Article:
Coyne, A.E., Constantino, M.J., Gaines, A.N. et al. (Accepted: 2025) Are some therapists more effective when they deliver one type of therapy versus another? Psychotherapy.
ISSN: 0033-3204 (In Press)

This is the final author version of the article accepted for publication in *Psychotherapy*:

Coyne, A. E., Constantino, M. J., Gaines, A. N., Wienicke, F. J., Mehta, A. H. P., Barkham, M., Stiles, W. B., Hardy, G. E., de Mattos Souza, L. D., Jansen, K., da Silva, R. A., dos Santos, É. N., Dekker, J. J. M., Peen, J., Van, H. L., Cohen, Z. D., & Driessen, E. (in press). Are some therapists more effective when they deliver one type of therapy versus another? *Psychotherapy.*

There may be very minor formatting or typing corrections compared with the final published version

Are Some Therapists More Effective When They Deliver One Type of Therapy Versus Another?

Alice E. Coyne[1], Michael J. Constantino[2], Averi N. Gaines[2], Frederik J. Wienicke[3], Anuj H. P. Mehta[2], Michael Barkham[4], William B. Stiles[5], Gillian E. Hardy[4], Luciano Dias de Mattos Souza[6], Karen Jansen[6], Ricardo Azevedo da Silva[6], Érico Nobre dos Santos[6], Jack J. M. Dekker[7], Jaap Peen[7], Henricus L. Van[8], Zachary D. Cohen[9], & Ellen Driessen[3,10]

[1] Department of Psychology, American University, Washington, D.C., United States of America

[2] Department of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, MA, United States of America

[3] Department of Clinical Psychology, Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

[4] Clinical and Applied Psychology Unit, School of Psychology, University of Sheffield, Sheffield, United Kingdom

[5] Department of Psychology, Miami University, Oxford, Ohio, United States; Department of Psychology, Appalachian State University, Boone, North Carolina, United States & Faculty of Research and Doctoral Programs, Metanoia Institute, London, United Kingdom

[6] Graduate Program in Health and Behavior, Catholic University of Pelotas, Pelotas, RS, Brazil

[7] Department of Research, Arkin Mental Health Care, Amsterdam, The Netherlands

[8] NPI, Arkin Mental Health Care, Amsterdam, The Netherlands

[9] Department of Psychology, University of Arizona, Tucson, AZ, United States of America

[10] Depression Expertise Centre, Pro Persona Mental Health Care, Nijmegen, The Netherlands

Author Note

Alice E. Coyne https://orcid.org/0000-0002-5950-0486

Michael J. Constantino https://orcid.org/0000-0003-3126-2575

Averi N. Gaines https://orcid.org/0000-0001-5856-7059

Frederik J. Wienicke https://orcid.org/0000-0002-1728-1203

Anuj H. P. Mehta https://orcid.org/0000-0001-6886-0064

Michael Barkham https://orcid.org/0000-0003-1687-6376

William B. Stiles https://orcid.org/0000-0001-7740-0609

Gillian E. Hardy https://orcid.org/0000-0002-9637-815X

Luciano D. de M. Souza https://orcid.org/0000-0001-9965-4837

Karen Jansen https://orcid.org/0000-0003-3494-8070

Ricardo Azevedo da Silva https://orcid.org/0000-0002-9114-7037

Érico Nobre dos Santos https://orcid.org/0000-0002-6773-8208

Zachary D. Cohen https://orcid.org/0000-0002-4883-1028

Ellen Driessen https://orcid.org/0000-0002-9602-1548

Recruitment of individual participant data for this work was supported by a Fund for Psychoanalytic Research of the American Psychoanalytic Association. This study was further financed by a research grant from the Netherlands Organisation for Scientific Research (NWO): 016.Veni.195.215 6806 to ED. The funders had no role in the study design; collection, analysis, or interpretation of data; writing the manuscript; or the decision to submit the manuscript for publication. Correspondence should be addressed to Alice E. Coyne, American University, 4400 Massachusetts Ave NW, Washington, DC 20016. Email: coyne@american.edu.

Abstract

Adding nuance to the *between*-therapist effect on patient outcomes, research has increasingly demonstrated that a given therapist can differ in their effectiveness depending on *who* they treat (e.g., patients with different racial/ethnic identities) and/or *what* they treat (e.g., patients with different presenting problems). This preregistered study examined whether individual therapists are also more or less effective depending on *how* they treat their patients; that is, delivering one type of therapy versus another. We did so in the context of an individual participant data meta-analysis of clinical trials that compared classes of cognitive-behavioral therapy (CBT) and psychodynamic therapy (PDT) for depression. The meta-analytic sample included 30 therapists who were crossed with treatment condition and 492 patients (M = 25.08 patients per therapist; SD = 15.77). Patients completed measures of depression at baseline and posttreatment. Multilevel structural equation models revealed significant variability in the within-therapist treatment condition–outcome association ($p < .001$), indicating that some therapists were more effective when delivering one treatment over the other. Descriptively, 53% of therapists had similar outcomes across both groups (ds < .20), whereas 47% had at least a small-sized treatment-type strength (ds ≥ .20; range = 0.21–0.65). Results inform the personalization of treatment usage to the individual provider's effectiveness data.

Keywords: within-therapist effectiveness differences, cognitive-behavioral therapy, psychodynamic therapy, depression, individual participant data meta-analysis

Are Some Therapists More Effective When They Deliver One Type of Therapy Versus Another?

Much research indicates that therapists can differ from one another in their average effectiveness (e.g., Johns et al., 2019). In addition to this *between*-therapist effect, research has revealed that a given therapist's effectiveness can vary (relative to themselves) as a function of different patient factors and treatment contexts (i.e., *within*-therapist effects; Coyne, 2024). For example, regarding *who* they treat, a therapist's effectiveness can vary based on their patients' sociocultural characteristics, such as race/ethnicity (e.g., White vs. a Person of Color; Kivlighan et al., 2019) or sexual orientation (e.g., gay or lesbian vs. heterosexual; Drinane et al., 2022). Additionally, regarding *what* they treat, a therapist's effectiveness can vary based on their patients' presenting mental health problems (e.g., depression vs. quality-of-life deficits; Constantino et al., 2021). Such findings present opportunities to extend personalized mental health care to the personal strengths of the therapist (Delgadillo et al., 2020).

Extending this within-therapist effectiveness variability notion, it is also possible a given therapist may be more effective depending on *how* they treat their patients; that is, delivering one type of treatment versus another. Although treatment type has been the primary focus of therapy personalization research (Nye et al., 2023), such efforts have typically focused on fitting specific treatments to the *patient* (e.g., Cohen et al., 2021). Yet, it seems comparably plausible that therapists who elect to administer more than a single treatment approach can also have optimal fits. If therapists were to exhibit meaningful effectiveness differences when delivering different treatment types, then treatment-selection algorithms could be personalized to fit both patient *and* provider—an approach that would help mitigate concerns that the provider is an often neglected factor in mental health care personalization efforts (Cook et al., 2017). Furthermore, such therapist-level data could also inform a therapist's ongoing professional development. For example, therapists might choose to specialize in their treatment strength(s) or obtain further training in treatments they use less effectively, potentially allowing them to treat a wider array of patients in an evidence-informed manner (Coyne, 2024).

Despite the *potential* relevance of treatment-focused within-therapist effectiveness differences, this question has yet to be studied. This gap is understandable given that uncovering such differences requires access to large datasets in which therapists delivered more than one well-defined and distinct treatment. Although therapies in comparative trials are typically well-defined and distinct, the therapists are usually nested within a single condition (versus being crossed with them). Conversely, whereas therapists in naturalistic care may use more than one treatment (Stiles et al., 2008), such approaches are typically less well-defined and distinctive than they are in trials. Moreover, even for datasets that transcend these methodological barriers, small therapist samples (and small numbers of patients nested within treatment type) have limited statistical power and thereby confidence in any conclusions (e.g., Schiefele et al., 2017).

One way to address such challenges involves accessing big-enough data in which therapists treated patients in more than one standardized and distinct treatment in comparative clinical trials that are meta-analytically integrated to increase power. In this vein, the present study examined whether individual therapists were more or less effective when administering one treatment versus another in the context of an individual participant data meta-analysis (IPDMA) that aggregated raw patient-level data from five independent trials that compared classes of cognitive-behavioral therapy (CBT) and psychodynamic therapy (PDT) for depression (see Driessen et al. 2018).[1] Such a comparison is ideal in that it is well-established (including with prior meta-analyses) that CBT and PDT are statistically comparably efficacious for the treatment of depression (e.g., Cuijpers et al., 2020). Therefore, it is unlikely that average treatment effects would confound any therapist-specific differences in outcome by treatment. Further, comparing classes of CBT and PDT is also ideal because they are theoretically and procedurally dissimilar in many ways. Whereas therapists in CBT generally take a directive stance to focus on thought patterns and related behavioral problems, therapists in PDT generally take a less directive stance to explore emotional processes and relational dynamics.

Method

Dataset Overview

As noted, we focused on the subset of five trials from the Driessen et al. (2018) IPDMA (see the online supplement for details about the search and selection process) that specifically compared CBT and PDT *and* for which some therapists treated patients in both treatment arms.[2] This allowed us to determine if therapists had a strength in administering one or the other treatment, while taking advantage of the greater statistical power and lower group-level bias that IPDMAs afford over individual studies or conventional meta-analyses (e.g., Riley et al., 2021). Across these five trials, for which the key study features and references are presented in Table 1, we first removed all cases for which the therapist identification variable was missing. After this exclusion, our sample included 119 therapists who treated a total of 728 patients. Next, we excluded the therapists who did not treat patients in both therapy conditions, leaving 42

[1] The parent IPDMA included all treatments that fell under the broad umbrella of psychodynamically oriented psychotherapies, including distinct approaches like psychodynamic-interpersonal therapy. For simplicity, hereafter, we refer to all psychodynamically oriented approaches as "PDT."
[2] Note that because the Driessen et al. (2018) IPDMA focused on monotherapies, patients who received both psychotherapy and *medication* as part of a given trial were excluded.

therapists who saw a total of 524 patients. Next, to ensure reliable modeling of each therapist's average effectiveness when delivering each treatment, we excluded therapists who treated fewer than two patients in each of CBT and PDT. Based on this criterion, we eliminated an additional 12 therapists and their patients, which resulted in the effective samples presented next.

Participants

Effective sample patients were 492 adults who were randomly assigned to either CBT (*n* = 255) or PDT (*n* = 237). They averaged 38.31 years of age (*SD* = 11.08) and were predominantly female (65.45%) and married or cohabiting (62.80%). Supplemental Table 1 shows this demographic information by treatment condition, for which there were no significant differences (all *p*s > .05).[3] All patients presented with elevated depression. Table 1 shows the inclusion criteria and assessment methods for each included trial. For additional context, effective sample patients did not differ from excluded patients on gender or age (*p*s > .05). However, included versus excluded patients were more likely to be married or co-habiting ($\chi^2[1] = 88.07$, $p < .0001$) and had more severe baseline depression ($t[719] = -7.10$, $p < .001$).

Effective sample therapists were 30 certified mental health care professionals who treated an average of 25.08 (*SD* = 15.77) total patients (*M* = 13.07, *SD* = 8.73 in CBT; *M* = 12.02, *SD* = 7.40 in PDT). Individual therapists' caseloads were largely balanced, with no more than an ~70% to 30% split favoring either treatment. Within each of the included trials, therapists received a similar amount of training in each approach and received intensive supervision in each treatment from PDT and CBT experts, respectively (see Table 1 for additional study-level details on the therapists). Two of the five included studies reported high levels of observer-rated therapist adherence across both conditions. The other three studies inferred high levels of

[3] Note that these three demographic variables were the only ones uniformly assessed across the trials.

adherence based on intensive training, practice cases, and ongoing supervision (with review of session audio/video) but did not include formal observer ratings. Unfortunately, because of the inconsistency in collecting individual therapist-level characteristics across the five trials, we are unable to report any additional aggregated descriptive statistics on them.

Treatments

Each of the five trials included manualized and individually administered CBT and PDT (see Table 1 for additional individual study treatment details and Supplemental Table 2 for additional details on and references for the treatment manuals). CBT broadly focused on identifying, challenging, and replacing maladaptive beliefs using common cognitive and behavioral techniques (e.g., cognitive restructuring, activity scheduling). In contrast, PDT broadly focused on uncovering unconscious thoughts and feelings using common psychodynamic techniques (e.g., exploratory reflection, interpretation), as well as on uncovering past and present interpersonal patterns that related to depressive symptoms. Treatment lengths varied by study (range = 3 to 18 sessions; see Table 1).

Measures

To assess therapist effectiveness, we focused on patients' depression severity as assessed at baseline and posttreatment. The five trials either administered the self-rated Beck Depression Inventory I (BDI-I; Beck et al., 1961), the self-rated Beck Depression Inventory II (BDI-II; Beck et al. 1996), or the observer-rated Hamilton Depression Rating Scale (HDRS; Hamilton, 1960). Given the variability in depression measures across the trials, we standardized the raw total scores by converting them to within-study z-scores (i.e., scores were standardized relative to patients' original samples). Thus, our resulting depression severity index for the total sample is on a scale of standard deviation units, with higher scores representing more severity.

Procedure

We accessed de-identified data from the five aforementioned trials for the following variables: therapist identification, patient treatment condition, patient pretreatment depression, patient posttreatment depression, and patient demographic variables. The parent IPDMA was preregistered through PROSPERO (CRD42017056029) and the present study was preregistered through OSF (https://osf.io/eu4hg/?view_only=f56aeec4b9bc40418883da244a0839b2).

Data Analytic Plan

First, we calculated descriptive statistics for all study variables to determine whether we needed to conduct any sensitivity analyses due to the presence of significant outliers (i.e., scores that were > ± 3 *SD*s from the mean). Second, as a preliminary analysis, we examined the amount of variance in patients' posttreatment depression that was attributable to differences among therapists. Finally, for our primary analyses, we used the Mplus 8.1 program (Muthén & Muthén, 1998–2017) to fit a multilevel structural equation model (MSEM). This modeling approach accounted for patients being nested within therapists and automatically parsed pre- and posttreatment depression scores into their latent within- and between-therapist components. Additionally, because random effects (the key focus of this study) are not normally distributed (variances cannot be < 0), we used the Bayesian estimator within Mplus. This estimator does not assume normality and therefore allows for more accurate significance tests (Muthén & Asparouhov, 2012). Also, with this approach, missing data are handled using the Bayesian corollary of full information maximum likelihood estimation. Accordingly, we were able to retain all patients who completed a measure of depression severity on at least one of the baseline or posttreatment occasions; therefore, our models included 486 patients (99% of the sample).[4]

[4] Note that ~30% of patients (n = 147) were missing their posttreatment depression score and 1% of patients (n = 7) were missing their pretreatment depression score. For those patients who had depression data at the other timepoint

Statistical significance is established based on 95% credible intervals (CIs); those that do not include 0 denote significance. We also used non-informative priors, which allowed the model to be estimated based solely on the data.

Within our two-level model, between-patient differences were modeled at level 1, and between-therapist differences were modeled at level 2.[5] The outcome variable was patient depression at posttreatment. At level 1, we controlled for patient-level differences in baseline depression severity. At level 2, we controlled for therapist caseload-level differences in their patient's average level of baseline depression severity. To test our research question, we included treatment type (CBT = 0; PDT = 1) as a level-1 predictor of posttreatment outcome. We treated this within-therapist association as a random slope and tested its significance. To quantify the size of any significant therapist-level variability in the within-therapist treatment-outcome slope, we output the empirical Bayes (EB) slope estimates for each therapist. These EB slopes represented a therapist's average difference in posttreatment outcome (after controlling for within- and between-therapist differences in patients' baseline depression severity) when treating patients in one condition versus the other. Because the depression variables were *z*-scores, this average difference is represented in *SD* units that can be considered an approximation for Cohen's *d,* for which standard interpretations apply (0.2 denotes a small effect, 0.5 denotes a medium effect, 0.8 denotes a large effect).

Results

Descriptive statistics revealed two possible outliers with pre- or posttreatment depression scores that were > 3 *SD*s above the mean. Therefore, we replicated our primary model without

(either at pre- or posttreatment), the model used the Bayesian version of full information maximum likelihood estimation to retain them in the analyses. Only patients who were missing both timepoints ($n = 6$) were excluded.

[5] Because only five studies met our inclusion criteria, we chose not to include study-level differences as a third level of analysis due to concerns about reliability.

these patients' data; the pattern, size, and significance of all results remained consistent. Given the stability of the results and the fact that these patients' original depression scores were plausible values, we had no reason to believe the scores were errors. Therefore, we included their data in our primary analyses.

Results of the unconditional model revealed that 4.8% of the variance in patients' posttreatment depression was attributable to differences among therapists ($\tau_{00} = 0.04$, $p < .001$, 95% CI [0.02, 0.13]). Next, as shown in Table 2, our primary MSEM revealed that the average within-therapist treatment condition-outcome association was not significant ($\gamma_{10} = 0.08$, $p = .622$, 95% CI -0.22, 0.37), indicating that patients' average outcomes did not differ by treatment type. However, for our primary aim, the size of the treatment condition-outcome association did vary significantly among therapists ($\tau_{11} = 0.28$, $p < .001$, 95% CI [0.07, 0.86]).[6]

Descriptively, this between-therapist variability is depicted in Figure 1. The size of the posttreatment outcome difference between the two conditions ranged from -0.35 (favoring PDT) to +0.65 (favoring CBT). In terms of raw differences, 30% of therapists ($n = 9$) had outcomes that were virtually indistinguishable between the two conditions (*d*s < 0.10). Another 23% ($n = 7$) also had largely comparable outcomes across both treatments (*d*s < 0.20). However, 37% ($n = 11$) had small-to-moderately sized expected differences in their outcomes between the treatments (*d*s ranged from .21 to .46) and 10% ($n = 3$) had moderately sized outcome differences when delivering CBT versus PDT (*d*s ranged from 0.51 to 0.65). Among the 14 therapists who had at least a small-sized predicted difference in their outcomes between treatments, 79% ($n = 11$) had better outcomes in CBT versus PDT (average $d = 0.39$), whereas only 21% ($n = 3$) had better

[6] A sensitivity analysis revealed that this finding held when including a third level of analysis that accounted for study-level nesting ($\tau_{11} = 0.23$, $p < .001$, 95% CI [0.03, 0.77]). Further, a second sensitivity analysis revealed that this finding was not explained by the proportion of cases a given therapist treated with PDT versus CBT (-0.16, $p = .954$; 95% CI [-4.13, 4.20]).

outcomes in PDT versus CBT (average $d = 0.29$). Further, it is worth noting that within the entire sample of therapists, only 30% ($n = 9$) had better outcomes in PDT versus CBT.[7]

Discussion

This study demonstrated that some therapists can be significantly more effective when using CBT or PDT relative to the other, despite similar training in each. Still other therapists, though, demonstrated comparable outcomes across the treatments; if replicated, it would mean a sizable number of therapists who are trained in both CBT and PDT would be able to flexibly switch between them without compromising their usual effectiveness. Even so, a meaningful number of therapists in the present sample had a treatment-specific effectiveness edge (to a small-to-moderate degree), which could inform a novel aspect of treatment personalization *to the provider's strength*. Notably, such relatively modest effect sizes may be consistent with existing direct tests of treatment personalization *to the patient*. For example, a meta-analysis comparing personalized versus non-personalized treatment approaches found a small, but statistically significant effect ($d = 0.22$; Nye et al., 2023).

When considered alongside the broader literature on within-therapist effectiveness differences (Coyne, 2024), the present findings suggest the importance of therapists measuring their patients' outcomes. Not only would doing so embody an evidence-based practice at the case level (e.g., de Jong et al., 2021), but it would also allow clinicians to discover their personal effectiveness strengths and weaknesses at the caseload/practice level (Muir et al., 2019). In addition to learning about *who* (based on patient characteristics; e.g., Drinane et al., 2022) and/or *what* (based on patient presenting problems; e.g., Constantino et al., 2021) therapists are most

[7] To check whether therapists' overall effectiveness may have confounded our results, we examined the covariance between therapists' caseload-level differences in posttreatment outcomes (i.e., their average effectiveness) and the therapist-specific treatment slope; results showed no significant correlation between global average effectiveness and treatment-specific effectiveness ($\tau_{12} = -0.12$, $p = .094$, 95% CI [-0.49, 0.02]).

adept at treating, such routine outcome measurement can also inform *how* they are most effective in using different treatment types—at least for therapists who do not restrict themselves to learning or delivering only one specific psychotherapy brand. Although there are certainly therapists who do make such restrictions, there are many others who do not (e.g., Norcross et al., 2023). For example, in training contexts, trainees may intentionally seek a breadth of theoretical inputs into their practice, or such breadth may be required through rotations. Even beyond training, many clinicians identify as integrative for which practicing from two or more distinct approaches is one manifestation. Thus, understanding their potential strengths and weaknesses when using distinct therapies is an important element of evidence-informed treatment and decision-making.

In the simplest case, therapists can follow the data, so to speak, by specializing to their own effectiveness strength. Alternatively, clinicians who learn they are comparably effective when using either CBT or PDT would be a versatile asset to the field in that the treatment they deliver with any given patient can be based more prominently on what the *patient* prefers (Swift et al., 2018) or on what a precision algorithm suggests would be most beneficial for that patient (Cohen et al., 2021). Such therapist flexibility or adaptiveness would create an enviable scenario in which personalization efforts would be informed by multiple intersecting evidence bases, with personalization being aimed at both the patient *and* provider.

However, in some situations, differential therapist effectiveness by treatment condition may have more nuanced implications, which could vary depending on a given therapist's *average* level of effectiveness. For example, if a therapist was relatively ineffective when using PDT and harmful when using CBT, they might need to consider learning/using another approach or seek additional training if they were wedded to using one or both of these specific treatments.

Alternatively, if a therapist was exceptional when using CBT and more average when using PDT, they may increase their use of CBT (given their exceptional outcomes) but need not abandon their personally less effective use of PDT (as it can still be a viable option for a patient who prefers it or appears to be a good candidate for it). In other words, some therapists' *relative weaknesses* may still be *normative strengths*, though more work is needed to replicate the present findings and to unpack these additional complexities.

The present results also have a methodological implication. Although comparative trials in which therapists deliver more than treatment are a relatively rare design (perhaps because of concerns about allegiance and contamination effects), our results highlight some of their advantages for future therapist effects research. Notably, although it is often believed that such therapist-crossed designs eliminate or control for global between-therapist effects, they do not (as further underscored herein with therapists explaining 5% of the variance in patient outcomes). Instead, these designs disentangle treatment and therapist effects, thereby allowing one to study both with more clarity and precision. And, regarding therapist effects, researchers can simultaneously examine both between- *and* within-therapist effects with more complexity, including the latter moving beyond treatment type. For example, therapists can also be crossed on case-assignment methods, patient problem type, level of care, etc. Such foci open up a new landscape for understanding and leveraging varied forms of therapist effectiveness differences.

Future work is also needed to uncover therapist-level characteristics and practices that explain *why* they may use certain treatments more effectively when they do. Although speculative, one potential determinant of this difference could be therapists' own preferred theoretical orientation. That is, it is possible that outside of comparative trials for which therapists are obligated to deliver more than one type of treatment, providers may actually prefer the treatments they are most effective in providing (a type of knowing thyself). Of course, this notion requires direct testing, especially given research showing that therapists are inaccurate judges of their own measurement-based strengths and weaknesses in other domains (i.e., patient problems; Constantino et al., 2023).

Additionally, although it does not directly speak to *why* a given therapist may use certain treatments more (or less) effectively, it is noteworthy that when therapists in our sample had an effectiveness edge, it tended to favor CBT; that is, 79% of these therapists had better outcomes in CBT versus PDT. Although the precise reason for this difference is unknown, the present study showed a small (non-significant) average posttreatment outcome difference favoring CBT over PDT (d = .08). Thus, to be classified as *more effective* in PDT versus CBT, a given therapist needed a stronger differential effect size to overcome the on-average small difference favoring CBT. In this vein, it is possible the aforementioned pattern was a methodological artifact. Alternatively, it could be that something about CBT itself genuinely allowed a greater proportion of the therapists to use it more effectively than PDT, though more research is needed to test this idea and identify such determinants.

The present study had several limitations. As noted, we had no information to characterize the therapist sample and could therefore not examine any therapist-level predictors of the variability in their effectiveness by treatment condition. We also had little systematically collected data on demographic and clinical characteristics for patients, which makes it difficult to estimate the generalizability of the findings. Also, because only two of five studies formally assessed therapist adherence, it is possible that contamination effects could have been present in the remaining studies. Further, although patients were randomly assigned to *treatments*, no trial indicated that therapists were then randomly assigned to treat a given patient with CBT or PDT.

Thus, there is the possibility that for some therapists, systematic bias existed between the patients they treated in one condition versus the other. Additionally, although drawing on an IPDMA yielded a large patient sample, we still had a relatively small number of therapists. It should also be noted that three of the five studies were conducted by the same research team, which could introduce researcher bias. Finally, the generalizability of our findings is limited to two empirically supported treatments for depression delivered in the context of clinical trials. Limitations notwithstanding, the present study contributes to a growing literature that encourages therapists to know their strengths and weaknesses, which can inform a more comprehensive form of personalized mental health care aimed not just at the patient but also the provider.

References

Barkham, M., Rees, A., Shapiro, D. A., Stiles, W. B., Agnew, R. M., Halstead, J., Culverwell, A., & Harrington, V. M. G. (1996). Outcomes of time-limited psychotherapy in applied settings: Replicating the Second Sheffield Psychotherapy Project. *Journal of Consulting and Clinical Psychology, 64*(5), 1079–1085. https://doi.org/10.1037/0022-006X.64.5.1079

Barkham, M., Shapiro, D. A., Hardy, G. E., & Rees, A. (1999). Psychotherapy in two-plus-one sessions: Outcomes of a randomized controlled trial of cognitive–behavioral and psychodynamic–interpersonal therapy for subsyndromal depression. *Journal of Consulting and Clinical Psychology, 67*(2), 201–211. https://doi.org/10.1037/0022-006X.67.2.201

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. The Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Cohen, Z. D., Delgadillo, J., & DeRubeis, R. J. (2021). Personalized treatment approaches. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change: 50th anniversary edition* (7th ed., pp. 673–703). John Wiley & Sons, Inc.

Cook, S. C., Schwartz, A. C., & Kaslow, N. J. (2017). Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics*, *14*(3), 537–545. https://doi.org/10.1007/s13311-017-0549-4

Constantino, M. J., Boswell, J. F., Coyne, A. E., Muir, H. J., Gaines, A. N., & Kraus, D. R. (2023). Therapist perceptions of their own measurement-based, problem-specific effectiveness. *Journal of Consulting and Clinical Psychology, 91*(8), 474–484. https://doi.org/10.1037/ccp0000813

Constantino, M. J., Boswell, J. F., Coyne, A. E., Swales, T. P., & Kraus, D. R. (2021). Effect of matching therapists to patients vs assignment as usual on adult psychotherapy outcomes: A randomized clinical trial. *JAMA Psychiatry, 78*(9), 960–969. https://doi.org/10.1001/jamapsychiatry.2021.1221

Coyne, A. E. (2024). Therapist performance report cards: Do clinicians differ in their specific effectiveness? In F. T. L. Leong, J. L. Callahan, J. Zimmerman, M. J. Constantino, & C. F. Eubanks (Eds.), *APA handbook of psychotherapy: Evidence-based practice, practice-based evidence, and contextual participant-driven practice*., Vol. 2. (pp. 227–242). American Psychological Association. https://doi.org/10.1037/0000354-015

Cuijpers, P., Karyotaki, E., de Wit, L., & Ebert, D. D. (2020). The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychotherapy Research*, *30*(3), 279–293. https://doi.org/10.1080/10503307.2019.1649732

de Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, *85*, 102002. https://doi.org/10.1016/j.cpr.2021.102002

Delgadillo, J., Rubel, J., & Barkham, M. (2020). Towards personalized allocation of patients to therapists. *Journal of Consulting and Clinical Psychology*, *88*(9), 799–808. https://doi.org/10.1037/ccp0000507 dos Santos, É. N., Molina, M. L., Mondin, T., Cardoso, T. A., Silva, R., Souza, L., & Jansen, K. (2020). Long-term effectiveness of two models of brief psychotherapy for depression: A three-year follow-up randomized clinical trial. *Psychiatry Research*, *286*, 112804. https://doi.org/10.1016/j.psychres.2020.112804

Driessen, E., Abbass, A. A., Barber, J. P., Connolly Gibbons, M. B., Dekker, J. J. M., Fokkema, M., Fonagy, P., Hollon, S. D., Jansma, E. P., de Maat, S. C. M., Town, J. M., Twisk, J. W. R., Van, H. L., Weitz, E., & Cuijpers, P. (2018). Which patients benefit specifically from short-term psychodynamic psychotherapy (STPP) for depression? Study protocol of a systematic review and meta-analysis of individual participant data. *BMJ Open*, *8*(2), e018900. https://doi.org/10.1136/bmjopen-2017-018900

Driessen, E., Van Henricus, L., Don, F. J., Peen, J., Kool, S., Westra, D., Hendriksen, M., Schoevers, R., Cuijpers, P., Twisk, J. W. R., & Dekker, J. J. M. (2013). The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: A randomized clinical trial. *American Journal of Psychiatry, 170*(9), 1041-1050. https://doi.org/10.1176/appi.ajp.2013.12070899

Drinane, J. M., Roberts, T., Winderman, K., Freeman, V. F., & Wang, Y. W. (2022). The myth of the safe space: Sexual orientation disparities in therapist effectiveness. *Journal of Counseling Psychology*, *69*(3), 268–275. https://doi.org/10.1037/cou0000584

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*(1), 56–62. https://doi.org/10.1136%2Fjnnp.23.1.56

Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review*, *67*, 78–93. https://doi.org/10.1016/j.cpr.2018.08.004

Kivlighan, D. M., Jung, A.-K., Berkowitz, A. B., Hammer, J. H., & Collins, N. M. (2019). 'To show or no-show'? Therapist racial-ethnic disparities in clients' nonattendance in therapy. *Journal of College Student Psychotherapy, 33*(1), 1–13. https://doi.org/10.1080/87568225.2018.1424597

Muir, H. J., Coyne, A. E., Morrison, N. R., Constantino, M. J., & Boswell, J. F. (2019). Ethical implications of routine outcomes monitoring for patients, psychotherapists, and mental health care systems. *Psychotherapy, 56*(4)*,* 459-469. https://doi.org/10.1037/pst0000246

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

Norcross, J. C., Rocha, M. N., & Chrysler, A. A. (2023). Psychologists conducting psychotherapy in 2022: Contemporary practices and historical patterns of the Society for the Advancement of Psychotherapy. *Psychotherapy, 60*(4), 587–592. https://doi.org/10.1037/pst0000493

Nye, A., Delgadillo, J., & Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*, *91*(7), 389–397. https://doi.org/10.1037/ccp0000820

Riley R. D., Stewart, L. A., Tierney, J. F. (2021). Individual participant data meta-analysis for healthcare research. In R. D. Riley, J. F. Tierney, & L. A. Stewart (Eds), *Individual participant data meta-analysis: A handbook for healthcare research* (pp. 1–6). Wiley.

Schiefele, A. K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., Kopta, M., Schulte, D., Saxon, D., Nielsen, S. L., & Lambert, M. J. (2017). Reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health*, *44*(5), 598–613. https://doi.org/10.1007/s10488-016-0736-3

Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology*, *62*(3), 522–534. https://doi.org/10.1037/0022-006x.62.3.522

Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary care routine practice: Replication in a larger sample. *Psychological Medicine, 38*(5)*,* 677–688. https://doi.org/10.1017/S0033291707001511.

Swift, J. K., Callahan, J. L., Cooper, M., & Parkin, S. R. (2018). The impact of accommodating client preference in psychotherapy: A meta-analysis. *Journal of Clinical Psychology*, *74*(11), 1924–1937. https://doi.org/10.1002/jclp.22680

Table 1

*Individual Study Characteristics of Included Trials (*k *= 5)*

Trials	Original Trial *N*[a]	Present Study *N*[b]	Inclusion Criteria	Exclusion Criteria	Inclusion Assessment	Therapist Information	Treatment Length	Depression Measure
Barkham et al. (1996)	36 patients 4 therapists	36 patients 4 therapists	BDI score ≥ 16; PSE ≥ 5; DSM-III MDD diagnosis; white-collar employment	Received more than 3 prior therapy sessions in past 5 years; psychotropic medication changes in the past 6 weeks; presence of psychotic, manic, or obsessional symptoms; depression attributable to organic illness	Observer-rated and self-report measures	Licensed psychologists Pre-trial training: 6 months of weekly training/supervision; treatment of ≥ 5 practice cases	8 or 16 sessions[c]	BDI-I
Barkham et al. (1999)	116 patients 3 therapists	116 patients 3 therapists	BDI score ≥ 4 and ≤ 25; professional, managerial and other white-collar workers who suffered from depression and stress	Received more than 3 prior therapy sessions in past 5 years; medication change within past 6 weeks; Presence of mania or psychotic symptoms	Self-report measure	Licensed psychologists Pre-trial training: practice cases	3 sessions; 2 delivered weekly and 1 delivered 12 weeks later	BDI-I
dos Santos et al. (2020)	75 patients[d] 28 therapists	219 patients[d] 17 therapists	Adults (18-60 years); diagnosed with MDD; ≥ 2 months without psychotherapeutic / pharmacological treatment	Moderate or severe suicide risk; history of abuse/ dependence of psychoactive substances (except alcohol and tobacco); psychotic symptoms; depressive episode due to bipolar disorder	Observer-rated	Master's and PhD level graduate students Pre-trial training: theoretical and practical training in both treatments	16 (CBT) or 18 (PDT) sessions	BDI-II
Driessen et al. (2013)	341 patients 93 therapists	4 patients 1 therapist	Age 18–65 years; HAM-D score ≥ 14; presence of a	Presence of psychotic symptoms or bipolar disorder; severe	Observer-rated	Psychiatrists or psychologists with at least a master's degree	16 sessions	HAM-D

			MDE according to DSM-IV criteria	suicidality warranting immediate intensive treatment/ hospitalization; substance misuse/ abuse in the past 6 months; pregnancy		Pre-trial training: treatment-specific workshops and courses; ≥ 1 practice case		
Shapiro et al. (1994)	117 patients 5 therapists	117 patients 5 therapists	BDI score ≥ 16; PSE ID score ≥ 5; DSM-III diagnosis of MDE within past 3 months; professional, managerial and other white-collar workers	Psychiatric disorder for more than 2 years; more than 3 prior therapy sessions in 5 years; psychotropic medication change within prior 6 weeks; psychotic, manic, or obsessional symptoms; depression attributed to organic illness	Clinician-rated and self-report measures	Clinical psychologists Pre-trial training: ≥ 2 practice cases in each approach; required to meet competency threshold	8 or 16 sessions[c]	BDI-I

Note. BDI = Beck Depression Inventory; PSE = Present State Examination; DSM-III = Diagnostic and Statistical Manual of Mental Disorders, Third Edition; MDD = Major Depressive Disorder; CBT = Cognitive-behavioral therapy; PDT = psychodynamic therapy; MDE = major depressive episode; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; HAM-D = Hamilton Depression Rating Scale; PSE ID= Present State Examination-Index of Definition.

[a] Sample sizes in this column are provided as reported in the published main outcomes manuscript from the referenced trial; in some instances, the total sample size provided by the authors for inclusion in the IPDMA differed from this number.

[b] Sample sizes in this column indicate the number of patients and therapists that were included in the present study's primary analyses.

[c] Patients were randomly assigned to receive either 8 or 16 weekly sessions of CBT or PDT.

[d] The dos Santos et al. (2020) study focused on long-term follow-up outcomes among a subset of the randomized patients who received CBT or PDT. Therefore, the present study included a larger sample than the referenced long-term outcomes paper because this study focused on posttreatment outcomes.

Table 2

*Within-Therapist Effectiveness Differences as a Function of Treatment Type (*n *= 486)*

	Coefficient (*SD*)	95% CI
Fixed effects		
Posttreatment depression, γ_{00}	-0.03 (*0.15*)	0.41, -0.32
Between-therapist pretreatment depression, γ_{01}	0.64 (*2.07*)	-4.00, 4.74
Treatment condition-outcome association, γ_{10}	0.08 (*0.15*)	-0.22, 0.37
Within-therapist pretreatment depression-outcome association, γ_{20}	0.39* (*0.08*)	0.22, 0.54
Random effects		
Within-therapist residual (level 1), σ^2	0.80* (*0.07*)	0.69, 0.96
Intercept (level 2), τ_{00}	0.11* (*0.10*)	0.02, 0.42
Treatment condition-outcome slope, τ_{11}	0.28* (*0.21*)	0.07, 0.86
Within-therapist pretreatment depression-outcome slope, τ_{22}	0.05* (*0.05*)	0.01, 0.19

Note. CI = credible interval

* Indicates that the 95% Bayesian credible interval does not include zero.

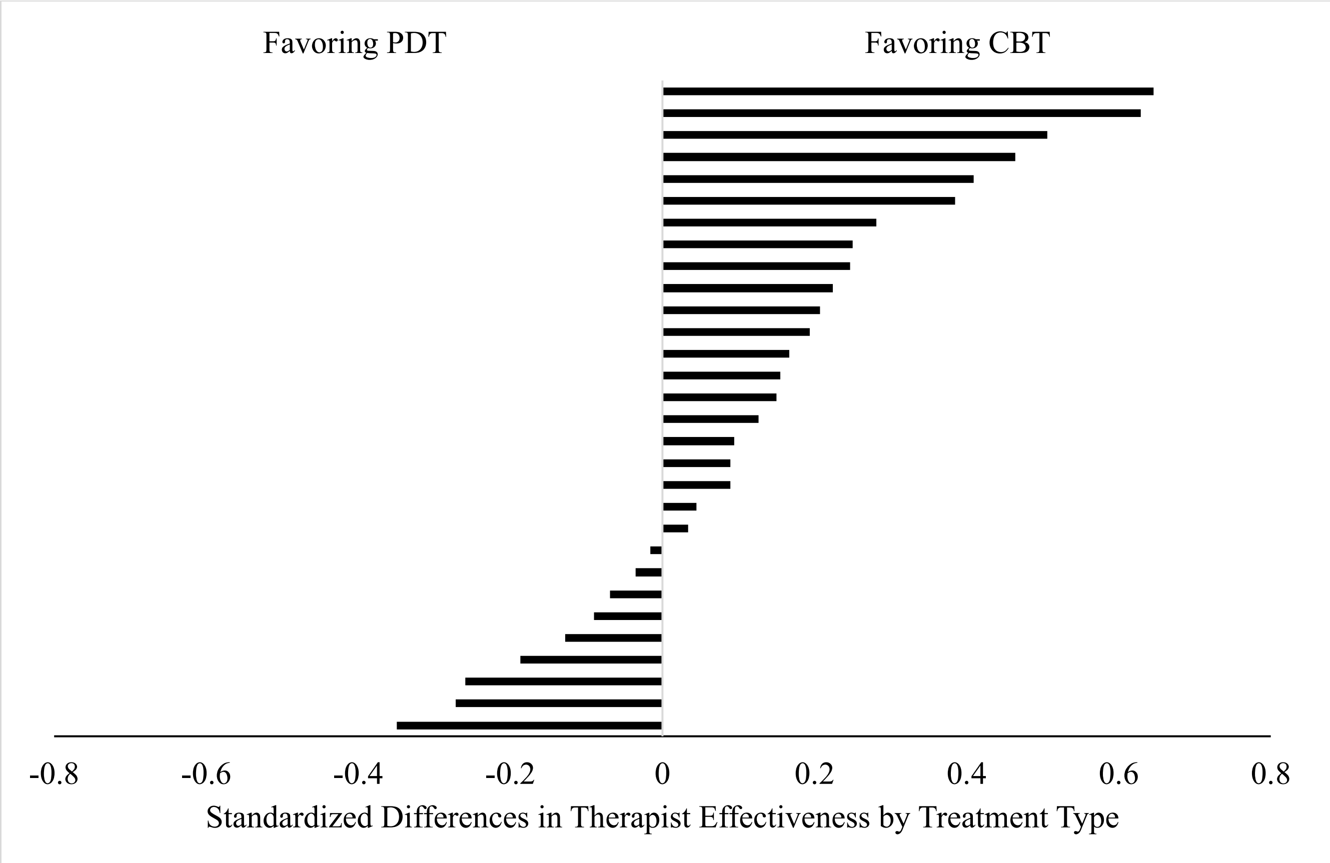

Figure 1. Standardized differences in individual therapists' average effectiveness when treatment patients with PDT vs. CBT. Each bar corresponds to a unique therapist. The length of each bar represents the magnitude of the treatment condition–outcome slope for a given therapist (i.e., the degree of average outcome difference when they delivered CBT vs. PDT). Bars favoring the left side (negative values) indicate that a given therapist was more effective, on average, when delivering PDT, whereas bars favoring the right side (positive values) indicate that a given therapist was more effective on average when delivering CBT. The standard deviation units depicted on the x-axis can be interpreted as an approximation of Cohen's *d.* Importantly, this figure does *not* depict each therapist's average posttreatment outcome across condition (i.e., between-therapist effectiveness differences, or each therapist's degree of effectiveness relative to other therapists in the sample).
CBT = cognitive behavioral therapy; PDT = psychodynamic therapy.