely raises instrument-specific remission rates. Finally, several subgroups were small (e.g., the Northeast region, N = 20), limiting the interpretability of subgroup comparisons, and the analysis did not include satisfaction, productivity, or cost-of-care outcomes. In addition, all authors are employees of or paid consultants to First Stop Health (see Disclosures) and analyzed the program's own records; although subgroups were pre-specified and the analysis was descriptive, this affiliation is a potential source of bias. Prospective, controlled studies with standardized repeat assessment and longer follow-up are needed to confirm these findings.

Implications for employers

For employers, these findings suggest that a virtual, integrated behavioral health benefit can address two persistent barriers to workforce MH care: delayed access and geographic shortage. Members reached care in a median of seven days, faster than the multi-week waits typical of community settings, and did so at comparable rates inside and outside provider shortage areas and across metro and non-metro geographies. This indicates that the model can extend consistent access to dispersed and rural employee populations that traditional networks underserve. The majority of members with clinically elevated symptoms achieved

meaningful improvement in both depression and anxiety, and members who engaged across multiple service types showed higher retention, consistent with coordinated, multi-modal care supporting engagement, although part of this difference reflects how dropout was defined. Because earlier access was associated with better outcomes at a descriptive level, investments that shorten time to first visit may be particularly worthwhile. While causal and financial conclusions await controlled study, the access, equity, and symptom-improvement patterns observed here are directly relevant to the productivity loss and disability burden that depression and anxiety impose on employers.

Conclusions

This descriptive analysis of the First Stop Health MH program suggests that a virtual, integrated employer-sponsored behavioral health model can provide timely access to care for a diverse employer population, sustain meaningful treatment engagement, particularly among members receiving care across integrated provider teams, and achieve clinically significant symptom improvement in the majority of members with elevated baseline depression and anxiety.

Access was comparable across geographic and rurality subgroups, including those in MH provider shortage areas, suggesting that virtual MH may effectively address traditional geographic barriers to behavioral health care. Earlier access to care was associated with better clinical outcomes at a descriptive level, supporting ongoing program investment in reducing time to first visit. Note that prospective studies with control arms and extended follow-up periods are needed to establish causal estimates of program effects and to identify the mechanisms by which integrated care teams drive engagement and outcomes.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Mary M. Cain, Alexandra P. Krueger, Elizabeth Blair, Jaynice N. Williams, Lee C. Barfield, Heather K. West

Drafting of the manuscript: Mary M. Cain, Alexandra P. Krueger, Elizabeth Blair, Jaynice N. Williams, Lee C. Barfield, Heather K. West

Critical review of the manuscript for important intellectual content: Mary M. Cain, Alexandra P. Krueger, Elizabeth Blair, Jaynice N. Williams, Lee C. Barfield, Heather K. West

Acquisition, analysis, or interpretation of data: Alexandra P. Krueger, Elizabeth Blair, Heather K. West

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. WCG Clinical Services issued approval 20263158. This study was reviewed by the WCG IRB and determined to be exempt human subjects research under 45 CFR 46.104(d)(4)(iii) (WCG protocol 20263158, 08/19/2026). All analyzed data were collected during the routine provision of care; members consented to treatment at the time of enrollment, and the analytic dataset was de-identified prior to analysis. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** Krueger, Williams, Blair, and Cain declare(s) employment from First Stop Health, LLC. Krueger, Williams, Blair, and Cain are employees of First Stop Health, LLC. Barfield declare(s) employment and stock/stock options from First Stop Health, LLC. Barfield is an employee of First Stop Health, LLC, and holds a <2% interest in the company's total equity. West declare(s) personal fees from First Stop Health, LLC. West is a paid consultant to First Stop Health, LLC. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC: The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry. 2015, 76:155-62. 10.4088/JCP.14m09298
2. Unum Group. Ten-year review of Unum's disability claims shows trends in workplace absences . (2018). Accessed: June 29, 2026: https://www.unumgroup.com/newsroom/2018/May/Ten-year-review-of-Unums-disability-claims-shows-trends-in-workplace-abse....
3. Centers for Disease Control and Prevention. Mental health and chronic diseases . (2026). Accessed: June 29, 2026: https://stacks.cdc.gov/view/cdc/105496.
4. Substance Abuse and Mental Health Services Administration. Key substance use and mental health

indicators in the United States: results from the 2023 National Survey on Drug Use and Health (NSDUH). (2024). Accessed: June 29, 2026: https://www.samhsa.gov/data/report/2023-nsduh-annual-national-report.
5. National Institute of Mental Health. Mental illness . (2023). Accessed: July 20, 2026: https://www.nimh.nih.gov/health/statistics/mental-illness.
6. Meiselbach MK, Ettman CK, Shen K, Castrucci BC, Galea S: Unmet need for mental health care is common across insurance market segments in the United States. Health Aff Sch. 2024, 2:qxae032. 10.1093/haschl/qxae032
7. Sun CF, Correll CU, Trestman RL, et al.: Low availability, long wait times, and high geographic disparity of psychiatric outpatient care in the US. Gen Hosp Psychiatry. 2023, 84:12-7. 10.1016/j.genhosppsych.2023.05.012
8. Gallucci G, Swartz W, Hackerman F: Impact of the wait for an initial appointment on the rate of kept appointments at a mental health center. Psychiatr Serv. 2005, 56:344-6. 10.1176/appi.ps.56.3.344
9. Health Resources and Services Administration. Health Workforce Shortage Areas . (2025). Accessed: June 29, 2026: https://data.hrsa.gov/topics/health-workforce/shortage-areas.
10. National Association of Worksite Health Centers, Mercer. Worksite health centers 2021 survey report . (2021). Accessed: July 20, 2026: https://www.mercer.com/assets/us/en_us/shared-assets/local/attachments/pdf-2021-worksite-health-centers-report.pdf.
11. Archer J, Bower P, Gilbody S, et al.: Collaborative care for depression and anxiety problems . Cochrane Database Syst Rev. 2012, 10:CD006525. 10.1002/14651858.CD006525.pub2
12. Fortney JC, Pyne JM, Mouden SB, et al.: Practice-based versus telemedicine-based collaborative care for depression in rural federally qualified health centers: a pragmatic randomized comparative effectiveness trial. Am J Psychiatry. 2013, 170:414-25. 10.1176/appi.ajp.2012.12050696
13. Fernandez E, Woldgabreal Y, Day A, Pham T, Gleich B, Aboujaoude E: Live psychotherapy by video versus in-person: a meta-analysis of efficacy and its relationship to types and targets of treatment. Clin Psychol Psychother. 2021, 28:1535-49. 10.1002/cpp.2594
14. Wright JR, Hagg HK, Watts SA, Nelson RS, Keller JM, Taylor K, Bravata DM: Integrating behavioral health services at employer-sponsored health clinics: a descriptive analysis. Fam Syst Health. 2020, 38:346-58. 10.1037/fsh0000545
15. Bondar J, Babich Morrow C, Gueorguieva R, et al.: Clinical and financial outcomes associated with a workplace mental health program before and during the COVID-19 pandemic. JAMA Netw Open. 2022, 5:e2216349. 10.1001/jamanetworkopen.2022.16349
16. Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CE: Blended care-cognitive behavioral therapy for depression and anxiety in real-world settings: pragmatic retrospective study. J Med Internet Res. 2020, 22:e18723. 10.2196/18723
17. Kroenke K, Spitzer RL, Williams JB: The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care. 2003, 41:1284-92. 10.1097/01.MLR.0000093487.78664.3C
18. Kroenke K, Spitzer RL, Williams JB: The PHQ-9: validity of a brief depression severity measure . J Gen Intern Med. 2001, 16:606-13. 10.1046/j.1525-1497.2001.016009606.x
19. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B: Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med. 2007, 146:317-25. 10.7326/0003-4819-146-5-200703060-00004
20. Spitzer RL, Kroenke K, Williams JB, Löwe B: A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006, 166:1092-7. 10.1001/archinte.166.10.1092
21. US Department of Agriculture, Economic Research Service. Rural-Urban Continuum Codes . (2026). Accessed: July 20, 2026: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/.
22. US Census Bureau. North American Industry Classification System (NAICS) . (2026). Accessed: July 20, 2026: https://www.census.gov/naics/.
23. Halfin A: Depression: the benefits of early and appropriate treatment . Am J Manag Care. 2007, 13:S92-7.
24. Giovanetti AK, Punt SE, Nelson EL, Ilardi SS: Teletherapy versus in-person psychotherapy for depression: a meta-analysis of randomized controlled trials. Telemed J E Health. 2022, 28:1077-89. 10.1089/tmj.2021.0294
25. Simon GE, Imel ZE, Ludman EJ, Steinfeld BJ: Is dropout after a first psychotherapy visit always a bad outcome?. Psychiatr Serv. 2012, 63:705-7. 10.1176/appi.ps.201100309
26. Swift JK, Greenberg RP: Premature discontinuation in adult psychotherapy: a meta-analysis . J Consult Clin Psychol. 2012, 80:547-59. 10.1037/a0028226
27. Maeng D, Cornell AE, Nasra GS: Utilization of an employee behavioral health program and its effects on outcomes for depression and anxiety disorders. J Occup Environ Med. 2019, 61:812-7. 10.1097/JOM.0000000000001678
28. Unützer J, Katon W, Callahan CM, et al.: Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. JAMA. 2002, 288:2836-45. 10.1001/jama.288.22.2836
29. Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A: The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. J Affect Disord. 2014, 159:118-26. 10.1016/j.jad.2014.02.026
30. Clark DM: Realizing the mass public benefit of evidence-based psychological therapies: the IAPT program . Annu Rev Clin Psychol. 2018, 14:159-83. 10.1146/annurev-clinpsy-050817-084833
31. Hendriks SM, Spijker J, Licht CM, et al.: Long-term work disability and absenteeism in anxiety and depressive disorders. J Affect Disord. 2015, 178:121-30. 10.1016/j.jad.2015.03.004
32. Hernán MA, Robins JM: Using big data to emulate a target trial when a randomized trial is not available . Am J Epidemiol. 2016, 183:758-64. 10.1093/aje/kwv254
33. Hernán MA, Hernández-Díaz S, Robins JM: A structural approach to selection bias . Epidemiology. 2004, 15:615-25. 10.1097/01.ede.0000135174.63482.43
34. Sterne JA, White IR, Carlin JB, et al.: Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009, 338:b2393. 10.1136/bmj.b2393
35. Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B: Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. Behav Res Ther. 2009, 47:910-20.

10.1016/j.brat.2009.07.010

36. Barnett AG, van der Pols JC, Dobson AJ: Regression to the mean: what it is and how to deal with it . Int J Epidemiol. 2005, 34:215-20. 10.1093/ije/dyh299