Effect of Matching Therapists to Patients vs Assignment as Usual on Adult Psychotherapy Outcomes
A Randomized Clinical Trial

Michael J. Constantino, PhD; James F. Boswell, PhD; Alice E. Coyne, MS; Thomas P. Swales, PhD; David R. Kraus, PhD

➕ **Visual Abstract**

➕ **Supplemental content**

IMPORTANCE Psychotherapists possess strengths and weaknesses in treating different mental health problems, yet performance information is rarely harnessed in mental health care (MHC). To our knowledge, no prior studies have tested the causal efficacy of prospectively matching patients to therapists with empirically derived strengths in treating patients' specific concerns.

OBJECTIVE To test the effect of measurement-based matching vs case assignment as usual (CAU) on psychotherapy outcomes.

DESIGN, SETTING, AND PARTICIPANTS In this randomized clinical trial, adult outpatients were recruited between November 2017 and April 2019. Assessments occurred at baseline and repeatedly during treatment at 6 community MHC clinics in Cleveland, Ohio. To be eligible, patients had to make their own MHC decisions. Of 1329 individuals screened, 288 were randomized. Excluding those who withdrew or provided no assessments beyond baseline, 218 patients treated by 48 therapists were included in the primary modified intent-to-treat analyses.

INTERVENTIONS Therapist performance was assessed pretrial across 15 or more historical cases based on patients' pre-post reporting across 12 problem domains of the routinely administered Treatment Outcome Package (TOP). Therapists were classified in each domain as effective (on average, patients' symptoms reliably improved), neutral (on average, patients' symptoms neither reliably improved nor deteriorated), or ineffective (on average, patients' symptoms reliably deteriorated). Trial patients were randomly assigned to good-fitting therapists (matched group) or were assigned to therapists pragmatically (CAU group). There were multiple match levels, ranging from therapists being effective on the 3 most elevated domains reported by patients and not ineffective on any others (highest) to not effective on the most elevated domains reported by patients but also not ineffective on any domain (lowest). Therapists treated patients in the matched and CAU groups, and treatment was unmanipulated.

MAIN OUTCOMES AND MEASURES General symptomatic and functional impairment across all TOP domains (average z scores relative to the general population mean; higher scores indicate greater impairment), global distress (Symptom Checklist-10; higher scores indicate greater distress), and domain-specific impairment on each individual's most elevated TOP-assessed problem.

RESULTS Of 218 patients, 147 (67.4%) were female, and 193 (88.5%) were White. The mean (SD) age was 33.9 (11.2) years. Multilevel modeling indicated a match effect on reductions in weekly general symptomatic and functional impairment ($\gamma_{110} = -0.03$; 95% CI, −0.05 to −0.01; $d = 0.75$), global distress ($\gamma_{110} = -0.16$; 95% CI, −0.30 to −0.02; $d = 0.50$), and domain-specific impairment ($\gamma_{110} = -0.01$; 95% CI −0.01 to −0.006; $d = 0.60$), with no adverse events.

CONCLUSIONS AND RELEVANCE Matching patients with therapists based on therapists' performance strengths can improve MHC outcomes.

TRIAL REGISTRATION ClinicalTrials.gov Identifier: NCT02990000

Author Affiliations: Department of Psychological and Brain Sciences, University of Massachusetts, Amherst (Constantino, Coyne); Department of Psychology, University at Albany, State University of New York, Albany (Boswell); Psychological and Behavioral Consultants, Beachwood, Ohio (Swales); Outcome Referrals Inc, Framingham, Massachusetts (Kraus).

Corresponding Author: Michael J. Constantino, PhD, Department of Psychological and Brain Sciences, University of Massachusetts, Amherst, 612 Tobin Hall, 135 Hicks Way, Amherst, MA 01003-9271 (constanm@umass.edu).

Mental illness is a major public health problem,[1-4] and even among people who engage mental health care (MHC), more than 60% do not benefit meaningfully from care received.[5,6] These estimates mainly derive from naturalistic studies where outcomes are collected on large patient samples. Such patient-generated data also yield information on MHC therapists, which reveals that therapists differ in their average effectiveness (the between-therapist effect).[7] For example, irrespective of treatment type, above-average therapists are up to twice as effective as below-average therapists.[8-10] Thus, research underscores that improvements in MHC can occur not just by using evidence-based interventions, but also by harnessing therapist performance data.[11]

To this end, multidimensional outcome measures that assess distinct symptom and functioning domains can also assess within-therapist differences in treating different mental health problems, and research has demonstrated that most MHC therapists have performance strengths and weaknesses. In 1 study, therapists who were, on average, reliably effective in treating patients with a given presenting problem had large positive pre-post effect sizes (d ranged from 1.00 to 1.52 across problem domains), whereas therapists who were reliably ineffective in treating a given problem had large negative effect sizes (d ranged from −0.91 to −1.42 across domains).[5] Moreover, underscoring that therapists possess distinct strengths and weaknesses, there were relatively low correlations among their domain-specific competencies. In another study, such effects were stable; therapists who demonstrated effectiveness or ineffectiveness in 1 wave of cases largely continued being effective or ineffective in those same domains over a second wave.[12] These data suggest that in any population of therapists, there is an opportunity to facilitate precision care through therapist specialization. Were therapists to specialize toward their measurement-informed skills and away from their shortcomings, it follows that population-level outcomes could be improved and harm could be reduced. Such personalized case assignment could be a readily scalable complement to efforts at improving patient-treatment matching or therapists' in-session interventions.

Historically, though, MHC stakeholders (ie, patients, therapists, and administrators) are unaware of therapists' outcome-based report cards, which represents a critical gap.[11,13] Without exploiting such information, suboptimal improvement rates may partly owe to a lack of measurement-based approaches to case assignment. That is, because current assignments are typically nonpersonalized and based on convenience or self-defined therapist expertise (which is often overestimated or inaccurate[14,15]), it is largely left to chance whether patients will be assigned to therapists who are historically exceptional or average at treating the patients' primary problems vs being below average in these areas. Conversely, there may be an advantage to intentionally matching patients to therapists based on therapist-level outcome data, and MHC patients have endorsed such matching as a valued priority.[16,17]

Accordingly, we developed a personalized match system based on therapist performance report cards determined with

Key Points

Question Can assigning patients to therapists with empirically determined strengths in treating the patients' specific mental health problem(s) (ie, measurement-based matching) improve the outcomes of naturalistic psychotherapy compared with case assignment as usual?

Findings In this 2-arm, double-blind randomized clinical trial including 48 therapists and 218 outpatients, measurement-based matching promoted significantly greater reductions in patients' general symptomatic and functional impairment, global psychological distress, and domain-specific impairment on patients' most elevated presenting problem over 16 weeks postintake.

Meaning In this study, mental health care was enhanced by prospectively assigning patients to empirically good-fitting therapists, which requires minimal disruptions within a mental health care system.

a multidimensional outcomes tool—the Treatment Outcome Package (TOP).[18] By collecting TOP data from enough patients treated by a given therapist, this outcomes tool can establish the domains in which that therapist is stably effective (historically, on average, their patients' symptoms reliably improved), neutral (historically, on average, their patients' symptoms neither reliably improved nor deteriorated), or ineffective (historically, on average, their patients' symptoms reliably deteriorated).[5,12] As routine patient outcomes monitoring continues to become more commonplace at the individual level,[11] such de facto therapist data can be generated within any MHC system with little to no additional burden, and used to implement our patient-therapist match algorithm. Specifically, after completing the TOP at intake, patients are assigned to therapists with previously established strengths in treating their primary problem(s).

For the first time, to our knowledge, the present study tested the efficacy of patient-therapist matching with an individual-level, double-blind, randomized clinical trial within a community MHC system. Adult outpatients were randomly assigned to matching or case assignment as usual (CAU), and treatment was delivered naturalistically. We hypothesized that patients assigned to empirically well-matched therapists would report greater reduction in general symptomatic and functional impairment (general impairment), global psychological distress, and domain-specific symptomatic or functional impairment on their most elevated (ie, severe) presenting problem domain (domain-specific impairment).

Methods

The trial was conducted across 6 clinics in a single health care system in Cleveland, Ohio. A total of 288 patients were enrolled from November 2017 to April 2019. The trial was approved by the institutional review board of the University of Massachusetts, Amherst, overseen by a data and safety monitoring board, and is registered at ClinicalTrials.gov. Written

informed consent was obtained for all participants. The protocol is in Supplement 1. The study followed the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline.

Participants and Procedures

To be included, therapists needed 15 or more historical cases with outcomes data to establish their pretrial performance profiles (exceeding the recommended minimum of 5 cases)[19] and needed to agree to keep roster openings until meeting their target number of trial cases (approximately 6). Therapists provided written informed consent before treating patients in both matched and CAU conditions and were unaware of their patients' condition assignments. This design minimized administrative disruptions and allowed us to test the within-therapist effect of matching—within a given therapist's study-based caseload, did patients who were matched have better outcomes than those who were not matched?

All adult patients aged 18 to 70 years who naturally presented during the trial period were eligible, as the only exclusion criterion was not making one's own MHC decisions. Once eligibility was confirmed, an intake specialist presented study information, including the focus on examining different case-assignment methods (although both therapists and patients would be unaware of what those methods were). Interested participants received a link to an online consent form and baseline TOP assessment to inform the match system. The project coordinator then allocated consenting patients to a condition based on a preestablished randomization sequence created with an online generator, which was concealed until participants were enrolled. The project coordinator was unaware of therapists' effectiveness report cards. Demographic characteristics were collected for both patients and therapists by self-reporting. Race/ethnicity and sexual orientation were self-reported according to predefined categories. Reporting race/ethnicity and sexual orientation data allowed us to provide information about the generalizability of the results in accordance with the Patient-Centered Outcomes Research Institute's methodology standards.

Intervention

Pretrial, enrolled therapists were classified as effective, neutral, or ineffective on each of 12 outcome domains. This classification was based on an established procedure that involved several steps.[5,10,12] First, in a reference sample of approximately 28 000 adult outpatients receiving naturalistic MHC in diverse settings (ie, outpatient clinics, hospitals, residential centers, or day treatments), machine learning determined the most important patient-level predictors of TOP-based change from pretreatment to posttreatment. For each TOP domain, the resulting best-fitting model generated normative, risk-adjusted change rates. Second, for any new patient, the algorithm compared their personally expected change, based on the aforementioned risk-adjusted norms, with their actual change on each TOP domain. Finally, these data were examined at the therapist level. For each therapist and each TOP domain, we calculated an 86% CI for the patients' mean difference from the risk-adjusted expected outcomes. Then, we used the reliable change index to determine whether therapists' domain-specific mean patient change rates exceeded the scale's measurement error.[20] Therefore, within each domain, an effective therapist was one whose patients, on average, reliably exceeded their expected outcome, a neutral therapist was one whose patients, on average, generally neither exceeded nor fell short of their expected improvement (ie, they changed as would be predicted within the 86% CI), and an ineffective therapist was one whose patients, on average, reliably fell short of their expected change. With the match system primed, new trial patients completed the TOP at intake to inform the intervention. Patients in the matched group were then assigned to a therapist based on an algorithm-generated shortlist of multiple clinicians according to 5 match levels, ranging from highest to lowest:
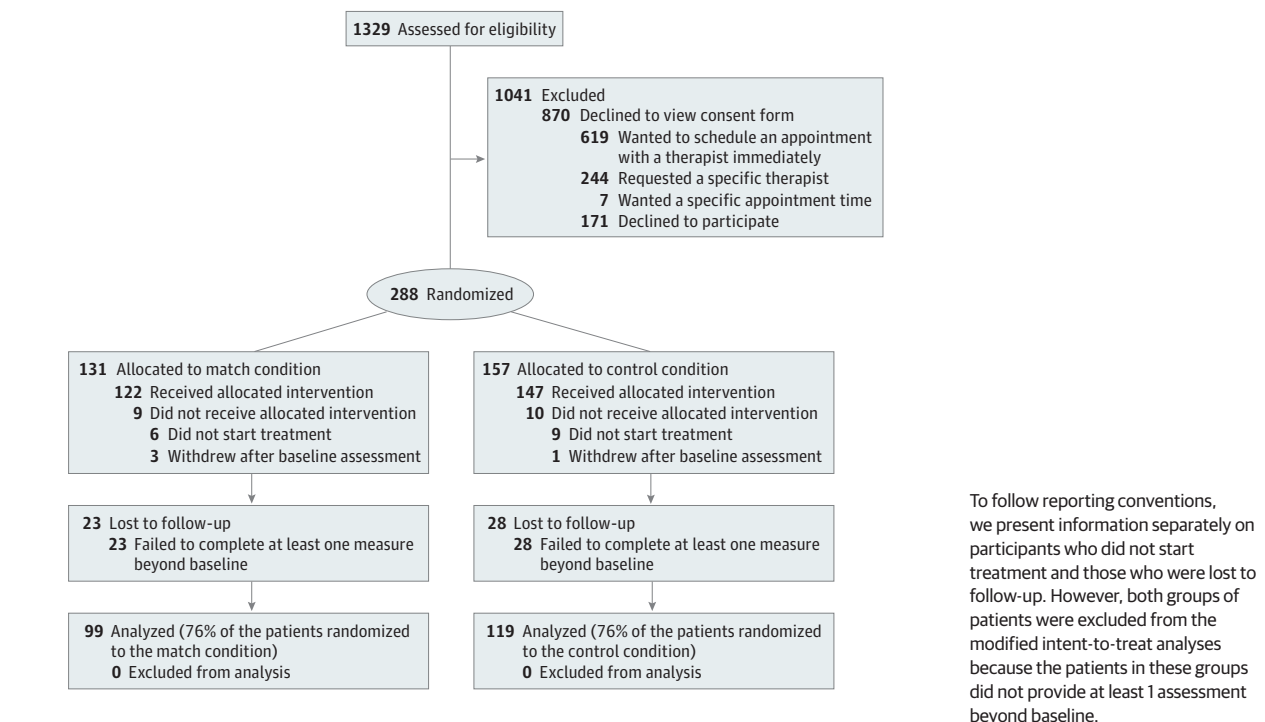
1. Therapist effective on patient's 3 most elevated domains, not ineffective on any domain.
2. Therapist effective on patient's single most elevated domain, not ineffective on any domain.
3. Therapist effective on patient's 3 most elevated domains, ineffective on at least 1 other.
4. Therapist effective on patient's single most elevated domain, ineffective on at least 1 other.
5. Therapist not effective on any elevated domain, but also not ineffective on any domain.

With the shortlist, the project coordinator could navigate from highest to lowest match until there was a therapist who met other necessary parameters for a patient (eg, accepted their insurance). Patients in the CAU group were assigned to therapists through typical pragmatic procedures, such as therapist availability. Following assignment, treatment was delivered naturalistically across both conditions.

Outcomes

The TOP,[18] which was used both for pretrial therapist classification and for 2 primary outcomes during the trial, includes 58 items assessing 12 symptomatic or functional domains in terms of how much of the time the person has experienced each of the following specific concerns in the past 2 weeks on a scale ranging from 0 (none) to 6 (all): depression, quality of life, mania, panic or somatic anxiety, psychosis, substance misuse, social conflict, sexual functioning, sleep, suicidality, violence, and work functioning. The general impairment index represents the mean of the domain-specific z scores (ie, SD units relative to the general population mean). The domain-specific impairment index represents individual z scores throughout treatment for each patient's most elevated baseline clinical scale. Higher scores represent greater impairment (eg, 0 represents the general population mean, 2 represents impairment 2 SDs above the general population mean). The Symptom Checklist-10 (SCL-10)[21] was used to assess global psychological distress. The 10 items are rated from 0 to 4 with a total score range of 0 to 40 (higher scores indicate greater distress). Patients completed the TOP and SCL-10 at baseline and biweekly, either through actual termination (ie, end of their treatment) or a maximum of 16 treatment weeks.

Figure 1. CONSORT Diagram Depicting the Patient Flow Through the Trial

To follow reporting conventions, we present information separately on participants who did not start treatment and those who were lost to follow-up. However, both groups of patients were excluded from the modified intent-to-treat analyses because the patients in these groups did not provide at least 1 assessment beyond baseline.

Statistical Analyses

A power analysis using the formula presented by Raudenbush and Liu,[22] as incorporated in the Optimal Design program version 3.01,[23] revealed that we needed 44 therapists and 211 patients to achieve a power of 0.80 to detect moderate condition effects (d = 0.50) on linear outcome change rates. Accounting for 25% attrition, our enrollment target was 281 patients. For this naturalistic trial, attrition connoted unusable cases in the primary analyses, that is, patients who consented but either actively withdrew consent to analyze their data after starting treatment, passively withdrew by never beginning treatment or receiving a therapist assignment, or provided no outcomes data beyond baseline (lost to follow-up).

To test condition effects, we used hierarchical linear modeling (HLM),[24] as facilitated by HLM software version 7 (Scientific Software International).[25] To address missing data, HLM uses maximum likelihood estimation,[24,26] resulting in a modified intent-to-treat (mITT) sample for our primary analyses, whereby all participants were retained who completed at least 1 outcome assessment beyond baseline, which equated to all nonattrited patients as per the 3 attrition categories described above and in **Figure 1**. (Had any randomized patients failed to start treatment but completed outcome assessments within the study period, we would have included them; however, there were no such cases.) Specifically, 3-level HLMs estimated within-patient (level 1), between-patient (level 2), and between-therapist (level 3) differences. At level 1, a linear change trajectory was fit to each individual's outcome scores. Because of natural variability in treatment length, we centered time at baseline so that the intercept represented patients' outcome level at the time point when all patients had a score. Random patient-level intercepts and slopes and a random therapist-level intercept were included across all models. Therapist-level random slopes were included if they were significant and/or improved model fit. After selecting the best-fitting growth model, we added assignment condition as a level 2 predictor of between-patient (within-therapist) differences in outcome, while accounting for global between-therapist differences at level 3. eAppendix 1 in Supplement 2 shows the equation. Results are presented with 95% CIs and 2-tailed P values. Significance was set at $P < .05$. Effect size is represented as Cohen d modified for a multilevel context,[25] which represents the number of SDs on the patient-level outcome by which the 2 groups differed.

Results

Therapist Sample

Of 48 therapists, 35 (73%) were female. Mean (SD) age was 48.0 (14.0) years. Therapists treated a mean (SD) of 4.54 (2.42) patients, with influences from varied theoretical orientations (**Table 1**). Therapists' pretrial effectiveness report cards were based on pre-post TOP data from a mean (SD) of 28.48 (3.00) patients. Overall, therapists had a mean (SD) of 1.56 (1.66) strengths (domains classified as effective) and 0.96 (1.65) weaknesses (domains classified as ineffective; no therapists were ineffective on 9 or more domains); 42 (87.5%) had at least 1 strength on which they could be matched or at least 1 weakness that could be avoided, and 6 (12.5%) were classified as neutral (neither effective nor ineffective) across all domains, allowing them to be matched at level 5 (eTables 1 and 2 in

Table 1. Demographic and Professional Characteristics of Therapists

Characteristic	Measure
Therapists, No.	48
Age, mean (SD), y	48.0 (14.0)
Sex, No. (%)	
Female	35 (73)
Male	13 (27)
Race/ethnicity, No. (%)[a]	
White	40 (83)
Hispanic	1 (2)
African American	6 (13)
Black	1 (2)
Highest academic degree, No. (%)	
Master's degree	32 (67)
Doctorate in psychology or counseling	15 (31)
Other	1 (2)
Theoretical orientation, mean (SD)[b]	
Psychodynamic/psychoanalytic	2.12 (1.74)
Cognitive-behavioral	5.19 (1.05)
Humanistic/experiential	3.31 (1.66)
Interpersonal	3.91 (1.56)
Systems	2.98 (1.35)
Integrative	4.31 (1.57)
Postdegree experience, mean (SD), y	15.81 (11.73)

[a] Therapists self-reported their race/ethnicity by checking 1 or more of the following predefined categories: White, Hispanic, African American, Black, Asian, Native American/American Indian, East Indian, and Pacific Islander. They could also write in their own description. Reporting therapist race/ethnicity data allowed us to provide information about the generalizability of the results in accordance with the Patient-Centered Outcomes Research Institute's methodology standards.

[b] Self-reported influence from different theoretical orientations was assessed on a scale ranging from 0 (not at all) to 6 (very much).

Supplement 2). Only 5 therapists (10.4%) were ineligible to treat patients in the match condition (ie, they were not classified as effective on any domains and were ineffective in at least 1 domain; eAppendix 2 in Supplement 2 shows analyses investigating the effect of this subgroup on the results, which was negligible).

Patient Flow and Descriptive Information

As per Figure 1, 288 patients were randomized after providing consent during the naturalistic intake process. Despite the minimal ask of eligible patients at baseline, of those who declined to participate, most simply wanted to be assigned to a therapist without having to leave the intake call to go online to consent and complete the TOP (moreover, without knowing the nature of the assignment methods being tested, patients had limited incentive to participate outside of financial compensation). Importantly, though, the study sample was closely aligned with the participating MHC network's average utilization data with regard to demographic characteristics, thereby increasing confidence in its representativeness. After allocation, 70 patients attrited for a final mITT sample of 218 for our primary analyses. This effective sample did not differ from attrited patients on clinical and demographic vari-

ables (eAppendix 3 in Supplement 2). **Table 2** presents baseline patient characteristics for the mITT sample. Patients in the CAU (n = 119) and matched (n = 99) conditions did not differ on any demographic or clinical variables (eAppendix 3 in Supplement 2). Among 218 patients, 147 (67.4%) were female, and 193 (88.5%) were White. The mean (SD) age was 33.9 (11.2) years. Patients provided a median (SD) of 12.14 (6.10) weeks of data, and 83 (38.1%) provided data through the full 16 treatment weeks.

Consistent with the intended manipulation, compared with patients in the CAU group and to a large degree, significantly more patients in the matched group were matched at higher levels (χ^2_4 = 49.47; P < .001; Cramer V = 0.48). Specifically, 43 patients in the CAU group (36.1%) were assigned to a therapist who was not a match at any level and was ineffective in at least 1 problem domain. Of the remaining 76 patients in the CAU group who received a chance match, 55 (72.4%) were matched at level 5, 2 (2.6%) at level 4, none at level 3, 17 (22.4%) at level 2, and 2 (2.6%) at level 1. In the match condition, 58 patients (58.6%) were matched at level 5, 4 (4.0%) at level 4, none at level 3, 28 (28.3%) at level 2, and 9 (9.1%) at level 1 (eFigure in Supplement 2). Finally, we tested the effects of different match levels on general impairment (eAppendix 4 in Supplement 2); results indicated that each level of matching outperformed no match, with higher levels showing the largest effects: d = 1.25 when patients were matched with therapists effective in treating the patients' top 3 elevated domains, d = 1.00 when patients were matched with therapists effective at treating the patients' single-most elevated domains, and d = 0.75 when patients were matched with therapists not ineffective in any domain.

All outcomes were normally distributed except domain-specific impairment, which was positively skewed (skewness = 2.12). Given that this variable included both negative and positive numbers, we added a constant before log transforming it (skewness = −0.18). Also, patients for whom substance misuse, suicidal ideation, or violence was the most elevated domain had more extreme values than patients for whom the most elevated domain was 1 of the other 9. Therefore, in the domain-specific impairment model, we included as covariates dummy-coded indicators of when a participant's most elevated problem was substance misuse, suicidal ideation, or violence. For context, there were no condition differences in the mean (SD) number of sessions attended (match: 5.80 [3.54]; CAU: 5.61 [3.02]; t_{216} = −0.42; SE = 0.44; P = .68) or in the mean (SD) number of weeks in the study (match: 11.09 [5.99]; CAU: 11.81 [6.20]); t_{216} = 0.61; SE = 0.83; P = .39).

Condition Effects on Primary Outcomes

Table 3 shows all fixed effects, and eTable 3 in Supplement 2 shows all random effects and model fit information. To a moderate to large degree, patients in the matched group experienced greater weekly reductions in their general impairment compared with patients in the CAU group (γ_{110} = −0.03; 95% CI, −0.05 to −0.01; P = .02; patient-level d = 0.75). **Figure 2** shows that this effect was clinically significant with the matched patients, on average, ending treatment in the nonclinical range (ie, average week 17 model-estimated

Table 2. Patient Characteristics at Baseline by Treatment Condition

Characteristic	No. (%)	
	CAU (n = 119)	Match (n = 99)
Age, mean (SD), y	34.42 (11.55)	33.33 (10.72)
Sex		
Female	81 (68.1)	66 (66.7)
Male	38 (31.9)	33 (33.3)
Race/ethnicity[a]		
White	106 (89.1)	87 (87.9)
Hispanic/Latino	3 (2.5)	3 (3.0)
African American/Black	6 (5.0)	7 (7.1)
Asian	2 (1.7)	1 (1.0)
East Indian	0	1 (1.0)
Other	2 (1.7)	0
Sexual orientation[b]		
Heterosexual	96 (80.7)	89 (89.9)
Bisexual	11 (9.2)	6 (6.1)
Gay or lesbian	4 (3.4)	3 (3.0)
Not sure	5 (4.2)	0
Missing	3 (2.5)	1 (1.0)
Annual household income, $		
<20 000	7 (5.9)	7 (7.1)
20 000-40 000	11 (9.2)	10 (10.1)
40 000-75 000	39 (32.8)	28 (28.3)
75 000-100 000	21 (17.6)	24 (24.2)
100 000-200 000	26 (21.9)	21 (21.2)
≥200 000	12 (10.1)	8 (8.1)
Missing	3 (2.5)	1 (1.0)
Education		
≤High school	14 (11.8)	18 (18.2)
Business or trade school	6 (5.0)	8 (8.1)
Associate's degree	11 (9.2)	13 (13.1)
Bachelor's degree	43 (36.1)	30 (30.3)
Master's degree or doctorate	34 (28.6)	22 (22.2)
Missing	11 (9.3)	8 (8.1)
Marital status		
Single	56 (47.1)	45 (45.5)
Married/cohabiting	53 (44.5)	44 (44.4)
Divorced/widowed/separated	7 (5.9)	9 (9.1)
Missing	3 (2.5)	1 (1.0)
Employment status		
Employed full-time	78 (65.6)	70 (70.7)
Employed part-time	13 (10.9)	14 (14.1)
Retired/unemployed but not looking for work/working but not for money	5 (4.2)	1 (1.0)
Full-time student	9 (7.6)	7 (7.1)
Unemployed, looking for work	8 (6.7)	6 (6.1)
Missing	6 (5.0)	1 (1.0)
Religious identification		
Christian	58 (48.7)	55 (55.6)
Jewish	7 (5.9)	3 (3.0)
Other (eg, Hindu, Muslim, Buddhist)	8 (6.7)	6 (6.1)
No religion	42 (35.3)	33 (33.3)
Missing	4 (3.4)	2 (2.0)

(continued)

Table 2. Patient Characteristics at Baseline by Treatment Condition (continued)

Characteristic	No. (%)	
	CAU (n = 119)	Match (n = 99)
Serious medical illness, mean (SD)[c]	5.34 (1.34)	5.54 (1.06)
Previous mental health hospitalization		
Yes	11 (9.3)	10 (10.1)
No	105 (88.2)	88 (88.9)
Missing	3 (2.5)	1 (1.0)
Previous therapists/courses of therapy, mean (SD)[d]	1.76 (1.95)	1.55 (1.50)
Currently using psychiatric medication		
Yes	37 (31.1)	26 (26.3)
No	57 (47.9)	52 (52.5)
Missing[e]	25 (21.0)	21 (21.2)
Primary problem/concern		
Quality of life	25 (21.0)	21 (21.2)
Depression	22 (18.5)	20 (20.2)
Substance misuse	20 (16.8)	18 (18.2)
Panic/somatic anxiety	15 (12.6)	8 (8.1)
Social functioning	7 (5.9)	8 (8.1)
Suicidal ideation	9 (7.6)	4 (4.0)
Sleep	7 (5.9)	5 (5.1)
Sexual functioning	4 (3.4)	7 (7.1)
Psychosis	6 (5.0)	2 (2.0)
Violence	2 (1.7)	2 (2.0)
Work functioning	1 (0.8)	3 (3.0)
Mania	1 (0.8)	1 (1.0)
Outcomes, mean (SD)		
TOP general impairment (z score)	0.86 (0.78)	1.04 (0.96)
SCL-10	15.48 (7.93)	16.08 (7.74)
TOP domain-specific impairment (z score)	3.97 (2.85)	3.98 (2.54)

Abbreviations: CAU, case assignment as usual; TOP, Treatment Outcome Package.

[a] Patients self-reported their race/ethnicity by checking 1 or more of the following predefined categories: White, Hispanic/Latino, African American/Black, Native American/American Indian, Asian, East Indian, and other. Reporting patient race/ethnicity data allowed us to provide information about the generalizability of the results in accordance with the Patient-Centered Outcomes Research Institute's Methodology Standards.

[b] Patients self-reported their sexual orientation by checking 1 or more of the following predefined categories: heterosexual, bisexual, gay or lesbian, or not sure. Reporting patient sexual orientation data allowed us to provide information about the generalizability of the results in accordance with the Patient-Centered Outcomes Research Institute's Methodology Standards.

[c] Serious medical illness was rated by patients on a scale from 1 (all) to 6 (none), and data for this variable were missing for 1 patient in the CAU group and 2 patients in the matched group.

[d] Data for this variable were missing for 4 patients in the CAU group and 1 patient in the matched group.

[e] The total sample size for the psychiatric medication item is 172 because of a technological error during data collection. Although all 218 patients completed the TOP–Case Mix form (which contained the item about psychiatric medications) at baseline, the responses to this item were not saved in the electronic database for patients who completed the item between October 25, 2018, and May 31, 2019. This was the only item for which this issue occurred.

z score = −0.03, or below the mean impairment level shown by individuals in non–treatment-seeking samples), and patients in the CAU group, on average, ending treatment in the clinically impaired range (ie, average week 17 model-estimated z score = 0.35). To a moderate degree, patients in the matched group experienced greater weekly reductions in psychological distress compared with patients in the CAU group (γ_{110} = −0.16; 95% CI, −0.30 to −0.02; P = .03; patient-level d = 0.50). Finally, to a moderate degree, patients in the matched group experienced greater weekly reductions in domain-specific impairment compared with patients in the CAU group (γ_{110} = −0.01; 95% CI, −0.01 to −0.006; P = .01; patient-level d = 0.60), controlling for which domain was elevated. No adverse events were reported.

To ensure that our results were not unduly influenced by our preregistered analytic plan for using the mITT sample, we conducted a supplemental analysis for the general impairment outcome. For this model, we included 48 of 51 patients who were lost to follow-up (3 of these patients did not complete the TOP at baseline and therefore had no outcomes data to include) for a sample size of 266. (For ethical reasons, we did not include the 4 patients who withdrew consent to analyze their data; we also could not include the 15 patients who never began treatment or received a therapist assignment, as these patients had no therapist identification variable.) For this sample of 266 patients, the match effect on general impairment reduction remained the same (γ_{110} = −0.03; 95% CI, −0.05 to −0.01; P = .01).

Table 3. Match Effect on General Impairment, Psychological Distress, and Domain-Specific Impairment Among 218 Mental Health Care Patients[a]

Fixed effects	TOP general impairment			SCL-10			TOP domain-specific impairment[b]		
	Coefficient (SE; 95% CI)	P value	ES[c]	Coefficient (SE; 95% CI)	P value	ES[c]	Coefficient (SE; 95% CI)	P value	ES[c]
Outcome at baseline (intercept), γ_{000}	0.83 (0.07; 0.69 to 0.97)	<.001	NA	15.33 (0.69; 13.98 to 16.68)	<.001	NA	0.33 (0.02; 0.29 to 0.37)	<.001	NA
Matched vs CAU, γ_{010}	0.11 (0.14; −0.16 to 0.38)	.44	0.13	0.28 (1.03; −1.74 to 2.30)	.79	0.04	0.02 (0.02; −0.02 to 0.06)	.37	0.13
Weekly outcome change (slope), γ_{100}	−0.03 (0.005; −0.04 to −0.02)	<.001	NA	−0.32 (0.05; −0.42 to −0.22)	<.001	NA	−0.01 (0.002; −0.01 to −0.006)	<.001	NA
Matched vs CAU, γ_{110}	−0.03 (0.01; −0.05 to −0.01)	.02	0.75	−0.16 (0.07; −0.30 to −0.02)	.03	0.50	−0.01 (0.002; −0.01 to −0.006)	.01	0.60

Abbreviations: CAU, case assignment as usual; ES, effect size; NA, not applicable; SCL-10, Symptom Checklist-10; TOP, Treatment Outcome Package.

[a] This table presents the results of 3 separate multilevel models (1 for each outcome variable) in the 3 main columns. The primary fixed effects represent the effect of condition (CAU = 0; matched = 1) on the intercept (baseline outcome level) and slope (weekly outcome change). See eTable 3 in Supplement 2 for the random effects and model fit information.

[b] The domain-specific outcome variable was log-transformed to correct a positive skew. This model included dummy-coded covariates that indicated

when participants reported substance misuse, suicidal ideation, or violence as their most elevated domain, because participants who endorsed 1 of these domains also tended to have more extreme values compared with those who endorsed the other domains.

[c] Effect sizes represent multilevel approximations of Cohen d; that is, match effects represent the number of patient-level standard deviations by which the groups differed at baseline (intercept) and in their weekly rates of change (slope).

Figure 2. Model-Estimated Match Effects on Change in Severity of General Impairment During Treatment Among 218 Mental Health Care Patients
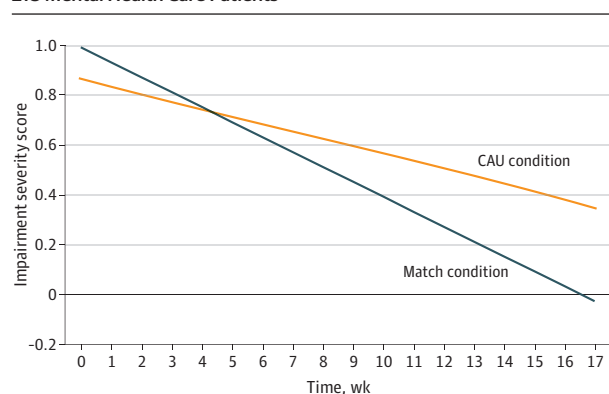
Patients in the matched group experienced significantly greater reductions in general impairment compared with patients in the case assignment as usual (CAU) condition across 17 weeks (16 weeks of treatment and 1 baseline week). In terms of overall impairment severity, a score of 0 represents the mean level of impairment shown by non–treatment-seeking individuals in the community. Therefore, a score of 1 represents an impairment level that is 1 SD higher than the community mean and a score of −1 represents impairment that is 1 SD lower than the community mean.

Discussion

In this 2-arm, double-blind randomized clinical trial, we tested the effectiveness of a personalized match system based on therapists' effectiveness report cards vs CAU prior to naturalistic MHC. As predicted, matched patients demonstrated greater reductions in 3 outcomes compared with patients in the CAU group. With most therapists treating patients in both conditions, our multilevel models tested a within-therapist effect of patient-specific matching,

while accounting for general between-therapist differences in effectiveness across all patients in the therapists' caseloads. That is, relative to their own average outcome, a given therapist achieved better outcomes when treating a matched patient vs a CAU patient. Thus, measurement-based matching represents a readily scalable innovation that can complement other evidence-based efforts to improve MHC.

These results suggest that measurement-based therapist report cards can help redirect MHC toward therapists' strengths.[11] Analogous to specialization movements in medicine,[27,28] such precision care in mental health can have a substantive positive effect. Importantly, dispelling the possibility that the match effect was solely a function of shared method variance, TOP-based matching also had a positive effect on SCL-10-based distress.

Notably, the good fit in this study came not from changing what the therapists did in their treatment, but rather who they treated. Capitalizing on whatever it is that a therapist historically does well when treating patients with certain mental health problems, the current data indicate that our match system can improve the effectiveness of that care, even with neither therapist nor patient being aware of their match status. To be used in this masked manner, the match system only requires that a multidimensional outcome tool be administered at baseline and follow-up—a practice that is becoming more common[29]—so that therapist performance data can prime matched case assignment without threatening clinician treatment autonomy. That said, it is plausible (although it requires testing) that there might be an even bigger effect were patients and therapists to be aware of their match status, perhaps through additive mechanisms, such as increased outcome expectation.[30] Whether masked or aware, patients have indicated that a MHC system's strategic use of therapist performance information would be a highly valued action.[16,17]

Limitations

This study had limitations. First, the naturalistic design resulted in limited information on treatment, and future research should examine what it is that well-performing therapists do when treating patients in a given strength domain. Second, the setting limited our control over a balanced design in which each therapist saw an equal number of patients in the matched group and in the CAU group. Therefore, we cannot fully rule out that between-therapist differences in effectiveness could have influenced the results, although both our crossed design and supplemental analyses suggested that this was unlikely. Third, with only 22% of eligible patients participating, it is possible this subgroup differed from the general population of treatment-seeking adults, which could reduce generalizability. However, it is plausible that participation was restricted simply because of the limited incentive to participate in research. Thus, future implementation work should examine ways in which multidimensional assessment can be most effectively incorporated into a system's modus operandi. Fourth, although all clinics in this trial were part of a single health care system, and clinicians may have worked fluidly between them in some instances, it is still possible that there were clinic-level effects on matching. Such additional nuance should be the focus of future work. Fifth, generalizability beyond a predominantly White, predominantly heterosexual sample within a specific MHC system may be limited. Sixth, although matching can theoretically be facilitated by any multidimensional outcome tool, this study only tested 1. Seventh, we report here only the main effects of 1 type of matching; future research should focus on potential moderators of (eg, patient motivation) or additions to (eg, also matching on cultural identities) the match system and other match or precision innovations in general.

Conclusions

This trial established a minimalist method of precision MHC. Namely, MHC can be substantially improved by using therapist performance data to determine who they treat. This method provides stakeholders (ie, patients, therapists, and administrators) a choice (as a shortlist) for optimizing care beyond chance levels, while also minimizing ineffectiveness.

REFERENCES

1. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annu Rev Public Health*. 2008;29:115-129. doi:10.1146/annurev.publhealth.29.020907.090847

2. Institute of Medicine. *Improving the Quality of Health Care for Mental and Substance-Use Conditions*. National Academies Press; 2006.

3. World Health Organization. *ICD-11: International Statistical Classification of Diseases and Related Health Problems*. 11th ed. World Health Organization; 2019.

4. Thorpe KE, Florence CS, Joski P. Which medical conditions account for the rise in health care spending? *Health Aff (Millwood)*. 2004;23(suppl 1):W4-437-445. doi:10.1377/hlthaff.W4.437

5. Kraus DR, Castonguay L, Boswell JF, Nordberg SS, Hayes JA. Therapist effectiveness: implications for accountability and patient care. *Psychother Res*. 2011;21(3):267-276. doi:10.1080/10503307.2011.563249

6. Lambert MJ. The efficacy and effectiveness of psychotherapy. In: Lambert MJ, ed. *Bergin & Garfield's Handbook of Psychotherapy and Behavior Change*. 6th ed. Wiley; 2013:169-218.

7. Johns RG, Barkham M, Kellett S, Saxon D. A systematic review of therapist effects: a critical narrative update and refinement to review. *Clin Psychol Rev*. 2019;67:78-93. doi:10.1016/j.cpr.2018.08.004

8. Firth N, Barkham M, Kellett S, Saxon D. Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: a multilevel modelling analysis. *Behav Res Ther*. 2015;69:54-62. doi:10.1016/j.brat.2015.04.001

9. Imel ZE, Sheng E, Baldwin SA, Atkins DC. Removing very low-performing therapists: a simulation of performance-based retention in psychotherapy. *Psychotherapy (Chic)*. 2015;52(3):329-336. doi:10.1037/pst0000023

10. Saxon D, Barkham M. Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *J Consult Clin Psychol*. 2012;80(4):535-546. doi:10.1037/a0028898

11. Muir HJ, Coyne AE, Morrison NR, Boswell JF, Constantino MJ. Ethical implications of routine outcomes monitoring for patients, psychotherapists, and mental health care systems. *Psychotherapy (Chic)*. 2019;56(4):459-469. doi:10.1037/pst0000246

12. Kraus DR, Bentley JH, Alexander PC, et al. Predicting therapist effectiveness from their own practice-based evidence. *J Consult Clin Psychol*. 2016;84(6):473-483. doi:10.1037/ccp0000083

13. Boswell JF, Constantino MJ, Kraus DR, Bugatti M, Oswald JM. The expanding relevance of routinely collected outcome data for mental health care decision making. *Adm Policy Ment Health*. 2016;43(4):482-491. doi:10.1007/s10488-015-0649-6

14. Coyne AE, Constantino MJ, Boswell JF, Romano FM, Kraus DR. Accuracy of therapist self-perceived effectiveness as a determinant of between-therapist effects. Paper presented at: Annual meeting of the Society for Psychotherapy Research; July 4, 2019; Buenos Aires, Argentina.

15. Walfish S, McAlister B, O'Donnell P, Lambert MJ. An investigation of self-assessment bias in mental health providers. *Psychol Rep*. 2012;110(2):639-644. doi:10.2466/02.07.17.PR0.110.2.639-644

16. Boswell JF, Constantino MJ, Oswald JM, et al. A multi-method study of mental health care patients' attitudes toward clinician-level performance information. *Psychiatr Serv*. 2021;72 (4):452-456. doi:10.1176/appi.ps.202000366

17. Boswell JF, Constantino MJ, Oswald JM, Bugatti M, Goodwin B, Yucel R. Mental health care consumers' relative valuing of clinician performance information. *J Consult Clin Psychol*. 2018;86(4): 301-308. doi:10.1037/ccp0000264

18. Kraus DR, Seligman DA, Jordan JR. Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: the Treatment Outcome Package. *J Clin Psychol*. 2005;61(3):285-314. doi:10.1002/jclp.20084

19. Wampold BE, Brown GS. Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. *J Consult Clin Psychol*. 2005;73(5):914-923. doi:10.1037/0022-006X.73.5.914

20. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59(1):12-19. doi:10.1037/0022-006X.59.1.12

21. Rosen CS, Drescher KD, Moos RH, Finney JW, Murphy RT, Gusman F. Six- and ten-item indexes of psychological distress based on the Symptom Checklist-90. *Assessment*. 2000;7(2):103-111. doi:10.1177/107319110000700201

22. Raudenbush SW, Liu X. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychol Methods*. 2001;6(4): 387-401. doi:10.1037/1082-989X.6.4.387

23. Raudenbush SW, Spybrook J, Bloom H, Congdon R, Hill C, Martínez A. Optimal Design software for multi-level and longitudinal research. Version 3.01. Accessed April 5, 2015. http://wtgrantfoundation.org/resource/optimal-design-with-empirical-information-od

24. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Sage; 2002.

25. Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, Raudenbush S. Optimal Design plus empirical evidence: documentation for the "Optimal Design" software. Accessed April 1, 2020. https://sites.google.com/site/optimaldesignsoftware/home

26. Allison PD. Missing data. In: Marsden PV, Wright JD, eds. *Handbook of Survey Research*. 2nd ed. Emerald Group Publishing; 2010:631-657.

27. Tu JV, Donovan LR, Lee DS, et al. Effectiveness of public report cards for improving the quality of cardiac care: the EFFECT study: a randomized trial. *JAMA*. 2009;302(21):2330-2337. doi:10.1001/jama.2009.1731

28. Farley DO, Elliott MN, Short PF, Damiano P, Kanouse DE, Hays RD. Effect of CAHPS performance information on health plan choices by Iowa Medicaid beneficiaries. *Med Care Res Rev*. 2002;59(3):319-336. doi:10.1177/107755870205900305

29. Rousmaniere T, Wright CV, Boswell J, et al. Keeping psychologists in the driver's seat: four perspectives on quality improvement and clinical data registries. *Psychotherapy (Chic)*. 2020;57(4): 562-573. doi:10.1037/pst0000227

30. Constantino MJ, Vîslă A, Coyne AE, Boswell JF. A meta-analysis of the association between patients' early treatment outcome expectation and their posttreatment outcomes. *Psychotherapy (Chic)*. 2018;55(4):473-485. doi:10.1037/pst0000169