Supplemental Online Content

Constantino MJ, Boswell JF, Coyne AE, Swales TP, Kraus DR. Effect of matching therapists to patients vs assignment as usual on adult psychotherapy outcomes: a randomized clinical trial. JAMA Psychiatry. Published online June 9, 2021. doi:10.1001/jamapsychiatry.2021.1221

eAppendix 1. General match effect equation
eTable 1. Therapist baseline strengths and weaknesses by domain (n = 48)
eTable 2. Therapists' baseline report cards (n = 48)
eAppendix 2. Examining the impact of between-therapist differences in global effectiveness on the study results
eAppendix 3. Baseline sample comparisons: full results
eFigure. Match level by condition (n = 218)
eAppendix 4. Correlational match level effect findings
eTable 3. Random effects and model fit information for the primary match effect models (n = 218)

This supplemental material has been provided by the authors to give readers additional information about their work.

eAppendix 1. General Match Effect Equation

Level-1 Model

$$\text{Outcome}_{ijk} = \pi_{0jk} + \pi_{1jk}*(\text{weeks}_{ijk}) + e_{ijk}$$

Level-2 Model

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}*(\text{match}_{jk}) + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k} + \beta_{11k}*(\text{match}_{jk}) + r_{1jk}$$

Level-3 Model

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

$$\beta_{01k} = \gamma_{010} + u_{01k}$$

$$\beta_{10k} = \gamma_{100} + u_{10k}$$

$$\beta_{11k} = \gamma_{110} + u_{11k}$$

At level 1, the relevant outcome variable (general impairment, global distress, domain-specific impairment) at time *i* for patient *j* treated by therapist *k* is predicted by each patient's estimated baseline (week 0) outcome level (π_{0jk}) and each patient's weekly rate of outcome change (π_{1jk}). At level 2, the patient-specific intercept and slope coefficients (π_{0jk}, π_{1jk}) drop down to become the outcome variables. Each patient's (*j*) baseline outcome level is predicted by the baseline outcome level of their therapist's (*k*) average CAU patient (β_{00k}) and treatment condition (β_{01k}; match = 1; CAU = 0). Similarly, each patient's (*j*) weekly change rate is predicted by the change rate of their therapist's (*k*) average CAU patient (β_{10k}) and treatment condition (β_{11k}). At level 3, each of these coefficients drop down to become the outcome variables, which are predicted by the sample averages (fixed effects); that is, γ_{000} represents the baseline outcome level for the average CAU patient, γ_{010} represents the average difference in baseline outcome level between match vs. CAU patients, γ_{100} represents the change rate for an average CAU patient, and γ_{110} represents the average difference in weekly outcome change between match vs. CAU patients.

At each level, random effects (e_{ijk}, r_{0jk}, r_{1jk}, u_{00k}, u_{01k}, u_{10k}, u_{11k}) allow for variability around the model estimates. At level 1, variability represents within-patient deviations from their own change trajectory plus measurement error (e_{ijk}). At level 2, random effects allow for between-patient variability around each therapist's average baseline outcome level and weekly change rate (i.e., r_{0jk} and r_{1jk}, respectively). At level 3, the random effects represent differences between therapists in their average baseline outcome levels and change rates with CAU patients (u_{00k} and u_{10k}, respectively) and differences between therapists in the effect of matching on their patients' baseline outcome levels and change rates (u_{01k} and u_{11k}, respectively). Given that our primary goal was to explain *between-patient* differences in outcome, level-2 random effects for both the intercept and slope were included across all models. As noted, we also included a therapist-level random intercept (u_{00k}) across all outcomes, but we included the other possible level-3 random effects only if they were significant and/or improved the fit of the model. This approach preserved power and parsimony.

Effect sizes were calculated according to information and formulas presented in the manual for the Optimal Design Program (Spybrook et al., 2011). More specifically, we used the relevant level 2 variance components for the intercept (r_{0jk}) and slope (r_{1jk}) from an unconditional linear model (with no predictors), to capture the total patient-level variability in our outcomes (i.e., weekly during-treatment outcome change, and posttreatment outcome level). Therefore, the effect sizes represent the number of standard deviations of patient-level (level 2) variability in

the intercept and slope by which the two groups are expected to differ (i.e., a multilevel approximation of Cohen *d*).

eTable 1. Therapist Baseline Strengths and Weaknesses by Domain (n = 48)

Domains	Effective No. (%)	Neutral No. (%)	Ineffective No. (%)
Depression	7 (15)	34 (71)	7 (15)
Panic/somatic anxiety	5 (10)	38 (79)	5 (10)
Mania	6 (13)	40 (83)	2 (4)
Substance misuse	3 (6)	44 (92)	1 (2)
Psychosis	4 (8)	40 (83)	4 (8)
Suicidality	10 (21)	35 (73)	3 (6)
Violence	14 (29)	33 (69)	1 (2)
Sexual functioning	8 (17)	34 (71)	6 (13)
Social functioning	1 (2)	45 (94)	2 (4)
Sleep	8 (17)	38 (79)	2 (4)
Work functioning	5 (10)	37 (77)	6 (13)
Quality of life	4 (8)	37 (77)	7 (15)

eTable 2. Therapists' Baseline Report Cards (n = 48)

Effective		Ineffective	
Number of Domains	% of Therapists	Number of Domains	% of Therapists
0	35	0	56
1	23	1	21
2	17	2	15
3	8	3	2
4	13	4	0
5	0	5	2
6	4	6	2
7	0	7	0
8	0	8	2
9	0	9	0
10	0	10	0
11	0	11	0
12	0	12	0

eAppendix 2. Examining the Impact of Between-Therapist Differences in Global Effectiveness on the Study Results

Despite the fact that the majority of therapists saw patients in both conditions, the naturalistic setting made it impossible for each therapist to treat a *perfectly* balanced number of cases in the match and CAU groups. Therefore, it remains possible that between-therapist differences in global effectiveness *could* have impacted the results. To investigate this possibility, we conducted several follow-up statistical and descriptive analyses. Perhaps most importantly, as described in the main text of this study, there were 5 study therapists who could not fulfill any level of the match (ie, they were not classified as effective on any TOP domains and were ineffective on at least one domain). Therefore, to test the possibility that the beneficial match effect could be better accounted for by the fact that these therapists (n = 5) could only treat CAU patients, we replicated our analysis for the general impairment severity outcome after removing these therapists and the 15 study patients they treated. In this subsample of 203 patients treated by the remaining 43 therapists, the beneficial match effect on during-treatment reduction in general impairment remained statistically significant and had the exact same effect size (γ_{010} = -0.03, SE = 0.01, P = .04; 95% CI -0.05, -0.01). Therefore, it is unlikely that the exclusion of a small number of therapists from the match condition accounted for the observed condition effect.

We also descriptively examined whether there was any evidence that more globally effective therapists treated more patients in the match vs. CAU conditions. Importantly, although not all therapists treated patients in both conditions (eg, therapists who saw only one study patient due to logistical factors), there was no evidence that therapists who were more globally effective with their trial patients were more likely to treat patients in one condition vs. the other; that is, when we selected the top 1/3 of therapists (n = 16) whose trial patients had the best posttreatment outcomes after adjusting for baseline severity, a descriptive examination revealed that these providers treated an equal number of patients in each of the two conditions (ns = 42 for both match and CAU).

Finally, although our multilevel models examined the condition effect on within-therapist differences in the study outcomes at level 2, while accounting for global between-therapist differences in effectiveness across all patients in their study caseloads at level 3, this method of statistical control does not work perfectly for therapists who only treated one study patient (and therefore had no within caseload variability). Therefore, we also replicated our general impairment severity analysis after removing the 6 therapists who each treated only one trial patient. In this subsample of 212 patients treated by 42 therapists, the beneficial match effect on during-treatment reduction in general impairment again remained statistically significant and had the exact same effect size (γ_{010} = -0.03, SE = 0.01, P = .03; 95% CI -0.05, -0.01).

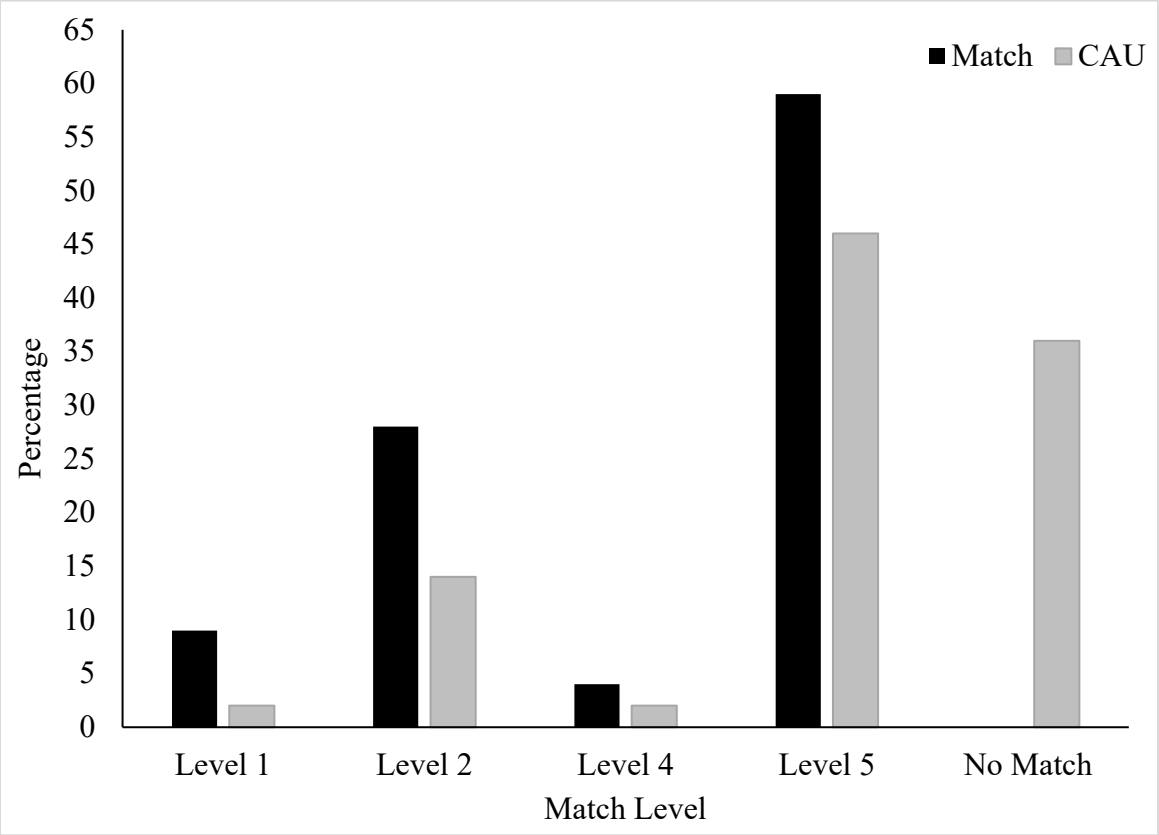
eAppendix 3. Baseline Sample Comparisons: Full Results

The final mITT sample ($N = 218$) did not differ from the patients who actively withdrew, passively withdrew, or were lost to follow-up ($N = 70$) on general impairment severity ($t[271] = -1.76$, $P = .08$; $d = .26$), gender ($\chi^2[1] = 0.004$, $P = .95$; phi = 0.02), age ($t[270] = -1.13$, $P = .26$; $d = .17$), or racial/ethnic minority (REM) status ($\chi^2[1] = 0.002$, $P = .97$; phi = 0.02). Additionally, all standardized group-mean differences were small (all ds < .30; all phis < .20). Importantly, the number of patients who withdrew or were lost to follow-up did not differ between the match ($n = 32$) and CAU ($n = 38$) conditions ($\chi^2[1] = 0.002$, $P = .97$; phi = 0.06).

Next, focusing on the final mITT sample, we compared the match ($n = 99$) and CAU ($n = 119$) conditions at baseline. There were no significant differences on the following demographics: gender ($\chi^2[1] = 0.05$, $p = .83$; phi = .02), age ($t[216] = 0.75$, $P = .48$; $d = .10$), REM status ($\chi^2[1] = 0.08$, $P = .78$; phi = 0.02), sexual orientation ($\chi^2[2] = 1.07$, $P = .59$; Cramer V = 0.07), marital status ($\chi^2[2] = 0.78$, $P = .68$; Cramer V = 0.06), family/household income ($\chi^2[8] = 3.22$, $P = .92$; Cramer V = 0.12), highest educational degree obtained ($\chi^2[4] = 4.42$, $P = .35$; Cramer V = 0.15), employment status ($\chi^2[4] = 2.62$, $P = .62$; Cramer V = 0.11), or religious identification ($\chi^2[3] = 1.53$, $P = .68$; Cramer V = 0.09). In terms of baseline clinical characteristics, there were also no differences between the conditions on psychiatric medication use ($\chi^2(1) = 0.67$, $P = .41$; phi = -0.06), presence of a serious medical illness ($t[212.71]^a = -1.20$, $P = .23$; $d = .17$), previous mental health hospitalization status ($\chi^2(1) = 0.03$, $P = .86$; phi = 0.01), number of previous therapists/courses of psychotherapy ($t[211] = 0.85$, $P = .40$; $d = .12$), general impairment ($t[216] = -1.55$, $P = .12$; $d = .21$), and global psychological distress ($t[216] = -0.56$, $P = .58$; $d = .08$). Additionally, all standardized group-mean differences were small (all ds < .30; all phis < .20; all Cramer V < .20).

[a] This analysis violated the Levene Test for Homogeneity of Variances, so we present the results using the corrected degrees of freedom with equal variances between the groups not assumed. Note that the non-parametric version of this analysis (a Mann-Whitney U) was also not significant ($P = .34$).

eFigure. Match Level by Condition (n = 218)



Note. A greater proportion of match vs. CAU patients were matched at level 1 (ie, a therapist who was effective on the patient's 3 most elevated domains and was not ineffective on any domain), level 2 (ie, a therapist who was effective on the patient's primary problem and was not ineffective on any domain), level 4 (ie, a therapist who was effective on the patient's primary problem, but who was ineffective on at least one other domain), and level 5 (ie, a generally not ineffective therapist who also possessed no strengths on any of the patient's elevated domains). In contrast, 36% (n = 43) of CAU patients (and, of course, 0% of match patients) were assigned to a therapist who was not a match at any level and who was ineffective on at least 1 problem domain.

eAppendix 4. Correlational Match Level Effect Findings

We used the same 3-level framework from our primary analyses to test whether match level (a level-2 variable) predicted outcome *across* both conditions. Given the possibility that differences between various levels of matching could be nonlinear, we dummy-coded the different levels, with non-matched patients (n = 43 or 36% of the CAU condition) as the reference category. Thus, we tested the extent to which each match level outperformed no match. As described in the main text of this manuscript, although there were actually 5 possible match levels, only 6 patients were matched at level 4 (ie, to a therapist who was effective at treating the patient's single most elevated TOP domain, but ineffective on at least one other domains) and no patients were matched at level 3 (ie, to a therapist who was effective in treating the patient's 3 most elevated TOP domains, but ineffective on at least one other domain). Thus, we removed match level 3, and we collapsed match levels 2 and 4 to create a single category that represents being matched with a therapist who is effective at treating a patient's most elevated TOP domain. Results indicated that across both conditions, patients who were matched with a therapist who was effective at treating their top 3 problem domains (γ_{110} = -0.05, SE = 0.02, P = .005; 95% CI -0.08, -0.01; d = 1.25), effective at treating their primary problem domain (γ_{120} = -0.04, SE = 0.01, P < .001; 95% CI -0.06, -0.02; d =1.00), and not ineffective on any domain (γ_{130} = -0.03, SE = 0.01, P = .005; 95% CI -0.05, -0.01; d = 0.75) had significantly steeper weekly reductions in general impairment than unmatched patients (γ_{100} = -0.01, SE = 0.01, P = .16; 95% CI -0.03, 0.01).

eTable 3. Random Effects and Model Fit Information for the Primary Match Effect Models (n = 218)

	TOP General Impairment Model			SCL-10 Model			TOP Domain-Specific Impairment[a]		
Random effects	Variance component	*P*	$\Delta\chi^2(df)$[b]	Variance component	*P*	$\Delta\chi^2(df)$[b]	Variance component	*P*	$\Delta\chi^2(df)$[b]
Level 1									
Residual, σ^2	0.14	--		15.56	--		0.02	--	
Level 2									
TOP intercept, $\tau_{\pi 00}$	0.50	< .001		43.95	< .001		0.02	< .001	
TOP slope, $\tau_{\pi 11}$	0.001	< .001		0.10	< .001		0.0001	< .001	
Level 3									
TOP intercept, $\tau_{\beta 0000}$	0.01	> .50		0.004	> .50		0.0001	.46	
Match effect intercept, $\tau_{\beta 0101}$	0.26	< .01		--	--		--	--	
TOP slope, $\tau_{\beta 1010}$	0.0001	> .50		--	--		--	--	
Match effect slope, $\tau_{\beta 1111}$	0.003	< .001		--	--		--	--	
Model deviance (df)	1583.42 (18)	--	20.18* (9)	5260.61 (9)	--	5.33 †(2)	-713.08 (15)	--	6.54*(2)

Note. Abbreviations: TOP, Treatment Outcome Package; SCL-10, Symptom Checklist-10, df, degrees of freedom. This table presents the random effects and model fit information from the three separate multilevel models (one for each outcome variable) in three main columns.

[a] As described in the main text of this manuscript, the domain-specific outcome variable was log-transformed to correct a positive skew. As also described, 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 one of these domains also tended to have more extreme values compared to the other domains.

[b] Significant χ^2 difference tests indicate that the match effect model is a significantly better fit to the data than the unconditional linear model, or that the match effect model is a significantly better fit to the data than the covariates only model (for the domain-specific impairment outcome).

* $P < .05$; † $P < .10$.