array

Clinical Outcomes and Quality Indicators at Array Behavioral Care

A Virtual Outpatient Behavioral Health Program Using Measurement-Based Care and Stratified Care Pathways

CLINICAL OUTCOMES REPORT | 2026

Authors

Jennifer Hettema, PhD
Senior Clinical Research Scientist at Array Behavioral Care

Sara Gotheridge, M.D.
Chief Medical Officer at Array Behavioral Care

Marlene McDermott, LMFT, PhD
VP of Therapy and Quality Services at Array Behavioral Care

Barry Lunemann
Director of Business Intelligence at Array Behavioral Care

Abstract

Background

Array Behavioral Care delivers a nationwide virtual outpatient behavioral health program grounded in measurement-based care (MBC), stratified care pathways, and care coordination. Although virtual behavioral health services have expanded rapidly, there remains limited large-scale, real-world evidence describing clinical outcomes, engagement, and patient experience in fully virtual outpatient care models operating under routine clinical conditions. This report summarizes internal quality-improvement findings from Array's "AtHome" program, describing patterns in symptom change, engagement, and patient experience among individuals receiving therapy, psychiatry, or combined services. The purpose is to inform ongoing program refinement and illustrate to current and potential partners how such a model performs in everyday practice.

Methods

We conducted a retrospective analysis of de-identified, aggregated data as part of routine quality improvement. The analysis included patients who completed an initial therapy or psychiatry visit between January 1 and June 30, 2025, with six months of follow-up. Repeated Patient Health Questionnaire – 9 (PHQ-9) and Generalized Anxiety Disorder – 7 (GAD-7) scores were used to assess symptom change. Outcomes included mean score change, remission, response, ≥50% response, clinically significant change, and reliable change. Subgroup analyses examined diagnostic categories, baseline symptom severity, and service line (therapy, psychiatry, or combined care). Engagement, accessibility, and patient experience, measured using Net Promoter Score (NPS), were also evaluated.

Results

The sample included 5,805 patients, of whom 4,715 completed at least one follow-up visit. Patients received therapy-only (27.9%), psychiatry-only (48.4%), and dual-service care (23.6%). At intake, 63% of patients reported moderate or higher symptoms of depression or anxiety (PHQ-9 or GAD-7 ≥10), and 87% reported at least mild symptoms (>5). Patients demonstrated clinically meaningful improvement across all service lines and diagnostic groups. Mean depression scores declined from 10.4 to 6.2 (d = 0.69), with larger reductions among patients with elevated baseline severity (PHQ-9 > 10: 15.8→8.5; d = 1.50). Mean anxiety scores declined from 10.1 to 6.0 (d = 0.71), with substantial improvement among patients with GAD-7 > 10 (15.1→7.9; d = 1.63). Engagement was high, and patient experience ratings were favorable (overall NPS = 86.95).

Conclusions

In this large, routine-care virtual outpatient cohort, patients experienced reliable and clinically significant reductions in depression and anxiety symptoms, with effect sizes in the moderate-to-large range. Outcomes followed expected patterns based on patient acuity and treatment intensity and were superior to benchmarks reported in integrated in-person care settings. These findings demonstrate that a virtual outpatient behavioral health model combining measurement-based care, stratified pathways, and care coordination can deliver high-quality, outcomes-driven care at scale without sacrificing patient experience.

Introduction

Array Behavioral Care

Array Behavioral Care is a national, multidisciplinary virtual behavioral health practice serving children, adolescents, adults, and geriatric populations. The organization delivers a comprehensive continuum of services spanning acute, outpatient, and home-based behavioral health care, supporting patients across a wide range of acuity levels through coordinated, virtual care.

AtHome Outpatient Care

Array's virtual outpatient "AtHome" program provides ongoing behavioral health treatment, including individual, group, couples, and family therapy, as well as medication management. Care is delivered by licensed therapists (e.g., social workers and counselors), psychiatrists, and nurse practitioners, supported by a dedicated Care Coordination team . Evidence-based psychotherapeutic approaches and the full spectrum of clinically appropriate pharmacotherapy options are offered within a unified virtual care environment.

Clinical Infrastructure

Array utilizes Epic, a unified electronic health record that supports high-quality behavioral health care by keeping clinical information organized, accessible, and consistent across encounters. Because Epic is widely used across medical, behavioral health, and community systems, it enables Array clinicians to collaborate seamlessly with acute care, primary care, specialty medicine, and other service partners.

In addition to this broad interoperability, Array has developed custom Epic templates specifically designed to operationalize behavioral health best practices, support measurement-based care (MBC), and enable rigorous outcomes measurement and analysis. Together, this infrastructure reduces fragmentation and supports timely communication by ensuring that all members of a patient's care team can view relevant information, track progress over time, and coordinate interventions. The result is more efficient workflows, fewer gaps in care, and a more holistic, person-centered approach that benefits both patients and providers.

Measurement-Based Care at Array

Care delivery at Array is guided by MBC, a clinical framework that uses systematic, repeated collection of patient-reported outcomes data to inform treatment decisions (Scott & Lewis, 2015). At Array, PHQ-9 (PHQ-9; Kroenke, 2001) and GAD-7 (GAD-7; Spitzer, 2005) measures are routinely collected at intake and every three weeks for all active patients, with additional diagnosis-specific measures deployed as clinically indicated. Embedded clinical decision support tools within Array's custom-designed Epic environment enable automated scoring, risk alerts, and treatment adjustment recommendations.

Meta-analyses and large controlled trials show that MBC is associated with higher response and remission rates, faster reduction of symptoms, and reduced risk of relapse compared to treatment as usual (Lambert et al., 2003; Fortney et al., 2017). These benefits have been replicated across a range of mental health diagnoses and treatment modalities.

Stratified Care Pathways

Array employs a proprietary stratified care framework, grounded in established clinical research and evidence-based care principles, designed to match patients to the most appropriate level and type of care. The framework incorporates multiple dimensions of clinical need, including symptom severity, risk profile, functional impairment, treatment utilization patterns, comorbid conditions, substance use, social determinants of health, and trauma history.

Care pathways function as structured clinical decision-support tools, outlining recommended treatment cadence, criteria for referral across service lines, and thresholds for care coordination involvement. Clear pathway reassignment criteria ensure that patients are continuously re-evaluated and that treatment plans are adjusted as symptoms and risks evolve, supporting delivery of the right level of care at the right time.

Care Coordination

Array's clinicians are supported by a dedicated care coordination team that works directly with patients and in close collaboration with treating providers to promote engagement and progress toward treatment goals. Care coordinators track attendance and help address practical barriers to participation, such as transportation needs, scheduling challenges, and required paperwork.

In addition to supporting logistics, care coordinators serve as a key communication link with primary care and other external providers by facilitating record sharing, obtaining necessary releases, and providing treatment updates. They also contribute to crisis stabilization efforts by identifying early warning signs, arranging urgent assessments or crisis services, and ensuring follow-up after emergency department visits or hospitalizations.

Through close collaboration with the clinical team, care coordinators gather and relay information that supports adherence to treatment recommendations, assist with referrals and insurance navigation, and facilitate transitions between levels of care, helping ensure continuity throughout the patient's treatment journey.

Measuring Clinical Quality in Outpatient Behavioral Health

Patient-reported outcome measures (PROMs) are standardized instruments that capture patients' direct reports of symptoms and mental health status (Johnston et al., 2024). Commonly used measures include the PHQ-9 for depression (Kroenke, 2001) and the GAD-7 for anxiety (Spitzer, 2005). While PROMs are widely used to guide individual clinical care, they are increasingly important at the organizational and systems level, where they support quality improvement initiatives, population health management, and operational decision-making. Since PROMs make both individual and organizational performance more visible, they have become central tools for assessing and improving the quality and effectiveness of mental health care delivery.

Evidence shows that routine PROM collection enables health systems to identify trends in symptom change, monitor outcomes across programs, and evaluate service performance over time (Bonsel et al., 2024). Aggregated PROM data can also inform resource allocation by signaling where additional staffing, training, or program redesign may be needed (Gelkopf et al., 2021). In mental health settings, PROMs provide a sensitive, patient-centered indicator of treatment effectiveness, allowing leaders to assess whether services are producing meaningful improvement at scale.

Early engagement represents another critical indicator of clinical quality in outpatient mental

health care. Psychotherapy follows a curvilinear dose-response pattern, with the greatest symptom improvement occurring in the first several sessions and diminishing returns over time (Robinson, 2020). Despite this, a substantial proportion of patients disengage before receiving sufficient treatment to benefit meaningfully. More than 70% of dropouts occur after the first or second visit (Olfson, 2009), and multiple studies demonstrate that the highest risk of dropout occurs during these early encounters (Fernandez, 2021). Earlier research has shown that approximately 50% of patients in outpatient care discontinue treatment within the first month (Fishman et al., 1999). Because only a minority of patients who enter care receive adequate treatment, and early dropout is a primary contributor to this gap, ensuring patient engagement during these initial, high-impact sessions is essential. Taken together, the disproportionate clinical gains that occur early in treatment and the high rates of premature disengagement make early engagement one of the most meaningful and actionable markers of mental health care quality.

In this report, we outline the methods used to evaluate clinical quality PROMs and engagement outcomes within Array's virtual outpatient "AtHome" program. We then describe the patient sample and report findings related to accessibility, engagement, patient experience, and depression and anxiety indicators. Finally, we contextualize these results by comparing Array's performance with published industry benchmarks from large, integrated care systems and highlighting how this evaluation speaks to the high quality of care offered by Array.

Methods / Approach

We conducted a retrospective analysis of clinical outcomes for Array Behavioral Care "AtHome" patients who completed an initial therapy or psychiatry visit between January 1 and June 30, 2025. Follow-up data were collected for six months post-intake. All data were de-identified and aggregated to ensure confidentiality. This evaluation was conducted as part of Array's internal program monitoring activities for the purposes of quality improvement and demonstrating quality to current and potential partners. Because the work was not designed to contribute to generalizable knowledge and involved only de-identified data, it does not meet the regulatory definition of human subjects research and therefore did not require Institutional Review Board (IRB) review or approval.

Subgroups

Patients were grouped for analysis based on diagnosis, baseline symptom severity, and service line. For depression outcomes, analyses included: all patients, patients with depression or dysthymia diagnosis, and patients with depression or dysthymia diagnosis and initial PHQ-9 score > 10.

For anxiety outcomes, analyses included: all patients, patients with an anxiety diagnosis, and patients with an anxiety diagnosis and an initial GAD-7 score > 10.

Results are further stratified by service line: therapy-only, psychiatry-only, and dual-service care.

Accessibility and Timeliness

Total visit volume during the study period was used as a measure of accessibility, reflecting patients' ability to engage with available clinical services.

Early Engagement

Engagement was assessed by calculating the proportion of patients who completed at least one, two, and three follow-up visits (psychiatry or therapy). We also report the mean number of visits during the six-month observation window overall and by service line.

Satisfaction and Alliance

Patient experience and loyalty were assessed using Net Promoter Score (NPS). Although originally developed for consumer industries, NPS is widely used in healthcare due to its brevity, high response rates, and ease of interpretation. A systematic review of NPS use in healthcare found that it is commonly employed to assess patient experience, though its validity varies by context (Adams et al., 2022).

NPS asks patients to rate how likely they are to recommend Array to a colleague or a friend on a 0–10 scale. Responses are categorized as promoters (9–10), passives (7–8), or detractors (0–6). The score is calculated by subtracting the percentage of detractors from the percentage of promoters, yielding a value from –100 to +100 that reflects overall sentiment toward the organization. NPS surveys were sent to patients following their intake appointment and again after every eighth visit.

Patient Reported Outcome Measures

Changes in PHQ-9 and GAD-7 scores over time were used to assess symptom improvement. To ensure robustness, multiple outcome metrics commonly used in the literature were evaluated, and concordance across metrics was examined to confirm that findings were not dependent on a single definition.

Evaluated metrics included:

- **Change score:** difference between mean intake and last follow-up mean score
- **Remission:** percentage of patients with a score less than 5 at the last follow-up
- **Response:** percentage of patients with a score less than 10 at the last follow-up
- **Response (>50%):** percentage of patients who achieved at least a 50% reduction in their score from baseline at the last follow-up
- **Clinically significant change:** percentage of patients with a score reduction greater than 5 points (for PHQ-9) and 4 points (for GAD-7) at the last follow-up
- **Reliable Change:** percentage of patients achieving both a 50% reduction in score and a final score less than 10 at the last follow-up

Magnitude of Effects

To characterize the magnitude of treatment effects, effect sizes for continuous outcomes were calculated using Cohen's d, a standardized measure expressed in standard-deviation units (Cohen, 1988). Effect sizes were interpreted using conventional benchmarks: small (0.20), medium (0.50), and large (0.80). In addition to within-group effect sizes, aggregate outcomes were compared with published benchmarks from large integrated systems, including Kaiser Permanente outcomes (Coley et al., 2020) and a large integrated care program (Jeffrey et al., 2021).

Results

Sample Description

The analytic sample consisted of 5,805 patients, of whom 4,715 completed at least one follow-up visit allowing for measurement of clinical outcomes. All analyses focus on the clinical outcomes sample with the exception of engagement analyses. During the six-month observation period, 48.4% (n=2,283) received psychiatry services only, 27.9 % (n=1,317) received therapy only, and 23.6% (n=1,115) received dual services, defined as at least one session of both therapy and psychiatry.

Among the 31.0% of patients (n=1,463) for whom ethnicity data were available, 77.3% (n=1,131) identified as not Hispanic and 22.5% (n=329) identified as Hispanic. The sample was predominately female (67%, n=3,142). The mean age was 39.4 (SD=15.1). Patients spanned a wide age range: 1.3% (n=59) were ages 5-12, 2.5% (n=117) were ages 13-17, 40.2% (n=1,895) were ages 18-34, 32.4% (n=1,528) were 35-49, 15.2% (n=715) were 50-64, and 8.5% (n=401) were 65 or older.

Clinical Characteristics

At intake, 63% of patients scored in the clinical range on either the PHQ-9 or the GAD-7 (score >10) and 87% reported at least mild symptoms of depression or anxiety (score >5), indicating a largely symptomatic population at treatment entry.

The most common diagnostic categories and the proportion of patients who had a diagnosis in each category at some point during the six-month treatment window were: 1) anxiety disorders (92.3%, n=4,354), such as *Generalized Anxiety Disorder* and *Post-Traumatic Stress Disorder*; 2) mood disorders (79.6%, n=3,755), such as *Major Depressive Disorder*; 3) attention-deficit, conduct, and disruptive behavior disorders (23.8%, n=1,124), primarily *Attention-Deficit Hyperactivity Disorder*; and 4) adjustment disorders (20.5%, n=965).

A smaller subset of patients met criteria for serious mental illness, including schizophrenia (2.6%, n=123), Bipolar I disorder (5.1%, n=240), and Bipolar II disorder (5.4%, n=253). Most patients (71.1%; n=3,352) had more than one diagnosis.

Accessibility and Timeliness

During the six-month study period, the 4,715 patients completed a total of 34,016 visits, including 18,868 therapy visits and 15,148 psychiatry visits.

Timeliness of access was measured as the number of days between an online or telephone scheduling request and the intake appointment. Across all patients, the mean time to first visit was 5.6 days (SD=7.7). Mean time to intake was 5.1 days (SD=7.3) for psychiatry patients and 6.3 days (SD=8.3) for therapy patients.

Engagement

Of the entire analytic sample (N=5,805), patients completed an average of 6.0 visits (SD=5.6) during the six-month period. Visit volume varied by service line: therapy-only patients completed an average of 6.3 visits (SD=5.6), psychiatry-only patients 3.7 visits (SD=2.4), and patients receiving dual services 11.7 visits (SD=7.3).

Overall, 69% of patients completed 1–6 visits, 20% completed 7–12 visits, and 12% completed 13 or more visits. Engagement beyond intake was high: 81.3% of patients returned for at least one follow-up visit, 70.0% completed a second follow-up, and 60.0% completed a third follow-up visit.

Satisfaction and Alliance

Patient experience, measured using Net Promoter Score (NPS), was consistently high. Overall NPS was 86.95 (SD=1.22). Scores were highest among psychiatry patients (88.54, SD=1.26) and remained strong among therapy patients (85.71, SD=1.13).

Depression Outcomes

As shown in Table 1, patients demonstrated clinically meaningful improvements in depression symptoms across all service lines. In the full patient cohort, mean PHQ-9 scores declined from 10.4 at intake to 6.2 at follow-up, corresponding to a moderate effect size ($d = 0.69$). Across treatment groups, average reductions ranged from 4 to 5 points, with effect sizes generally in the moderate-to-large range, indicating reliable symptom improvement over time.

Improvement followed a clear and consistent pattern based on baseline clinical acuity. Patients with a documented depression diagnosis entered care with higher initial PHQ-9 scores than the overall sample, as did those with baseline PHQ-9 scores ≥ 10. Both higher-acuity groups experienced larger absolute reductions in symptom severity, reflecting greater clinical gains among patients with more severe depression at intake (≥ 10), and both groups demonstrated greater absolute change.

Service line was also strongly associated with both baseline severity at intake and the magnitude of improvement. Patients receiving therapy-only services had the lowest initial symptom severity and showed more modest reductions over time. Patients treated with psychiatry services entered care with higher baseline PHQ-9 scores and demonstrated correspondingly larger improvements. The highest acuity was observed among patients receiving dual services (therapy plus psychiatry), who began treatment with the most elevated symptom severity and experienced the largest overall reductions. The greatest improvement was observed among patients with both a depression diagnosis and baseline PHQ-9 scores above the moderate threshold who received dual services; this subgroup demonstrated an average 7-point reduction in PHQ-9 scores, corresponding to a very large effect size ($d = 1.33$).

Clinical outcome indicators further reinforced these patterns. Remission rates ranged from 18% to 25%, while response rates varied from 32% to 61%, depending on service line and outcome definition. Rates of clinically significant change exceeded 60% for most service lines. Across all metrics, outcomes reflected a graded relationship between baseline acuity and treatment gains, with patients presenting with more severe symptoms achieving proportionally larger absolute improvements.

Table 1. Depression Outcomes

Metric	Description	All Patients	Therapy Patients	Psychiatry Patients	Dual Service Patients
All Patients					
		N= 4,517	N=1,359	N=2,322	N=1,083
Score Change (All)	Initial Mean (SD) → Last Follow-Up Mean (SD); Effect Size (d)	10.4(6.5) → 6.2(5.3); d=.69	9.6 (6.2) → 6.4 (5.3); d= .53	10.1(6.5) → 5.4(4.8); d= .79	12.4 (6.4) → 7.5 (5.9); d=.79
Patients with Depression or Dysthymia Diagnosis					
		N=3,158	N=1,025	N=1,449	N=872
Change Score	Initial Mean (SD) → Last Follow-Up Mean (SD); Effect Size (d)	11.6 (6.4) → 6.8 (5.4); d=.79	10.4 (6.1) → 6.7 (5.3); d=.63	11.7 (6.4) → 6.3 (4.9); d=.92	13.0 (6.4) → 7.8 (5.9); d=.84
Patients with Depression or Dysthymia Diagnosis and Initial PHQ > 9					
		N=1,868	N=524	N=889	N=586
Change Score	Initial Mean (SD) → Last Follow-Up Mean (SD); Effect Size (d)	15.8 (4.4) → 8.5 (5.6); d=1.50	15.3 (4.2) → 8.6 (5.8); d=1.45	15.8 (4.3) → 7.8 (5.1); d=1.73	16.4 (4.6) → 9.4 (6.1); d=1.33
Remission	% with Score < 5 at Last Follow-Up	22.4% N=419/1,868	18.1% N=95/524	24.7% N=220/889	21.8% N=128/586
Response	% with Score < 10 at Last Follow-Up	56.7% N=1,059/1,868	49.0% N=257/524	61.3% N=545/889	54.8% N=321/586
Response (>50%)	% with 50% Reduction in Score at Last Follow-Up	41.6% N=777/1,868	31.9% N=167/524	46.5% N= 413/889	42.5% N= 249/586
Clinically Significant Change	% with Score change of ≥ 5 at Last Follow-Up	61.2% N=1,143/1,868	50.0% N= 262/889	66.9% N=595/889	61.9% N= 363/586
Reliable Change	50% reduction in score and a final score < 10 at Last Follow-Up	40.1% N=749 /1,868	30.9% N= 162/889	44.7% N=397/889	40.8% N= 239/586

Anxiety Outcomes

As summarized in Table 2, patients experienced clinically meaningful reductions in anxiety symptoms across all service lines. In the full patient cohort, mean GAD-7 scores declined from 10.1 at intake to 6.0 at follow-up, corresponding to a moderate effect size (d = 0.71). Reductions of similar or greater magnitude were observed across diagnostic and service-line subgroups.

Symptom improvement closely tracked baseline clinical acuity. Patients with a documented anxiety diagnosis entered care with higher mean GAD-7 scores than the overall sample and demonstrated larger absolute reductions. This pattern was most pronounced among patients with elevated baseline anxiety (GAD-7 > 9). Within this higher-acuity group, mean scores declined from 15.1 to 7.9, corresponding to a very large effect size (d = 1.63). Improvements were robust across service lines, with the largest absolute reductions observed among patients receiving psychiatry services, whose scores decreased from 15.1 to 7.2 (d = 1.83).

Differences by service line aligned with expected gradients of clinical need. Therapy-only patients began treatment with the lowest levels of anxiety and showed smaller, though still meaningful, reductions over time. Patients receiving psychiatry services entered care with more severe symptoms and experienced greater improvement. Patients engaged in dual-service care generally presented with the highest baseline GAD-7 scores and demonstrated substantial symptom reduction over the course of treatment.

Clinical outcome indicators reinforced these patterns. Among patients with an anxiety diagnosis and elevated baseline GAD-7 scores, remission rates (final score < 5) ranged from 18% to 32%, with the highest remission observed among psychiatry-only patients. Response rates (final score < 10) were consistently strong across service lines, ranging from approximately 49% to 65%. The proportion of patients achieving a ≥50% reduction in symptoms was also notable, ranging from 31% to 50%, again highest among those receiving psychiatry services. Rates of clinically significant change (>4-point reduction) exceeded 63% across most pathways and reached 73% among psychiatry-only patients. Rates of reliable change (≥50% reduction and a final score <10) followed a similar pattern, with the highest rates observed in the psychiatry cohort.

Table 2. Anxiety Outcomes

Metric	Description	All Patients	Therapy Patients	Psychiatry Patients	Dual Service Patients
All Patients					
		N=4,409	**N=1,313**	**N=2,281**	**N=1,058**
Score Change (All)	Initial Mean → Last Follow-Up Mean; Effect Size (d)	10.1 (6.0) → 6.0 (5.3); d=.71	9.2(5.8) → 6.4(5.4); d=.49	9.9 (6.1) → 5.2 (5.0); d=.80	12.0 (5.7) → 7.4 (5.7); d=.81
Patients with Anxiety Diagnosis					
		N=3,646	**N=1,073**	**N=1,818**	**N=971**
Change Score	Initial Mean → Last Follow-Up Mean; Effect Size (d)	10.7 (5.9) → 6.3 (5.3); d=.73	9.6 (5.7) → 6.5 (5.4); d=.54	10.8 (5.9) → 5.7 (5.1); d=.86	12.2 (5.7) → 7.5 (5.7); d=.88
Patients with Anxiety Diagnosis and Initial GAD > 9					
		N=2,020	**N=492**	**N=1,035**	**N=638**
Change Score	Initial Mean → Last Follow-Up Mean; Effect Size (d)	15.1 (3.5) → 7.9 (5.7); d=1.63	14.8 (3.4) → 8.9 (5.9); d=1.45	15.1 (3.5) → 7.2 (5.4); d=1.83	15.6 (3.6) → 8.9 (5.9); d=1.43
Remission	% with Score < 5 at Last Follow-Up	27.5% N=556/2,020	17.7% N=87/492	32.2% N=333/1,035	26.3% N=168/638
Response	% with Score < 10 at Last Follow-Up	59.3% N=1,197/2,020	48.6% N=239/492	65.0% N=673/1,035	55.0% N=351/638
Response (>50%)	% with 50% Reduction in Score at Last Follow-Up	43.9% N=887/2,020	30.7% N=151/492	50.0% N=518/1,035	42.2% N=269/638
Clinically Significant Change	% with Score change of ≥ 4 at Last Follow-Up	66.2% N=1,338/2,020	52.4% N=258/492	72.8% N=753/1,035	63.9% N=408/638
Reliable Change	50% reduction in score and a final score < 10 at Last Follow-Up	43.8% N=884/2,020	30.7% N=151/492	49.8% N=515/1,035	42.2% N=269/638

Comparison with National Benchmarks

As shown in Table 3, Array's depression outcomes compare favorably with published benchmarks from large, integrated healthcare systems. When evaluated using consistent, measurement-based definitions of remission and response, outcomes from Array's virtual outpatient program exceed those reported in established in-person integrated care settings.

One benchmark study by Coley et al. (2020) examined outcomes for more than 20,000 patients receiving depression care within a large integrated healthcare system that implemented routine PHQ-9 monitoring. This study established widely used definitions of depression remission and response within MBC and reported remission rates of approximately 22% and response rates near 46% among patients with initial moderate-to-severe baseline symptoms. The study is frequently cited as a gold standard, and the findings often serve as a reference point for evaluating outcomes in mature MBC programs.

A second benchmark study by Jeffrey et al. (2021) examined depression outcomes within a large integrated primary care behavioral health program using a retrospective cohort design. Among patients with elevated baseline PHQ-9 scores, the authors reported a 16% remission rate and a 47% response rate at six months. This study is particularly relevant given its real-world design and use of multidisciplinary care pathways, similar to those used in large health systems.

Against these national comparators, Array's outcomes are notably strong. Among patients with a depression or dysthymia diagnosis and elevated baseline PHQ-9 (>9), Array achieved a 27% remission rate, exceeding both the large integrated healthcare system benchmark (22%) and the integrated care program benchmark (16%). Similarly, Array's 60% response rate surpassed those reported in both comparator studies (46% and 47%, respectively). These differences are clinically meaningful, particularly given the high acuity and diagnostically complex population treated within a fully virtual outpatient setting.

Table 3. Comparison to National Benchmarks

Metric	Description	Array Outcome	Large Integrated Healthcare System[1]	Large Integrated Care Program Outcome[2]
Patients with Depression or Dysthymia Diagnosis and Initial PHQ > 9				
Remission	% with Score < 5 at Last Follow-Up	27%	22%	16%
Response	% with Score < 10 at Last Follow-Up	60%	46%	47%

[1]Coley et al, 2020

[2]Jeffery et al., 2021

Discussion

Main Findings

In this large, real-world cohort from a nationwide virtual outpatient behavioral health program, patients experienced clinically meaningful improvement in both depression and anxiety symptoms. Across the full sample, symptom severity declined substantially over six months, with moderate to large effect sizes and stronger improvements observed among patients with higher baseline acuity.

A clear and consistent gradient emerged by clinical acuity. Patients entering care with elevated PHQ-9 or GAD-7 scores achieved larger absolute reductions in symptoms than those with milder presentations. Differences across service lines followed expected clinical patterns: patients receiving therapy-only services entered care with lower baseline severity and showed more modest improvement, while those receiving psychiatry or combined therapy-plus-psychiatry services presented with higher acuity and demonstrated greater symptom reduction. These gradients were observed for both depression and anxiety outcomes.

These findings align with the underlying design of Array's care model. MBC care and stratified pathways are intended to match patients to treatment intensity based on clinical need and to support timely treatment adjustment as symptoms change. The observed gradients by baseline acuity and service mix suggest that higher-need patients were appropriately matched to medication management and/or combined services and benefited from this alignment.

Comparison with National Benchmarks

When compared with published benchmarks from large, integrated healthcare systems, Array's outcomes were comparable to or exceeded those reported in established in-person settings. Among patients with moderate to severe depression, remission and response rates met or surpassed benchmarks from both a large integrated health system and an integrated primary care behavioral health program.

Importantly, these outcomes were achieved within a fully virtual outpatient setting and among a clinically complex population with high baseline symptom burden and frequent comorbidity. Together, these findings suggest that Array's virtual outpatient behavioral health program, designed around systematic outcome monitoring, stratified care pathways, and care coordination, can deliver clinical outcomes on par with leading integrated in-person models while offering the scalability and accessibility advantages of virtual care.

Array Differentiators

Several features of Array's care model likely contribute to the strong outcomes demonstrated in this evaluation. First, MBC is embedded at scale, with routine PHQ-9 and GAD-7 assessments collected at intake and on a three-week cadence. The data from these assessments flow directly into a unified electronic health record, enabling automated scoring, real-time clinical decision support, and early detection of nonresponse so clinicians can adjust treatment proactively.

Second, Array's stratified care pathways guide initial placement and ongoing treatment decisions, ensuring that patients are matched to the appropriate level of care and service line (therapy, psychiatry,

or combined care) and that care intensity is escalated or de-escalated as clinical needs evolve.

Third, a robust care coordination infrastructure supports continuity by conducting outreach after missed visits, coordinating with external providers, and offering focused support to patients during the early phase of treatment when engagement is most critical.

Finally, patient experience is treated as a core design principle rather than an afterthought; consistently high Net Promoter Scores indicate strong patient satisfaction and reinforce the factors that support therapeutic alliance, retention, and ultimately better clinical outcomes.

Together, these differentiators position Array as a scalable virtual care platform capable of delivering measurable symptom improvement, population-level quality management, and consistently high patient experience. This is an uncommon combination in behavioral health and one that is particularly difficult to achieve across geographically distributed outpatient networks.

Practical and Policy Implications

For health systems and payers, these findings indicate that virtual outpatient behavioral health can deliver benchmark-level remission and above-benchmark response for common conditions at scale when anchored in measurement-base care, stratified pathways, and care coordination. This has implications for value-based arrangements that reward measurable improvement, as Array's care model routinely generates the PROMs and quality signals needed for outcomes-based contracting and population health management.

Conclusion

In a large virtual outpatient program operated under routine clinical conditions, patients experienced robust and clinically meaningful improvements in depression and anxiety, with effect sizes in the moderate-to-large range and response rates comparable to or exceeding published integrated care benchmarks. These outcomes are consistent with a care model that combines MBC, stratified pathways, and care coordination within a unified digital infrastructure. The present findings suggest that high-quality, outcomes-driven behavioral health care can be delivered virtually at scale without sacrificing patient experience.

References

Adams, C., Walpola, R., Schembri, A. M., & Harrison, R. (2022). The ultimate question? Evaluating the use of Net Promoter Score in healthcare: A systematic review. Health Expectations, 25(5), 2328–2339. https://doi.org/10.1111/hex.13577

Bonsel, J. M., Itiola, A. J., Huberts, A. S., Bonsel, G. J., & Penton, H. (2024). The use of patient-reported outcome measures to improve patient-related outcomes – A systematic review. Health and Quality of Life Outcomes, 22(101). https://doi.org/10.1186/s12955-024-02312-4

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Coley, R. Y., Boggs, J. M., Beck, A., Hartzler, A. L., & Simon, G. E. (2020). Defining Success in Measurement-Based Care for Depression: A Comparison of Common Metrics. Psychiatric services (Washington, D.C.), 71(4), 312–318. https://doi.org/10.1176/appi.ps.201900295

Coley, R. Y., Boggs, J. M., Beck, A., Hartzler, A. L., & Simon, G. E. (2020). Defining Success in Measurement-Based Care for Depression: A Comparison of Common Metrics. Psychiatric services (Washington, D.C.), 71(4), 312–318. https://doi.org/10.1176/appi.ps.201900295

Fernández, D., Vigo, D., Sampson, N. A., Hwang, I., Aguilar-Gaxiola, S., Al-Hamzawi, A. O., Alonso, J., Andrade, L. H., Bromet, E. J., de Girolamo, G., de Jonge, P., Florescu, S., Gureje, O., Hinkov, H., Hu, C., Karam, E. G., Karam, G., Kawakami, N., Kiejna, A., Kovess-Masfety, V., … Haro, J. M. (2021). Patterns of care and dropout rates from outpatient mental healthcare in low-, middle- and high-income countries from the World Health Organization's World Mental Health Survey Initiative. Psychological medicine, 51(12), 2104–2116. https://doi.org/10.1017/S0033291720000884

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement based care. Psychiatric Services, 68(2), 179–188. https://doi.org/10.1176/appi.ps.201500439

Gelkopf, M., Mazor, Y., & Roe, D. (2021). A systematic review of patient-reported outcome measurement (PROM) and provider assessment in mental health: Goals, implementation, setting, measurement characteristics and barriers. International Journal for Quality in Health Care, 34(Supplement_1), ii13–ii27. https://doi.org/10.1093/intqhc/mzz133

Jeffrey, J., Klomhaus, A., Aralis, H., Barrera, W., Rosenberg, S., Grossman, M., & Lester, P. (2021). Factors associated with response and remission from depression at 6-months of treatment in a retrospective cohort treated within an integrated care program. BMC health services research, 21(1), 703. https://doi.org/10.1186/s12913-021-06729-1

Johnston, B. C., Patrick, D. L., Devji, T., Maxwell, L. J., Bingham, C. O., Beaton, D., … Guyatt, G. H. (2024). Chapter 18: Patient reported outcomes. In J. P. Higgins et al. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions (Version 6.5). Cochrane. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-18

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x 9: Validity of a brief depression severity measure.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta analysis. Clinical Psychology: Science and Practice, 10(3), 288–301. https://doi.org/10.1093/clipsy.bpg025

Robinson, L., Delgadillo, J., & Kellett, S. (2020). The dose-response effect in routinely delivered psychological therapies: A systematic review. Psychotherapy research : journal of the Society for Psychotherapy Research, 30(1), 79–96. https://doi.org/10.1080/10503307.2019.1566676

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092 7.

Scott, K., & Lewis, C. C. (2015). Using measurement based care to enhance any treatment. Cognitive and Behavioral Practice, 22(1), 49–59. https://doi.org/10.1016/j.cbpra.2014.01.010

Wang PS, Lane M, Olfson M, et al: Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. Archives of General Psychiatry 62:629–640, 2005