Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: www.tandfonline.com/journals/tpsr20

The effectiveness and efficiency of brief strategic therapy: A comparison of in-person, online, and hybrid treatment delivery modalities

Jeffrey B. Jackson, Rytis Pakrosnis, Grégoire Vitry, Zoie J. Bartling, Laurent Collongues, Padraic Gibson, Claudette Portelli & Matteo Papantuono

To cite this article: Jeffrey B. Jackson, Rytis Pakrosnis, Grégoire Vitry, Zoie J. Bartling, Laurent Collongues, Padraic Gibson, Claudette Portelli & Matteo Papantuono (22 May 2026): The effectiveness and efficiency of brief strategic therapy: A comparison of in-person, online, and hybrid treatment delivery modalities, Psychotherapy Research, DOI: 10.1080/10503307.2026.2666649

To link to this article: https://doi.org/10.1080/10503307.2026.2666649

Psychotherapy Research, 2026
https://doi.org/10.1080/10503307.2026.2666649

RESEARCH ARTICLE

The effectiveness and efficiency of brief strategic therapy: A comparison of in-person, online, and hybrid treatment delivery modalities

JEFFREY B. JACKSON [1], RYTIS PAKROSNIS [2], GRÉGOIRE VITRY [3], ZOIE J. BARTLING [1], LAURENT COLLONGUES [3], PADRAIC GIBSON [4], CLAUDETTE PORTELLI [5,6], & MATTEO PAPANTUONO [7]

[1]*School of Family Life, Brigham Young University, Provo, UT, USA;* [2]*Department of Psychology, Vytautas Magnus University, Kaunas, Lithuania;* [3]*La Clinique de L'Action pour la Transformation, Paris, France;* [4]*OCD Clinic, Dublin, Ireland;* [5]*University of Malta, Tal-Qroqq Msida, Malta;* [6]*OCD Clinic, Naxxar, Malta &* [7]*OCD Clinic, Ancona, Italy*

(Received 21 August 2025; revised 16 April 2026; accepted 22 April 2026)

Abstract
Objective: This study evaluated the effectiveness and efficiency of brief strategic therapy (BST) in routine care settings and examined whether outcomes and therapeutic alliance, as an indicator of positive therapeutic process, differed across in-person, online, and hybrid delivery modalities.
Method: Systemic Practice Research Network (SYPRENE) data included 1,933 clients from Western Europe who received BST in one of the modalities. The nonrandomized prospective pretest-posttest naturalistic design included measures of therapist-rated and client-rated problem resolution, as well as client-rated psychological impairment, overall well-being, and therapeutic alliance.
Results: On average, problem resolution was high (client-rated $M = 6.71/10$; therapist-rated $M = 5.76/10$), psychological impairment significantly decreased ($d = -2.20$), overall well-being significantly increased ($d = 0.87$), and therapeutic alliance was high with no detectable pretreatment-posttreatment change ($M = 36.91/40$). Modality was not a significant predictor of problem resolution or change in psychological impairment, overall well-being, or therapeutic alliance; clinical significance analyses largely supported these patterns. Efficiency did not differ by modality, except in-person therapy was modestly more efficient than hybrid therapy for improving overall well-being.
Conclusion: BST showed effectiveness, positive therapeutic process, and efficiency across delivery modalities; modality differences were negligible or small, although some were statistically significant. Findings indicate technology-mediated BST is a feasible therapeutic alternative to expand access to care.

Keywords: brief strategic therapy; systemic therapy; teletherapy; telehealth; treatment delivery modalities; effectiveness

Clinical or methodological significance of this article: Brief strategic therapy demonstrated substantial effectiveness and efficiency in routine care when delivered in-person, online, or in a hybrid format, helping many clients achieve meaningful improvements in their presenting problems, psychological health, and overall well-being. Strong client-therapist relationships were observed at both the beginning and end of therapy across all formats. Attrition and small longitudinal subsamples warrant cautious interpretation of between-modality comparisons. The findings provide provisional evidence that online and hybrid brief strategic therapy can help expand access to effective mental health care by providing flexible options for people who face barriers to in-person treatment.

Introduction

The integration of digital technology into healthcare has transformed mental health services. Online therapy—psychotherapy delivered remotely through digital platforms such as video conferencing, phone, email, or text-based communication—was once met

Correspondence concerning this article should be addressed to Jeffrey B. Jackson jeff_jackson@byu.edu Brigham Young University, 273 TLRB, Provo, UT 84602, USA
Supplementary data for this data can be accessed at https://doi.org/10.1080/10503307.2026.2666649

with skepticism but is now generally recognized as a flexible and effective treatment delivery modality (Aafjes-van Doorn et al., 2023; Connolly et al., 2020). Despite the growing adoption of technology-mediated therapy, growing evidence supporting online therapy's effectiveness, and the expansion of hybrid delivery models, research directly comparing the outcomes of specific psychotherapy approaches across delivery modalities remains limited, particularly for systemic therapies.

Brief strategic therapy (BST) is well-suited for examining modality differences. BST is a solution-focused, problem-oriented intervention grounded in systems theory that targets maladaptive perceptive-reactive systems and interactional patterns through structured techniques such as reframing, symptom prescription, and behavioral directives, typically achieving meaningful change within 10 sessions (Nardone & Portelli, 2005). Because BST relies primarily on verbal communication and therapeutic directives rather than in-person-specific elements (e.g., physical proximity, real-time nonverbal observation), it is theoretically amenable to cross-modality delivery. BST's established effectiveness across diverse client populations in both controlled and naturalistic settings (Jackson et al., 2018; Nardone & Watzlawick, 2005; Vitry et al., 2021) allows examination of whether outcomes differ meaningfully across delivery modalities in routine practice.

This study evaluated the effectiveness and efficiency of BST in routine care clinical settings across three treatment delivery modalities: in-person therapy (face-to-face sessions in a shared physical space), online therapy (sessions delivered virtually via video conferencing), and hybrid therapy (combining at least one in-person and one online session within the same course of treatment). The Systemic Practice Research Network (SYPRENE)—a practice-based research network established to evaluate BST effectiveness in naturalistic settings (Vitry et al., 2020, 2021)—has systematically collected outcome data across all three modalities since inception. This existing naturalistic dataset provided an opportunity to examine modality differences without requiring artificial modality assignment.

Advantages and disadvantages of treatment delivery modalities

Online therapy offers convenience and reduced barriers including scheduling, transportation, cost, and stigma (Aafjes-van Doorn et al., 2021; Barak et al., 2008; Cronin et al., 2021; Griffith et al., 2023; Kafali et al., 2014), and facilitates participation from multiple family members (Benz et al., 2022). However, challenges include limited visibility, technical disruptions, and privacy concerns (Burgoyne & Cohn, 2020; Connolly et al., 2020; Heiden-Rootes et al., 2021; Lin et al., 2022). The hybrid modality combines advantages of both online and in-person approaches (Aafjes-van Doorn et al., 2023; Babiano-Espinosa et al., 2021; Benz et al., 2022; Conroy et al., 2024; Dunn & Wilson, 2021). Recent research suggests no significant difference in therapeutic alliance between online and in-person therapy (Davis et al., 2023; Greenwood et al., 2022), although alliance development may be slower or more complex when conducted online, particularly in relational contexts or with trainee therapists (Bartle-Haring et al., 2021; Bradford et al., 2023; Lin et al., 2022; Sucala et al., 2012).

Treatment outcome evaluative criteria

When assessing the success of different psychotherapy delivery modalities, it is important to consider three distinct but related treatment outcome evaluative criteria: *effectiveness* (treatment performance in real-world clinical settings), *efficacy* (performance under controlled conditions, such as randomized clinical trials [RCTs]), and *efficiency* (ratio of resources expended to effectiveness achieved; Cochrane, 1972). Most research has examined non-systemic therapies, leaving limited evidence on systemic approaches, though emerging meta-analyses suggest comparable outcomes between online and in-person modalities despite scarce direct comparisons (Erasmus et al., 2025; McLean et al., 2021).

Treatment delivery modality. Meta-analytic research consistently finds no significant differences in efficacy between online and in-person therapy (Giovanetti et al., 2022; Lin et al., 2022). Meta-analytic evidence similarly shows that in-person, online, and hybrid therapies are equally effective across a range of disorders and client populations, supporting digital delivery as a viable option for mental health care (Barak et al., 2008; Greenwood et al., 2022; Hjelle et al., 2023; Miu et al., 2023; Wootton, 2016). In terms of efficiency, online therapy offers greater convenience and lower costs than in-person care while reducing barriers such as travel, scheduling conflicts, and stigma, and may be especially beneficial for populations facing logistical, financial, or accessibility challenges (Adam et al., 2022; Greenwood et al., 2022; Hensel et al., 2021; Kafali et al., 2014).

Brief strategic therapy (BST). BST has demonstrated efficacy, effectiveness, and efficiency across diverse populations. For example, 87% of 3,484 cates at the Centro di Terapia Strategica [Center for Strategic Therapy] in Italy reported

problem resolution in an average of about seven therapy sessions across a range of presenting problems, with gains largely maintained at one-year follow-up and low rates of relapse (Nardone & Watzlawick, 2005). Similarly, 76% of 1,150 SYPRENE clients in France and Italy with a range of presenting problems achieved clinically significant change in an average of 5.4 sessions over 5.3 months (Vitry et al., 2021). An RCT comparing BST to cognitive–behavioral therapy for binge eating disorder found that a 16-session BST hybrid protocol produced superior one-year outcomes, with 53% demonstrating reduced binge frequency, 90% with reduced weight, and 93% with improved global functioning (Jackson et al., 2018).

Study aims

Given (a) research has consistently shown no difference in efficacy, effectiveness, or efficiency between in-person therapy and online therapy, (b) other systemic models have demonstrated success in online and hybrid modes of therapy, and (c) emerging evidence supports the effectiveness and efficiency of BST, online and hybrid delivery of BST warrant examination as viable alternatives to traditional in-person BST. Accordingly, this study evaluated BST across in-person, online, and hybrid treatment delivery modalities, addressing two research questions: (1) *In naturalistic practice-based settings, is BST effective and efficient in improving client outcomes (i.e., problem resolution, psychological impairment, and overall well-being), and is the therapeutic process (i.e., therapeutic alliance) favorable?* (2) *Do the effectiveness, efficiency, and therapeutic process of BST differ across treatment delivery modalities (i.e., in-person, online, and hybrid)?* Due to the naturalistic design and exploratory nature of the modality comparisons, we examined whether observed differences in effectiveness, efficiency, and therapeutic process across delivery modalities were statistically detectable and clinically meaningful, rather than specifying directional hypotheses.

Method

Study design

The present study used a three-group prospective naturalistic-setting pretest-posttest design with modified available data analysis for evaluating the effectiveness and efficiency of BST across the treatment delivery modalities of in-person, online, and hybrid. The nonrandomized naturalistic design of this study prioritized ecological validity by analyzing data from SYPRENE, which included clients across in-person, online, and hybrid therapy modalities; this approach mirrors real-world clinical conditions, enhancing generalizability to diverse practice settings (Verster et al., 2019). As a purely observational study with no investigator-controlled treatment assignment, SYPRENE and the present investigation were not registered as a clinical trial (International Committee of Medical Journal Editors, 2024).

Over the course of its development, SYPRENE assessment instruments have been phased in, resulting in a dataset without client data for certain assessment instruments based on when they were added to SYPRENE. To enhance the external validity of our findings and preserve the integrity of the naturalistic research design, we employed a modified available data analysis. This approach retained all clients who met study inclusion criteria, including those who discontinued treatment prematurely. By doing so, we reflected practice-based patterns of attrition and varying levels of treatment adherence, thereby increasing the generalizability of our results to typical clinical settings. Unlike standard available-case analysis, our method prioritized triangulation of treatment indicators (i.e., effectiveness indicators, a therapeutic process indicator, and efficiency indicators) over maximal data retention and excluded only cases lacking essential indicator information. This strategy avoided data imputation while still acknowledging missing-not-at-random patterns common in practice-based research (e.g., phased instrument introduction in SYPRENE). Consistent with best practices for handling missing data in psychotherapy research (Schafer & Graham, 2002), the modified available data approach balances methodological rigor with naturalistic applicability, capturing the heterogeneity of clinical populations in routine care and improving the interpretability of treatment effects under real-world conditions (Little, 2024).

The inclusion criteria were (a) therapist completion of the Problem Resolution Scale (PRS) at posttreatment and (b) client completion of at least one of the following treatment indicator assessments at both pretreatment and posttreatment: the General Health Questionnaire (GHQ-12), Outcome Rating Scale (ORS), or Session Rating Scale (SRS). This multi-informant, multi-domain triangulation strategy served both substantive and methodological purposes. Substantively, comparing treatment outcomes and therapeutic process across informants and domains enabled examination of whether findings were robust across measures or selective to particular informants or treatment indicators. Methodologically, this approach supported our modified available-case analysis by ensuring each case contributed evidence from multiple perspectives, reducing dependence on single measures and

enhancing validity of treatment effect estimates in this naturalistic sample.

Participants

Clients. Of 3,552 clients participating in SYPRENE between 2012 and 2024, 1,933 (54%) met inclusion criteria: 891 (46%) in-person only, 171 (9%) online only, and 871 (45%) hybrid. All clients resided in Western Europe (primarily France and Italy); race/ethnicity data were not collected per EU privacy regulations. Demographic characteristics (age, gender, relationship status, number of children) did not differ significantly across delivery modalities (all $ps \geq .088$; see supplemental material for additional details). About 51% of clients presented with a single clinical issue and 49% presented with multiple issues; the most prevalent clinical issues spanned anxiety disorders (44%), relationship problems (43%), low self-confidence (14%), and mood disorders and disturbances (11%). Approximately 37% of clients discontinued therapy after three or fewer sessions; 74% initiated therapy before COVID-19 (77% in-person, 31% online, 78% hybrid).

Therapists. Analysis and results are based on data generated by the 33 registered SYPRENE therapists (23 females and 10 males) who contributed data on at least one completed case. Their mean age was 49.73 years ($SD = 8.36$), ranging from 30 to 63 years. Therapists reported an average of 12.3 years of experience as psychotherapists. All identified themselves as strategic systemic therapists. Due to the naturalistic design of SYPRENE, treatment delivery modality was not systematically distributed across therapists; rather, 15 therapists (45%) had cases in only a single modality (13 hybrid-only, 1 in-person-only, 1 online-only), and only 3 therapists (9%) contributed cases across all three modalities with at least 20 cases in each modality.

Procedure

Ethics approval was obtained from a university research ethics committee in France, and informed consent was obtained from all participants. SYPRENE therapists received training via monthly practice exchange groups (minimum one year) on questionnaire administration. Although the protocol eventually collected assessment data before and after each session, measurement schedules evolved over time; consequently, pretreatment-posttreatment data were available for substantially more clients than complete session-by-session series. Accordingly, analyses used pretreatment and posttreatment data only. The GHQ-12 and ORS were administered immediately before the first session and after the last session, whereas the SRS was administered immediately after the first and last sessions; for simplicity, we refer to these first- and last-session assessments as *pretreatment* and *posttreatment*, respectively.

Measures

Treatment effectiveness indicators. Problem Resolution. The Problem Resolution Scale (PRS; Vitry et al., 2024) is a single-item measure assessing the extent to which clients' presenting problem was perceived as resolved during therapy. Originating from strategic systemic therapy approaches, the PRS item was adapted and adopted in SYPRENE and then formally validated using SYPRENE data (Vitry et al., 2024). Ratings are provided by both therapist and client on an 11-point scale (0 = *problem not resolved* to 10 = *problem resolved*), with higher scores indicating greater problem resolution. Validation analyses ($n = 747$ clients) demonstrated good construct validity (therapist-client $r = .71$, $p < .001$), criterion validity correlating with GHQ-12 ($r = -.63$ to $-.66$) and ORS ($r = .44$ to $.51$), all $ps < .001$, and robustness across five major presenting problems. Test-retest reliability could not be assessed due to the scale's design, but the PRS provided time- and cost-effective assessment across all presenting problems.

Psychological Impairment. Client psychological impairment was assessed using the 12-item version of the General Health Questionnaire (GHQ-12; Goldberg, 1972; Goldberg & Williams, 1991), a self-report measure. The measure typically uses a four-point scale in which *not at all* is scored as 0, *no more than usual* is scored as 1, *rather more than usual* is scored as 2, and *much more than usual* is scored as 3; for the present study, a symmetrical five-point scale was used to provide a *less than usual* option between *not at all* and *no more than usual*. Client total GHQ-12 scores were subsequently rescaled to coincide with the four-point scale. A cumulative measure of client psychological impairment is calculated by summing the scores from the 12 items (range: 0–36); higher scores on the GHQ-12 indicate higher levels of psychological impairment. Test-retest reliability ranged from .79 to .84 in an Italian sample (Piccinelli et al., 1993). Internal consistency scores on the GHQ-12 range from .76 to .94 (Lesage et al., 2011; Sánchez-López & Dresch, 2008; Werneke et al., 2000). For the current study, internal consistency scores were not calculated due

to simplifications in GHQ-12 data collection, which only provided total scale level information, thus preventing the calculation of internal consistency scores.

Overall Well-being. The Outcome Rating Scale (ORS; Miller & Duncan, 2004) is a self-report measure of general subjective client well-being. The ORS employs a visual analog scale to rate four areas of client functioning: individual, interpersonal, social, and overall well-being over the past week (Miller et al., 2020). A cumulative measure of client overall well-being is calculated by summing the scores across the four areas (range: 0–40); higher scores on the ORS signify greater levels of client overall well-being. Test-retest reliability for non-clinical adults ranges from .58 (Miller et al., 2003) to .80 (Bringhurst et al., 2006). Internal consistency for non-clinical adults ranges from .87 to .96 (Miller et al., 2003) and has been found to be .90 for clinical adults (Campbell & Hemsley, 2009). The internal consistency score for the present study was .81 at pretreatment and .89 at posttreatment.

Treatment process indicator. Therapeutic Alliance. Therapeutic alliance served as an indicator of positive therapeutic process. The Session Rating Scale (SRS; Duncan et al., 2003) is a four-item visual analog self-report measure to assess the therapeutic alliance in four domains: relationship, goals and topics, approach or method, and an overall session evaluation. A cumulative score is calculated ranging from 0 to 40 points with higher scores indicating better evaluation of the therapeutic alliance. The validity and reliability of the scale have been established by previous studies (Duncan et al., 2003; Murphy et al., 2020). The internal consistency score for the present study was .81 at pretreatment and .88 at posttreatment.

Treatment dose variables. Therapy duration and number of therapy sessions were recorded as treatment dose variables and were later used to compute treatment efficiency indicators (i.e., improvement in treatment effectiveness indicators per unit of therapy). Therapy duration was measured as the total amount of time in months elapsed from the first therapy session to the last. The total number of therapy sessions over the course of therapy was recorded for each client.

Data analysis

Analyses were conducted using IBM Statistical Package for the Social Sciences (SPSS) Version 29 to evaluate the effectiveness, therapeutic process, and efficiency of BST, both overall and across specific treatment delivery modalities (i.e., in-person, online, and hybrid). Model equations and representative syntax appear in the supplemental material. To assess the magnitude of therapist-level clustering, intraclass correlation coefficients (*ICC*s) were calculated for each indicator using unconditional multilevel models with therapist identity as the sole random effect. Therapy duration ($M = 5.16$ months), number of therapy sessions ($M = 5.95$ sessions), and client age ($M = 43.28$ years) were grand-mean-centered prior to analysis by subtracting each variable's sample mean from individual values; grand mean centering improves the interpretability of model intercepts and coefficients (representing values at the sample mean rather than zero) and reduces potential multicollinearity among predictors and interaction terms. Residual diagnostics were conducted to evaluate the assumptions of normality and homoscedasticity for all analyzed models.

To examine problem resolution, two multiple linear regression models were estimated predicting client-rated and therapist-rated problem resolution scores. Predictor variables included therapy duration, number of therapy sessions, client age, client gender, and treatment delivery modality (represented by two dummy-coded variables with in-person as the reference category). The overall effect of treatment delivery modality was tested using a hierarchical regression approach with an omnibus *F*-test (2 *df*). Post hoc power analyses (G*Power 3.1) were conducted to evaluate statistical power for treatment delivery modality comparisons using Cohen's f^2 (negligible < 0.02; small 0.02–0.14; medium 0.15–0.34; large ≥ 0.35; Cohen, 1988).

To examine longitudinal indicators (psychological impairment, overall well-being, therapeutic alliance), general linear models (GLMs) with fixed effects were selected as the appropriate analytic approach for repeated measurements (pretreatment and posttreatment) nested within clients, providing stable parameter estimation and direct estimation of average pretreatment-posttreatment change. Random slope estimation was not feasible with only two timepoints per client, as it requires within-person variability across multiple observations to be identified, which a two-timepoint design cannot provide (Meteyard & Davies, 2020). Random intercept models were also not warranted given the absence of significant pretreatment differences across treatment delivery modalities, as confirmed by ANOVA analyses: psychological impairment ($F(2, 247) = 0.66$, $p = .520$), overall well-being ($F(2, 273) = 0.88$, $p = .415$), and therapeutic alliance ($F(2, 214) = 0.78$, $p = .459$). Models included time, treatment delivery modality, and their interaction as primary predictors,

with therapy duration, number of sessions, age, and gender as covariates. Time was reverse-coded so that pretreatment served as the reference; negative coefficients indicate decreases and positive coefficients indicate increases from pretreatment to posttreatment. Unstandardized coefficients were calculated for all variables; standardized coefficients (β), growth modeling analysis standardized slope coefficients (GMA *d*), and classical Cohen's *d* were calculated for time only, with 95% confidence intervals. GMA *d* quantifies within-person change relative to baseline variability by adapting Feingold's (2009) framework for single-group designs, whereas Cohen's *d* provides the standardized pretreatment-posttreatment difference. Effect sizes were interpreted using established benchmarks: negligible ($|d| < 0.20$), small ($0.20 \leq |d| < 0.50$), medium ($0.50 \leq |d| < 0.80$), large ($0.80 \leq |d| < 1.20$), very large ($1.20 \leq |d| < 2.00$), or huge ($|d| \geq 2.00$; Cohen, 1988; Sawilowsky, 2009). Post hoc power analyses (G*Power 3.1) were conducted to evaluate statistical power for treatment delivery modality comparisons using Cohen's *f* (negligible < 0.10; small 0.10–0.24; medium 0.25–0.39; large ≥ 0.40; Cohen, 1988).

In terms of clinical significance, chi-square goodness-of-fit tests examined whether each treatment modality demonstrated clinical outcome rates that differed from the overall sample rates for each clinical significance level. Client- and therapist-rated problem resolution scores were categorized as unresolved (0–3), partially resolved (4–6), or resolved (7–10) in accordance with established interpretive ranges (Vitry et al., 2021). Psychological impairment and overall well-being were assessed using a categorical classification system based on reliable change and clinical cutoff criteria (Jacobson & Truax, 1991). Cases with pretreatment scores above the clinical cutoff score were omitted from the clinical significance analyses. Clients who demonstrated reliable improvement were categorized as either recovered (if their posttreatment score crossed the clinical cutoff into the non-clinical range) or non-recovered (if their score improved reliably but remained within the clinical range). Clients whose scores exhibited no reliable change were classified as unchanged, whereas those who demonstrated reliable deterioration were categorized as deteriorated. Clinical significance was evaluated separately for psychological impairment and overall well-being. For psychological impairment, a clinical cutoff of 17.36 and a reliable change index (RCI) of 8.72 were established using reliability and normative data from French and Spanish populations (Lesage et al., 2011; Sánchez-López & Dresch, 2008). For overall well-being, a clinical cutoff of 25 and an RCI of 5 were used, based on U.S. normative data (Seidel & Miller, 2012). Chi-square tests were conducted to examine differences in clinical significance outcomes across treatment delivery modalities.

Therapy efficiency was operationalized as the amount of clinical improvement per unit of therapy. For each client, efficiency per month was calculated as the problem resolution score divided by the total therapy duration in months and the difference between posttreatment and pretreatment scores divided by total therapy duration in months for psychological impairment and overall well-being. Similarly, efficiency per session was calculated as the problem resolution score divided by the total number of sessions completed and the difference between posttreatment and pretreatment scores divided by the total number of sessions completed for psychological impairment and overall well-being. Group differences in treatment efficiency indicators across treatment delivery modalities were examined using one-way ANOVA tests for normally distributed variables and Kruskal–Wallis tests for non-normally distributed variables. No missing or zero values were present for the number of therapy sessions or therapy duration. All treatment efficiency indicators were computed such that cases with missing pretreatment or posttreatment values were assigned missing values for the relevant efficiency indicators. All analyses were conducted using listwise deletion for missing data. Post hoc power analyses using G*Power 3.1 were conducted to evaluate statistical power for treatment delivery modality efficiency comparisons.

Results

The descriptive statistics of the treatment indicators for the entire sample are presented in Supplemental Table S2 and by delivery modality in Supplemental Table S3. Pairwise analyses indicated no pretreatment differences between clients who initiated therapy before COVID-19 and those who initiated during and after on psychological impairment, overall well-being, and therapeutic alliance, although some of the analyses were underpowered because those measures were instituted only several months before COVID-19. *ICC* values for therapist effects ranged from .00 to .26, indicating that between-therapist variance accounted for 0% to 26% of indicator variance depending on the measure. Larger therapist effects were observed for therapist-rated problem resolution (*ICC* = .22) and posttreatment psychological impairment (*ICC* = .26), whereas the other indicators showed negligible clustering (*ICC*

≤ .07). Although these *ICC*s suggested meaningful therapist-level variability for some indicators, the naturalistic design resulted in substantial confounding between therapist and treatment modality, preventing reliable multilevel modeling (see limitations).

Because only a minority of clients had complete pre–post self-report data (12.9% for psychological impairment, 14.3% for overall well-being, 11.2% for therapeutic alliance), we conducted attrition analyses to examine representativeness of the longitudinal subsamples, given that missingness likely reflected both the phased introduction of outcome instruments within SYPRENE, which introduced administrative missingness, and some degree of differential treatment engagement, which introduced attrition-related missingness. Clients with complete psychological impairment data attended slightly more sessions and had somewhat longer treatment (M = 6.66 sessions and 6.66 months vs. 5.84 sessions and 4.92 months), corresponding to roughly one extra session and under two additional months, whereas for overall well-being and therapeutic alliance, differences in sessions and duration between complete and incomplete cases were small and not statistically or clinically meaningful (with only a trivial Mann–Whitney difference for overall well-being duration). Early discontinuation (≤3 sessions) was substantially less common among clients with complete psychological impairment data (11.2% vs. 40.9%) and among those providing client-rated problem resolution at the last session (14.4% vs. 45.1%), indicating that psychological impairment and client-rated problem resolution under-represent early discontinuation rates, whereas early discontinuation rates were comparable for clients with and without complete overall well-being and therapeutic alliance data, suggesting that missingness for overall well-being and therapeutic alliance was more consistent with instrument phasing than with differential attrition. Full results of these attrition analyses are reported in the Supplemental Material.

Problem resolution indicators

To evaluate predictors of treatment-related problem resolution, two multiple regression models were estimated using client-rated and therapist-rated PRS scores as indicators. Gender was initially included as a predictor in the client-rated model but was removed in a refined model due to its clear nonsignificance (p = .648) and negligible contribution to model fit; this adjustment resulted in a significant improvement in model performance. Residual diagnostics confirmed regression assumptions were adequately satisfied. Visual inspection of histograms and Q-Q plots indicated approximately normal residuals with minor tail deviations. Although formal normality tests were statistically significant for both models (p < .001), this likely reflects sample size sensitivity rather than meaningful violations. Visual evidence and the documented robustness of regression to such minor deviations (Knief & Forstmeier, 2021) supported retention of ordinary least squares estimation. Scatterplots showed random residual dispersion (homoscedasticity), and multicollinearity diagnostics indicated no problematic collinearity among predictors (all Condition Indices < 3.4, well below the threshold of 10.0). Multiple regression model results are summarized in Table I.

Table I. Multiple regression predicting client- and therapist-rated problem resolution.

Predictor	*B*	*SE*	β	*t*	*p*
Client-rated problem resolution (n = 384)					
Constant	**6.51**	**0.17**	**-**	**39.02**	**< .001**
Treatment delivery modality					.695
Hybrid vs. in-person	0.30	0.24	0.06	1.27	.206
Online vs. in-person	0.09	0.33	0.02	0.29	.733
Therapy duration (in months)	**0.07**	**0.02**	**0.24**	**2.96**	**.003**
Number of therapy sessions	**−0.06**	**0.02**	**−0.20**	**−2.44**	**.015**
Age	−0.01	0.01	−0.07	−1.27	.204
Therapist-rated problem resolution (n = 1,294)					
Constant	**5.57**	**0.12**	**-**	**49.56**	**< .001**
Treatment delivery modality					.499
Hybrid vs. in-person	0.17	0.15	0.03	1.09	.275
Online vs. in-person	−0.04	0.26	0.00	−0.14	.890
Therapy duration (in months)	**0.10**	**0.02**	**0.23**	**5.29**	**<.001**
Number of therapy sessions	**0.07**	**0.02**	**0.16**	**3.78**	**<.001**
Age	0.00	0.01	0.01	0.35	.730
Gender	−0.16	0.15	−0.03	−1.04	.297

Note. B = unstandardized regression coefficient; SE = standard error; β = standardized regression coefficient; t = t-value; p = p-value. Therapy duration, number of therapy sessions, and age were grand mean centered; gender (0 = female, 1 = male). Treatment delivery modality p-value represents the omnibus F-test (2 df). Bolded rows have p-values that are significant at the p < .05 level.

The average posttreatment rating of the extent to which clinical issues were resolved for the entire sample was 6.71 (SD = 2.40; n = 507) based on client evaluations and 5.76 (SD = 2.71; n = 1,933) based on therapist evaluations (Table S2). Among

the 507 clients for whom both therapist and client ratings were available, a paired-samples *t*-test indicated a statistically significant difference between therapist-rated ($M = 6.92$, $SD = 2.10$) and client-rated ($M = 6.71$, $SD = 2.40$) evaluations ($t(506) = 2.67$, $p = .008$); however, the negligible effect size ($g = 0.12$; 95% *CI* [0.03, 0.21]) suggests that the discrepancy between therapist and client ratings is of limited practical significance.

In the client-rated problem resolution model (Table I), which included 384 clients with complete data, the predictors accounted for approximately 2.9% of the variance in client-rated problem resolution scores ($F(5, 378) = 2.26$, $p = .048$). Treatment delivery modality was not a significant predictor of client-rated problem resolution ($F(2, 373) = 0.36$, $f^2 = .002$, $p = .695$). Post hoc power analysis indicated adequate power (.69) to detect a small effect ($f^2 = .02$), exceeding the conventional .80 threshold for a marginally larger small effect ($f^2 = .025$). Given the nonsignificant negligible observed effect, the finding likely reflects an absence of modality differences rather than Type II error. Longer therapy duration was significantly associated with greater perceived problem resolution when controlling for number of sessions, client age, and treatment delivery modality ($B = 0.07$, $SE = 0.02$, $\beta = .24$, $p = .003$); on average, each additional month of therapy was linked to a 0.07-point increase in client ratings (PRS scores range from 0 to 10, with 10 indicating complete problem resolution). Conversely, a greater number of sessions was associated with slightly lower perceived resolution when controlling for therapy duration, client age, and treatment delivery modality ($B = -0.06$, $SE = 0.02$, $\beta = -.20$, $p = .015$); on average, each additional therapy session was associated with a 0.06-point decrease in client-rated scores. Although this was a small effect, it may reflect greater treatment difficulty among clients who required more sessions rather than indicating that additional sessions were detrimental (e.g., clients requiring more sessions may have presented with more complex or persistent problems). Client age was not a significant predictor of client-rated problem resolution ($p = .204$).

In the therapist-rated problem resolution model (Table I), which included 1,294 clients, predictors explained 14.1% of the variance in therapist ratings ($F(6, 1{,}287) = 35.09$, $p < .001$). Treatment delivery modality was not a significant predictor of therapist-rated problem resolution ($F(2, 1{,}287) = 0.70$, $f^2 = .001$, $p = .499$); the analysis achieved excellent power (.99) to detect a small effect ($f^2 = .02$). Longer therapy duration was associated with greater therapist-rated problem resolution ($B = 0.10$, $SE = 0.02$, $\beta = .23$, $p < .001$) when controlling for all other variables in the model, with each additional month of therapy on average associated with a 0.10-point increase in therapist-rated scores. A greater number of sessions was also associated with higher therapist-rated problem resolution ($B = 0.07$, $SE = 0.02$, $\beta = .16$, $p < .001$) when controlling for all other variables in the model, with each additional session on average corresponding to a 0.07-point increase in therapist-rated problem resolution. Gender ($p = .297$) and client age ($p = .730$) were not significant predictors.

Clinical significance rates by treatment modality are presented in Table III. In terms of clinical significance based on client-reported problem resolution ($N = 507$), 60.5% were classified as resolved, 29.0% as partially resolved, and 10.5% as unresolved. The majority of clients reported scores classified as problem resolution regardless of treatment modality, with resolution rates falling between 52% and 64%. The overall association between client-rated clinical significance and treatment delivery modality was statistically significant ($\chi^2(4, N = 507) = 9.57$, $p = .048$, Cramér's $V = .10$), indicating a small effect size. Resolved cases ($\chi^2(2, N = 307) = 1.24$, $p = .538$) and unresolved cases ($\chi^2(2, N = 53) = 2.14$, $p = .343$) demonstrated comparable rates across modalities, indicating that rates were consistent across treatment modalities relative to their sample representation. Conversely, partially resolved cases showed a significant departure from expected rates ($\chi^2(2, N = 147) = 6.19$, $p = .045$), with online clients (41.7%) more likely than expected to achieve partial resolution and hybrid clients (24.2%) less likely than expected to achieve partial resolution.

In terms of clinical significance based on therapist-reported problem resolution ($N = 1{,}933$), 46.2% were classified as resolved, 33.6% as partially resolved, and 20.2% as unresolved. Therapist ratings revealed significant modality differences across all three outcome levels ($\chi^2(4, N = 1{,}933) = 49.32$, $p < .001$, Cramér's $V = .11$), representing a small effect size. Resolved cases ($\chi^2(2, N = 893) = 8.26$, $p = .016$) and unresolved cases ($\chi^2(2, N = 391) = 9.18$, $p = .010$) showed significant modality differences, with hybrid clients showing higher than expected rates (51.1% and 23.5% respectively) and in-person clients showing lower than expected rates (42.3% and 17.1% respectively). Partially resolved cases demonstrated the strongest modality effect ($\chi^2(2, N = 649) = 31.94$, $p < .001$), with hybrid clients (25.4%) less likely than expected to be classified as reaching partial resolution and in-person clients (40.6%) more likely than expected to be classified as reaching partial resolution.

The findings suggest differential outcome patterns across treatment delivery modalities, particularly for

therapist-rated assessments. These differences, although statistically significant, should be interpreted cautiously given the potential for Type I error inflation across multiple comparisons. The observed Cramér's V values (.10 for client ratings and .11 for therapist ratings) indicate small effect sizes, suggesting that treatment delivery modality accounts for minimal variance in problem resolution and does not substantially impact therapeutic effectiveness.

Longitudinal indicators

Visual inspection of histograms, Q–Q plots, and residual scatterplots for all three pretreatment-posttreatment indicators revealed approximate normality and homoscedasticity, with only minor deviations at distribution tails, supporting the use of standard GLM procedures. Although formal normality tests (Kolmogorov–Smirnov, Shapiro–Wilk) were statistically significant for some residuals, this was judged inconsequential given the minor visual deviations and the known robustness of GLM and F-tests to such violations. Multicollinearity diagnostics for centered therapy duration and number of therapy sessions were well within acceptable limits across all models (all variance inflation factor values < 2.6, all tolerance values > 0.38), indicating no problematic collinearity among covariates. No transformations or robust estimators were applied. GLM

Table II. General linear models predicting within-person change over time in longitudinal indicators.

Predictor	Estimate (B)	SE	95% CI	t	F	p
Psychological impairment (n = 250)						
Intercept (pretreatment)	**23.27**	**1.50**	**[20.32, 26.22]**			**<.001**
Time (posttreatment vs pretreatment)	**−13.03**	**1.93**	**[−16.83, −9.23]**			**<.001**
Treatment delivery modality					1.73	.179
Online vs. in-person	0.17	1.50	[−2.79, 3.12]	0.11		.912
Hybrid vs. in-person	−0.94	0.89	[−2.69, 0.80]	−1.06		.289
Time × treatment delivery modality (interaction)					0.05	.956
Time × online vs. in-person	0.12	2.09	[−3.99, 4.23]	0.06		.955
Time × hybrid vs. in-person	−0.34	1.24	[−2.78, 2.11]	−0.27		.787
Therapy duration (in months)	−0.07	0.08	[−0.23, 0.09]			.403
Number of therapy sessions	0.01	0.11	[−0.22, 0.23]			.952
Age	−0.02	0.02	[−0.06, 0.02]			.414
Gender	1.00	0.64	[−0.25, 2.25]			.117
Overall well-being (n = 276)						
Intercept (pretreatment)	**16.69**	**1.39**	**[13.96, 19.42]**			**<.001**
Time (posttreatment vs pretreatment)	**7.55**	**1.74**	**[4.13, 10.98]**			**<.001**
Treatment delivery modality					**5.00**	**.007**
Online vs. in-person	−1.50	1.46	[−4.36, 1.36]	−1.03		.304
Hybrid vs. in-person	−1.54	1.33	[−4.15, 1.07]	−1.16		.247
Time × treatment delivery modality (interaction)					0.92	.398
Time × online vs. in-person	−1.19	2.05	[−5.22, 2.84]	−0.58		.563
Time × hybrid vs. in-person	−2.50	1.86	[−6.14, 1.15]	−1.35		.179
Therapy duration (in months)	−0.01	0.15	[−0.30, 0.28]		−0.16	.926
Number of therapy sessions	0.00	0.16	[−0.33, 0.32]			.989
Age	−0.01	0.03	[−0.07, 0.04]			.650
Gender	−1.20	0.85	[−2.87, 0.46]			.156
Therapeutic alliance (n = 217)						
Intercept (pretreatment)	**37.01**	**0.77**	**[35.51, 38.52]**			**<.001**
Time (posttreatment vs pretreatment)	0.69	0.96	[−1.20, 2.58]			.474
Treatment delivery modality					2.70	.068
Online vs. in-person	0.47	0.81	[−1.12, 2.06]	0.58		.563
Hybrid vs. in-person	−0.01	0.72	[−1.43, 1.40]	−0.02		.988
Time × treatment delivery modality (interaction)					1.95	.144
Time × online vs. in-person	1.23	1.13	[−0.99, 3.45]	1.09		.278
Time × hybrid vs. in-person	1.92	1.01	[−0.06, 3.91]	1.90		.058
Therapy duration (in months)	**−0.17**	**0.09**	**[−0.34, 0.00]**			**.050**
Number of therapy sessions	0.16	0.09	[−0.01, 0.33]			.059
Age	−0.02	0.02	[−0.05, 0.01]			.267
Gender	0.30	0.45	[−0.60, 1.19]			.514

Note. B = unstandardized regression coefficient; SE = standard error; CI = confidence interval; t = t-value; F = F-value; p = p-value. Time (0 = pretreatment, 1 = posttreatment); gender (0 = female, 1 = male); therapy duration, number of therapy sessions, and age were grand mean centered. Bolded rows have p-values that are significant at the $p < .05$ level.

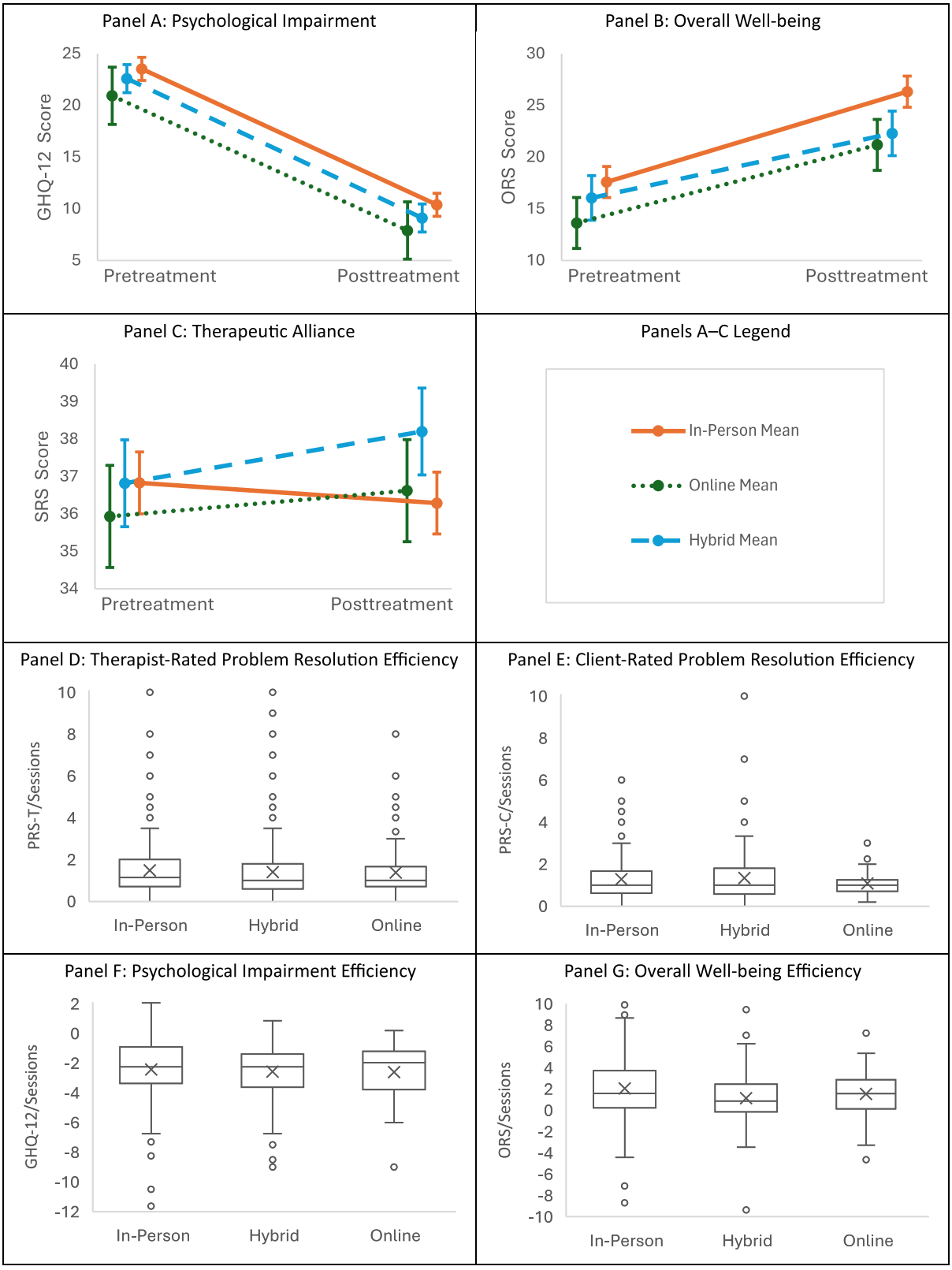

Figure 1. *Treatment Effectiveness, Process, and Efficiency by Delivery Modality.*
Note. Line graphs (Panels A–C) display pretreatment-posttreatment change in longitudinal indicators, showing estimated marginal means with 95% confidence intervals from general linear models; horizontal offsets enhance visibility of overlapping confidence intervals. Boxplots (Panels D–G) display treatment efficiency per session, showing the interquartile range (middle 50% of efficiency scores within each box), median (horizontal line), mean (×), and outliers (circles) for each treatment modality; vertical lines extending from each box indicate the range of non-outlier values.

model results are summarized in Table II, which reports omnibus *F*-values for each predictor as well as unstandardized coefficients, standard errors, 95% confidence intervals, *t*-values, and *p*-values for online and hybrid modalities relative to in-person and for the corresponding time × modality interaction terms.

Psychological impairment. The model accounted for 56.4% of the variance in psychological impairment scores. Psychological impairment significantly decreased ($F(1, 408) = 297.82$, $p < .001$) from pretreatment ($M = 23.70$) to posttreatment ($M = 10.67$), representing a very large to huge effect ($B = -13.15$, $\beta = -1.06$, GMA $d = -2.11$, 95% *CI* [−2.72, −1.49], $p < .001$; Cohen's $d = -2.20$, 95% *CI* [−2.57, −1.88], $p < .001$; Figure 1, Panel A). None of the covariates (therapy duration, number of sessions, age, gender) significantly predicted change in impairment. Treatment delivery modality did not significantly affect psychological impairment. The main effect of modality was nonsignificant ($F(2, 248) = 1.73$, $f = .09$, $p = .179$), as was the time × modality interaction ($F(2, 248) = 0.05$, $f = .01$, $p = .956$). Post hoc power analysis indicated power of .27 to detect the lowest bound of a small modality effect ($f = .10$) but adequate power (.80) for larger small effects ($f > .19$). Although the observed modality interaction effect was negligible and nonsignificant, the combination of limited power, significant attrition, and imbalanced subgroup sizes —particularly for the online modality—restricts confidence in this null finding. Estimated marginal means indicated substantial within-modality improvements across all modalities: in-person ($\Delta = 13.15$, $p < .001$), online ($\Delta = 13.03$, $p < .001$), and hybrid ($\Delta = 13.48$, $p < .001$), with overlapping confidence intervals suggesting practical equivalence.

In terms of clinical significance (Table III), 43 clients (17.2%) were omitted due to pretreatment scores above the clinical cutoff, leaving 207 clients for analysis. The distribution of omitted cases did not differ significantly across modalities ($\chi^2(2, N = 250) = 1.15$, $p = .564$). Across the entire sample, 82.6% of clients demonstrated reliable improvement in psychological impairment (80.7% met recovery criteria and 1.9% were classified as non-recovered), 16.9% remained unchanged, and 0.5% deteriorated. The overall between-group comparison across recovered, non-recovered, unchanged, and deteriorated categories was not statistically significant ($\chi^2(6, N = 207) = 4.47$, $p = .613$), indicating comparable clinical outcomes across treatment delivery modalities.

Overall well-being. The model accounted for 19.9% of the variance in overall well-being scores. Overall well-being scores significantly increased ($F(1, 466) = 78.35$, $p < .001$) from pretreatment

Table III. Treatment outcome effectiveness: clinical significance.

Modality	In-person		Online		Hybrid				
	n	%	*n*	%	*n*	%	χ^2	*df*	*p*
Client-rated problem resolution									
Resolved	142	60.9	44	52.4	121	63.7	1.24	2	.538
Partially resolved	**66**	**28.3**	**35**	**41.7**	**46**	**24.2**	**6.19**	**2**	**.045**
Unresolved	25	10.7	5	6.0	23	12.1	2.14	2	.343
Therapist-rated problem resolution									
Resolved	**377**	**42.3**	**71**	**41.5**	**445**	**51.1**	**8.26**	**2**	**.016**
Partially resolved	**362**	**40.6**	**66**	**38.6**	**221**	**25.4**	**31.94**	**2**	**< .001**
Unresolved	**152**	**17.1**	**34**	**19.9**	**205**	**23.5**	**9.18**	**2**	**.010**
Psychological impairment									
Improved	93	81.6	17	77.3	61	85.9	1.06	2	.588
Recovered	91	79.8	16	72.7	60	84.5	1.61	2	.446
Non-recovered	2	1.8	1	4.5	1	1.4	0.92	2	.633
Unchanged	21	18.4	5	22.7	9	12.7	1.62	2	.445
Deteriorated	0	0	0	0	1	1.4	1.93	2	.382
Overall well-being									
Improved	**82**	**70.1**	**28**	**54.9**	**34**	**51.5**	**7.36**	**2**	**.025**
Recovered	**63**	**53.8**	**19**	**37.3**	**24**	**36.4**	**6.91**	**2**	**.032**
Non-recovered	19	16.2	9	17.6	10	15.2	0.13	2	.936
Unchanged	**29**	**24.8**	**19**	**37.3**	**28**	**42.4**	**6.66**	**2**	**.036**
Deteriorated	6	5.1	4	7.8	4	6.1	0.47	2	.792

Note. Improved (reliable change in the positive direction), non-recovered (improved but did not cross the clinical cutoff score), recovered (improved and crossed the clinical cutoff score), unchanged (no reliable change), and deteriorated (reliable change in a worse direction). Bolded rows have *p*-values that are significant at the $p < .05$ level.

($M = 16.69$) to posttreatment ($M = 24.24$), representing a medium to very large effect ($B = 7.55$, $\beta = 0.51$, GMA $d = 1.01$, 95% CI [0.55, 1.47], $p < .001$; Cohen's $d = 0.87$, 95% CI [0.71, 1.03], $p < .001$; Figure 1, Panel B). None of the covariates (therapy duration, number of sessions, age, gender) significantly predicted change in well-being. Although treatment delivery modality demonstrated a small significant main effect ($F(2) = 5.00$, $f = 0.15$, $p = .007$; reflecting higher average well-being levels in the in-person group, particularly at posttreatment; see Table S3), modality did not moderate improvement from pretreatment to posttreatment (time × modality interaction: $F(2) = 0.92$, $f = 0.06$, $p = .398$). Post hoc power analysis indicated power of .30 to detect the lowest bound of a small modality effect ($f = .10$) but adequate power (.80) for larger small effects ($f > .18$); although the nonsignificant observed modality interaction effect was negligible ($f = 0.06$) and limited power restricts confidence in this null finding, even a statistically significant negligible effect would likely lack clinical meaningfulness. Estimated marginal means indicated all modalities showed substantial improvements: in-person ($\Delta = 8.74$, $p < .001$), online ($\Delta = 7.55$, $p < .001$), and hybrid ($\Delta = 6.24$, $p < .001$), with overlapping confidence intervals suggesting practical equivalence in improvement.

In terms of clinical significance (Table III), 42 clients (15.2%) were omitted due to pretreatment scores above the clinical cutoff, leaving 234 clients for analysis. The distribution of omitted cases was not significantly different across modalities ($\chi^2(2, N = 276) = 1.33$, $p = .514$). Across the entire sample, 59.8% of clients showed reliable improvement in overall well-being (44.9% achieved recovery and 16.3% were classified as non-recovered), 34.2% remained unchanged, and 6.0% deteriorated. The overall between-group comparison of clinical significance categories approached significance ($\chi^2(6, N = 234) = 8.83$, $p = .184$). Significant differences emerged when comparing specific categories, with in-person clients more likely to improve (70.1%) compared to online (54.9%) and hybrid (51.5%) modalities ($\chi^2(2) = 7.36$, $p = .025$). Similarly, recovery rates were significantly higher for in-person clients (53.8%) than for online (37.3%) and hybrid (36.4%) clients ($\chi^2(2) = 6.91$, $p = .032$). Clients in the online and hybrid conditions were more likely to remain unchanged (37.3% and 42.4%, respectively, vs. 24.8% for in-person; $\chi^2(2) = 6.66$, $p = .036$), whereas deterioration rates were low and did not differ significantly across modalities ($\chi^2(2) = 0.47$, $p = .792$).

Therapeutic alliance. Therapeutic alliance was examined as an indicator of positive therapeutic process. Alliance scores did not differ significantly between pretreatment and posttreatment ($B = 0.69$, $\beta = 0.10$, GMA $d = 0.20$, 95% CI [−0.35, 0.76], $p = .474$; Cohen's $d = 0.08$, 95% CI [−0.08, 0.24], $p = .331$; Figure 1, Panel C). The model accounted for 4.6% of the variance in alliance scores. Change from pretreatment to posttreatment was nonsignificant ($F(1, 362) = 1.21$, $p = .272$), with estimated marginal means of 37.01 at pretreatment and 37.70 at posttreatment. Of the covariates, only therapy duration significantly predicted alliance scores ($B = -0.17$, 95% CI [−0.34, 0.00], $p = .050$), indicating a small decline per additional month of treatment. Treatment delivery modality did not significantly predict alliance ($F(2, 362) = 2.70$, $f = 0.12$, $p = .068$) or moderate change from pretreatment to posttreatment (time × modality interaction: $F(2) = 1.95$, $f = 0.10$, $p = .144$). Post hoc power analysis revealed power of .24 to detect the lowest bound of a small modality effect ($f = 0.10$) but adequate power (.80) for larger small effects ($f > .21$). Although the observed modality interaction effect was borderline negligible and non-significant, power levels restrict confidence in this null finding; however, even if adequately powered, a significant effect size of this magnitude would suggest little practical difference in therapeutic alliance across modalities. Estimated marginal means showed nonsignificant changes across modalities: in-person ($\Delta = 0.54$, $p = .363$), online ($\Delta = 0.69$, $p = .474$), and hybrid ($\Delta = 1.38$, $p = .092$).

Treatment efficiency indicators

Duration-based treatment efficiency. Therapy duration was non-normally distributed across the sample. The mean duration was 5.16 months ($SD = 6.60$), with a median of 2.76 months and a mode of 1 d (i.e., a single session). The first quartile of clients completed less than 0.92 months of therapy, the second quartile between 0.92 and 2.76 months, the third quartile between 2.76 and 6.67 months, and the fourth quartile more than 6.67 months. Mean ranks for therapy duration were 976.48 for in-person, 964.09 for hybrid, and 932.43 for online delivery. A Kruskal–Wallis test indicated no statistically significant differences in therapy duration across treatment delivery modalities ($H(2, N = 1{,}933) = 0.94$, $p = .626$); therefore, post hoc pairwise comparisons were not conducted.

Treatment efficiency based on therapy duration was calculated as the average pretreatment-posttreatment change per month of therapy, computed by dividing the magnitude of pretreatment-posttreatment change by the total duration of therapy in

months for psychological impairment and overall well-being. Because problem resolution was measured using a single-item 0–10 scale referencing perceived problem resolution from the start of therapy, dividing the final score by therapy duration can yield extreme and potentially uninterpretable efficiency values, especially for very brief therapies; to address this, we calculated the median efficiency scores for therapist-rated problem resolution and client-rated problem resolution. On average, each month of therapy was associated with an 1.70-point increase in therapist-rated problem resolution, a 1.31-point increase in client-rated problem resolution, a 3.96-point decrease in psychological impairment ($SD = 4.80$), and a 3.21-point increase in overall well-being.

All duration-based treatment efficiency indicators were non-normally distributed: therapist-rated problem resolution efficiency per month (skewness = 4.61, kurtosis = 21.80), client-rated problem resolution efficiency per month (skewness = 11.40, kurtosis = 144.96), psychological impairment efficiency per month (skewness = −2.18, kurtosis = 6.24), and overall well-being efficiency per month (skewness = 0.57, kurtosis = 9.31). Results from Kruskal–Wallis tests indicated no significant treatment delivery modality differences for therapist-rated problem resolution efficiency ($H(2, 1{,}931) = 1.56$, $p = .458$), client-rated problem resolution efficiency ($H(2, 505) = 1.82$, $p = .403$), and psychological impairment efficiency ($H(2, 247) = 1.72$, $p = .423$); however, results from a Kruskal–Wallis test revealed a significant difference in efficiency per month for overall well-being across treatment delivery modalities ($H(2, 273) = 6.89$, $p = .032$). The mean rank for efficiency was highest in the in-person group (148.57), followed by the online group (140.16), and lowest in the hybrid group (119.08). Follow-up Mann–Whitney *U*-tests with Bonferroni correction (adjusted $\alpha = .017$) indicated that in-person therapy was significantly more efficient per month than hybrid therapy ($U = 4{,}342.00$, $p = .010$), whereas differences between in-person and online ($U = 3{,}755.00$, $p = .470$) and between hybrid and online ($U = 1{,}865.00$, $p = .111$) were not statistically significant. Post hoc power analyses revealed variable statistical power across efficiency comparisons, with adequate power (.65) for the significant overall well-being efficiency finding ($f = .16$, $p = .032$); however, the remaining duration-based efficiency comparisons demonstrated negligible effect sizes (therapist-rated problem resolution efficiency: $f = .03$, $p = .458$); client-rated problem resolution efficiency: ($f = .06$, $p = .403$); psychological impairment efficiency: ($f = .08$, $p = .423$), and even if adequately powered, effect sizes of this magnitude would lack meaningful clinical implications for treatment.

Session-based treatment efficiency. The number of therapy sessions was also non-normally distributed. The mean was 5.95 sessions ($SD = 5.35$), with a median of 4 and a mode of 2 sessions. The first quartile of clients completed 3 or fewer sessions, the second quartile 3–4 sessions, the third quartile 5–7 sessions, and the fourth quartile 8 or more sessions. Reflecting our modified available data analysis approach, 7.8% of clients completed only one session, 16.9% completed two sessions, and 12.3% completed three sessions; therefore, a total of 37% of clients completed between one and three sessions. Mean ranks were 983.58 for hybrid, 964.01 for online, and 951.36 for in-person. A Kruskal–Wallis test indicated no statistically significant differences in the number of sessions completed across treatment delivery modalities ($H(2, N = 1{,}933) = 1.49$, $p = .475$); therefore, post hoc pairwise comparisons were not conducted.

Treatment efficiency based on session count was calculated as the average pretreatment-posttreatment change per session by dividing the magnitude of pretreatment-posttreatment change by the total number of completed sessions. On average, each therapy session was associated with a 1.44-point increase in therapist-rated problem resolution ($SD = 1.30$), a 1.27-point increase in client-rated problem resolution ($SD = 1.04$), a 2.50-point decrease in psychological impairment ($SD = 2.08$), and a 1.69-point increase in overall well-being ($SD = 2.71$). Treatment efficiency distributions per session for all indicators are shown in Figure 1, Panels D–G.

Psychological impairment efficiency (skewness = −1.12, kurtosis = 2.25) and overall well-being efficiency (skewness = −0.03, kurtosis = 2.12) were normally distributed; therapist-rated problem resolution efficiency (skewness = 2.42, kurtosis = 8.41) and client-rated problem resolution efficiency (skewness = 2.75, kurtosis = 13.44) were non-normally distributed. Levene's test indicated variances were equal across treatment delivery modalities for psychological impairment efficiency ($F(2, 247) = 0.40$, $p = .674$) and overall well-being efficiency ($F(2, 273) = 1.42$, $p = .244$); therefore, a standard one-way ANOVA with Tukey post hoc tests was conducted. Results indicated no significant treatment delivery modality differences for psychological impairment efficiency ($F(2, 247) = 0.15$, $p = .859$) and overall well-being efficiency ($F(2, 273) = 2.94$, $p = .055$); however, pairwise Tukey post hoc tests indicated that efficiency per session for overall well-being was

significantly lower for the hybrid modality compared to in-person (mean difference = −0.90, 95% *CI* [−1.80, −0.01], $p = .047$), with no other statistically significant pairwise comparisons. Results from Kruskal–Wallis tests indicated no significant treatment delivery modality differences for therapist-rated problem resolution efficiency ($H(2, 1{,}931) = 5.82$, $p = .055$) and client-rated problem resolution efficiency ($H(2, 505) = 0.69$, $p = .710$). Post hoc power analyses revealed variable statistical power across efficiency comparisons, with adequate power (.60) for the small borderline overall well-being efficiency effect ($f = .15$, $p = .055$); the other session efficiency comparisons demonstrated negligible nonsignificant effect sizes (therapist-rated problem resolution efficiency: $f = .06$, $p = .055$); client-rated problem resolution efficiency: ($f = .04$, $p = .710$); psychological impairment efficiency: ($f = .04$, $p = .859$). These results suggest that modality differences in treatment efficiency, even if statistically significant with adequate power, would likely lack practical clinical significance given the negligible to small magnitude of between-modality differences.

Discussion

The purpose of this study was to evaluate the effectiveness, therapeutic process, and efficiency of BST delivered via in-person, online, and hybrid modalities in a naturalistic clinical setting. By leveraging a large, ecologically valid dataset and employing a modified available data analysis, we aimed to provide a comprehensive practice-based assessment of BST treatment indicators across diverse delivery modalities using both therapist- and client-reported measures. Findings from this practice-based study provide preliminary evidence that BST can be effective when delivered in-person, online, or in hybrid modalities, with substantial pretreatment-posttreatment improvements across modalities. Across treatment indicators, between-modality effects were negligible and nonsignificant or small, with a subset of small effects statistically significant; notably, no moderate or large differences emerged. These findings are especially relevant in the context of increasing adoption of online and hybrid therapy, and align with prior research indicating that some psychotherapies can be successfully adapted to a range of delivery methods (Barak et al., 2008; Greenwood et al., 2022).

Problem resolution ratings at posttreatment were generally high, with clients and therapists reporting mean scores of 6.71 and 5.76, respectively, on a 0–10 scale. Although clients and therapists differed statistically in their ratings, the effect size was negligible and likely of limited clinical importance. Notably, treatment delivery modality did not predict problem resolution from either the client or therapist perspective. Longer therapy duration was consistently associated with greater perceived problem resolution, though the number of sessions played a more nuanced role: therapists associated more sessions with higher resolution, whereas clients reported slightly lower resolution with more sessions, possibly reflecting greater treatment difficulty among clients requiring extended care. Demographic factors such as age and gender were not significant predictors. The modest proportion of explained variance in both client-rated (3%) and therapist-rated (14%) problem resolution models is consistent with broader psychotherapy outcomes research and likely reflects unmeasured client factors (e.g., motivation, dimensional baseline severity, contextual life changes, demographic characteristics) and therapist factors that influence treatment outcomes (Wampold & Imel, 2015) independently of delivery modality. Overall, these results suggest that the extent of problem resolution achieved in therapy is more strongly related to therapy duration than to delivery modality, with some nuanced differences in how clients and therapists interpret the impact of session count.

With regard to problem resolution clinical significance, therapists rated substantially fewer clients as achieving problem resolution compared to client self-reports (44.6% vs. 60.5%), which may reflect therapists' more conservative assessments of progress or could be influenced by the fact that many clients attended only a few therapy sessions, limiting therapists' observation of improvement trajectories. Therapist-rated assessments revealed a distinctive pattern whereby hybrid delivery appeared associated with more polarized outcomes (i.e., clients were more likely to be rated as either unresolved or resolved), whereas in-person clients were more frequently classified in the intermediate partial resolution category. Interestingly, this polarization pattern was not replicated in client self-reports, where it was online therapy recipients who were more likely to report partial resolution outcomes compared to the other treatment modalities, suggesting potential discrepancies in how clients and therapists perceive progress across different delivery modalities. Overall, problem resolution rates were broadly similar across modalities, particularly for client reports. The slight variations in rates across modalities may reflect measurement characteristics, sampling variability, or unmeasured differences in client or therapist characteristics, rather than systematic clinically meaningful differences in therapeutic effectiveness; however, given the study's observational design, these possibilities cannot be definitively distinguished.

The present study found that BST was associated with substantial improvements in psychological impairment and overall well-being, with effect sizes that ranged from very large to huge for impairment and from moderate to very large for well-being. These substantial gains were achieved over a relatively brief average of approximately six sessions over roughly five months, highlighting the efficiency of BST in facilitating meaningful change within a short timeframe. Improvements in psychological impairment were observed across all modalities and were not linked to duration, sessions, age, or gender, suggesting potential generalizability across modalities and client characteristics. Clinical significance analyses for psychological impairment confirmed that most clients reliably improved and achieved recovery, with no differences across treatment delivery modalities. Although overall well-being improved significantly for all clients and the time × modality interaction was nonsignificant, clinical significance analyses indicated that in-person clients were somewhat more likely to achieve reliable improvement and recovery in well-being; this categorical difference may reflect distributional features or subgroup variability rather than a systematic difference in improvement rates across modalities. Therapeutic alliance scores, examined as an indicator of positive therapeutic process, were high at both assessment points and showed no significant pretreatment-posttreatment difference. Given the two-timepoint design, this pattern indicates no detectable pre–post difference rather than confirming alliance stability across the full course of treatment. These findings suggest therapists were generally able to establish a strong alliance regardless of modality. Notably, longer therapy duration was marginally associated with lower alliance ratings, which may reflect gradual attenuation of the alliance in more protracted cases. Collectively, these results underscore the adaptability, effectiveness, and efficiency of BST in real-world clinical settings, and provide preliminary support for its use across multiple delivery platforms as an accessible and effective approach to care.

Across modalities, each month and each session of therapy were associated with improvements in problem resolution, psychological impairment, and overall well-being, though distributions were often non-normal and highly skewed. Importantly, efficiency did not significantly differ by treatment delivery modality for therapist- or client-rated problem resolution or psychological impairment; however, between-group comparison indicated that in-person therapy was significantly more efficient than hybrid therapy in improving overall well-being, with no significant efficiency differences observed between in-person and online or between hybrid and online modalities. These findings suggest that efficiency was largely comparable across delivery modalities, apart from overall well-being, where in-person therapy demonstrated a modest advantage over hybrid care.

Clinical implications

The present findings offer several important implications for providing BST in practice-based settings. First, the present findings suggest that BST can be delivered effectively and, in many cases, efficiently via in-person, online, and hybrid modalities in routine practice, allowing clinicians to flexibly align delivery modality with client needs and preferences, although formal non-inferiority was not tested. The generally high levels of problem resolution reported by both therapists and clients, coupled with the strong clinical gains observed over a relatively brief treatment duration, highlight the potency and time-efficiency of BST. The robust improvements in psychological impairment and overall well-being, regardless of delivery modality support the integration of technology-mediated approaches into routine practice. The absence of meaningful between-modality differences in therapeutic alliance suggests that a positive therapeutic process can be established across in-person, online, and hybrid BST, complementing the broader pattern of favorable clinical outcomes. The generally small and non-significant modality differences, aside from a modest in-person well-being advantage in clinical significance rates, indicate that clinicians can reasonably consider client circumstances, access, and preferences when selecting a modality, while recognizing that more rigorous comparative trials are still needed.

These results reinforce the adaptability of BST and support its continued use and dissemination across diverse service delivery platforms, particularly as the field embraces digital and hybrid models of care (Burgoyne & Cohn, 2020). Importantly, the demonstrated effectiveness and efficiency of BST across all modalities suggests that technology-mediated therapy can play a critical role in expanding access and reducing barriers to care for underserved and marginalized minoritized groups, offering flexible treatment options that can be tailored to individual needs and potentially improving engagement and outcomes in populations historically facing disparities in mental health service utilization (Kafali et al., 2014).

Limitations

Despite the strengths and ecological validity of this large practice-based study, several limitations should be considered. First, although the uncontrolled and non-randomized naturalistic design

enhances generalizability to everyday clinical settings, it precludes causal inferences about the relative effectiveness or efficiency of BST across treatment modalities. The absence of a control condition does not allow estimating the extent to which the intervention accounts for obtained results. The lack of random assignment increases the risk of selection bias and unmeasured confounding, as clients were not randomly allocated to in-person, online, or hybrid modalities, and external variables such as therapist-client dynamics and environmental factors were not controlled (Verster et al., 2019). Furthermore, nonsignificant differences between treatment modalities should not be interpreted as definitive evidence of equivalence. Our findings suggesting comparable results across modalities are better characterized as preliminary evidence from comparative effectiveness research in naturalistic settings, warranting confirmation through adequately powered non-inferiority trials with pre-registered margins and appropriate sample sizes. Additionally, therapist effects were not accounted for, and clustering by therapist was not modeled, which may have influenced results.

The study's reliance on a modified available data analysis introduced further limitations. A substantial proportion (46%) of SYPRENE cases were excluded from analyses primarily because the phased introduction of indicator instruments during the study period led to missing data, raising the possibility of selection bias if excluded clients differed systematically from those retained. The phased introduction of indicator instruments across the study period also resulted in varying sample sizes for different analyses and fewer data for more recently added measures, potentially introducing measurement bias. The absence of data imputation preserved ecological validity but may have biased estimates if missingness was related to unobserved factors such as treatment dissatisfaction or premature discontinuation (Smith et al., 2014). Moreover, the voluntary nature of therapist participation and the addition of measures over time contributed to non-uniform data collection and further variability in analytic sample sizes.

The generalizability of the findings is limited by several factors. All participants resided in Western Europe (primarily France and Italy), restricting applicability to other cultural contexts. Race, ethnicity, and socioeconomic status data were not collected due to privacy regulations, precluding analysis of indicators by these demographic variables. Additionally, the online modality subgroup may have been differentially affected by the variability in timing of data collection relative to the COVID-19 pandemic, given that 31% of online clients initiated therapy before the pandemic compared with 77% to 78% of in-person and hybrid clients, respectively. Because BST is a brief and structured approach, findings may not generalize to longer-term psychotherapies.

Sample sizes varied substantially across indicators and treatment delivery modalities. Client-rated problem resolution had a considerably smaller subsample than therapist-rated problem resolution, and the online modality subgroup was consistently the smallest across all measures (Table S3). For the longitudinal indicators, only 250, 276, and 217 clients (approximately 11–14% of the 1,933 eligible SYPRENE clients) had complete psychological impairment, overall well-being, and therapeutic alliance data, respectively, with especially small online subsamples (i.e., 25 clients for psychological impairment, 57 clients for overall well-being, and 46 clients for therapeutic alliance); as a result, statistical power for detecting small between-modality effects in these analyses was low, and standard errors and confidence intervals for the online modality in particular were correspondingly large. However, all observed modality effect sizes were negligible or at the lower bound of the small-effect range ($f = .01–.16$) and power was adequate ($\geq .80$) to detect effect sizes at the upper bound of the small-effect range ($f > .18–.21$), suggesting that the observed between-modality effects across indicators in this sample are unlikely to reflect undetected moderate or large differences. Nevertheless, the possibility of small effects cannot be ruled out given the study's power and attrition limitations. Although nonsignificant findings always carry the possibility of Type II error, nonsignificant modality differences—especially for the longitudinal indicators—should therefore be interpreted as inconclusive rather than as evidence of equivalence. In addition, although observed between-modality effect sizes were nonsignificant, negligible, or small, these estimates may be biased due to non-random attrition and underrepresentation of early treatment discontinuation, which may attenuate or distort true between-modality differences. At the same time, the observed pattern and magnitude of effect sizes indicate that any underlying modality differences, if present, are likely to be negligible or small and of limited clinical importance. This interpretation is supported by the substantial and statistically significant within-modality treatment effects observed across all three delivery modalities (Table S4; Table S5). The magnitude of these within-modality improvements far exceeded the modest between-modality differences, suggesting that BST's effectiveness may not be substantially dependent on treatment delivery modality. Future research with larger, adequately powered, and balanced samples across modalities would strengthen confidence in these conclusions.

Although only a minority of clients had complete pre–post self-report data on psychological impairment, overall well-being, and therapeutic alliance—partly because these instruments were phased in over time within SYPRENE—attrition analyses indicated that clients with available psychological impairment and client-rated problem resolution data were modestly more treatment-engaged (approximately one to two additional sessions and one to three extra months) and substantially less likely to discontinue early than those without, suggesting some selection effects for these indicators, such that these subsamples may under-represent clients who discontinue very early. In contrast, clients with complete overall well-being and therapeutic alliance data did not differ meaningfully from those without in terms of sessions, duration, or early discontinuation. Because BST is designed as a brief intervention, some clients may discontinue after relatively few sessions due to sufficient improvement, so early discontinuation in this sample may reflect a mixture of premature discontinuation and successful early termination rather than uniformly poor outcomes or therapeutic alliance. However, the small size of all longitudinal subsamples still warrants conservative interpretation of pre–post self-report findings and indicates that these longitudinal estimates generalize most clearly to clients who remain in treatment beyond the earliest sessions, highlighting the need for future research with more systematic treatment process and outcome monitoring.

Measurement limitations were also present. Therapist-reported clinical issue data were missing for a notable portion of clients, and client-reported indicators may be biased toward those who remained in therapy longer, as clients who prematurely discontinued therapy were less likely to complete posttreatment measures. There was no formal assessment of treatment fidelity or adherence to the BST model, so variability in how BST was delivered across therapists is unknown.

Statistical and analytic limitations should also be noted. The therapeutic alliance, therapy duration, and therapy sessions variables exhibited non-normal distributions. Therapeutic alliance data were likely non-normally distributed due to commonly observed alliance ceiling effects, where client ratings cluster toward the top of the scale early in treatment (Meier & Feeley, 2021). Although non-parametric analyses addressed non-normality in treatment process and efficiency indicators, ceiling effects and extreme distributions may have reduced statistical power to detect modality differences, particularly for brief treatments with substantial variability in duration. Treatment indicator triangulation across multiple informants and domains provided robustness to instrument-specific missingness but did not resolve inherent pretest-posttest design limitations. With clinical problem resolution based on a single timepoint and longitudinal indicators on two timepoints, analyses cannot address session-by-session dynamics of symptom change or alliance development. Phased instrument introduction and evolving administration schedules precluded session-by-session growth-curve modeling, as restricting to clients with dense longitudinal data would substantially reduce sample size and introduce systematic bias.

Finally, the regression models explained only a small proportion of variance in client- (2.8%) and therapist-rated (14.0%) problem resolution, suggesting that additional clinically relevant factors not included in the models (e.g., initial symptom severity, client motivation, therapist characteristics) likely contribute to variability in treatment outcomes. Therapist identity accounted for meaningful variance in therapist-rated problem resolution and posttreatment psychological impairment, suggesting that therapist effects should ideally be modeled. However, the naturalistic design resulted in substantial confounding between therapist identity and treatment delivery modality within the sample, with many therapists contributing cases to only a single modality. This design characteristic prevented the use of multilevel models to separate therapist effects from modality effects, as most therapists lacked sufficient within-therapist cross-modality variability in the current dataset to estimate reliable slopes. Consequently, observed modality differences may partially reflect therapist-level characteristics (e.g., training, practice setting, client selection) rather than pure effects of modality. Additionally, indicators were assessed only at pretreatment and posttreatment, with no long-term follow-up to determine the durability of treatment effects.

Recommendations for future research

Future research should employ rigorous designs—such as RCTs or propensity score matching—to strengthen causal inferences regarding BST effectiveness and efficiency across modalities. Longitudinal studies with extended follow-up are needed to assess the durability of treatment gains and whether modality influences outcome maintenance over time. Given the observed modest in-person advantage for well-being improvement, further investigation into underlying mechanisms is warranted, including client engagement and nonverbal communication. Future studies should identify client characteristics (e.g., clinical severity, technological comfort, sociodemographic factors) that predict better treatment outcomes and process in specific modalities and examine potential

moderating variables such as therapist experience, treatment fidelity, and contextual factors (e.g., pandemic-related stressors) that clarify when each modality is most effective. Research should also explore the optimal balance of in-person and online sessions in hybrid modalities. Additionally, distinguishing between clients who discontinue therapy following rapid improvement and those who discontinue prematurely without benefit would clarify implications for treatment efficiency indicators and clinical decision-making. Collecting comprehensive demographic data, including race, ethnicity, and socioeconomic status, will be critical for addressing disparities in treatment access, process, and outcomes. Finally, incorporating client and therapist satisfaction measures and qualitative feedback would enhance understanding of the subjective experience and acceptability of different modalities, informing personalized modality selection in routine practice.

Conclusion

This study suggests that, in a large naturalistic sample, BST can be delivered via online and hybrid modalities with clinical outcomes broadly similar in magnitude to those observed in traditional in-person therapy, alongside modest advantages for in-person care on overall well-being. Therapeutic alliance findings further suggest that a positive therapeutic process can be established across delivery formats. These findings are consistent with viewing technology-mediated delivery as a potentially viable option for many clients; however, more rigorous comparative and non-inferiority trials are needed before firm conclusions can be drawn about relative efficacy across modalities. Nevertheless, the present results indicate that integrating technology-mediated BST into clinical services may help broaden options for receiving therapy.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Declaration of generative AI and AI-assisted technologies

During the preparation of this work the authors used ChatGPT (Version 4.0) and Claude (3.5 Sonnet) to provide suggested revisions for writing clarity that were taken under advisement. The authors take full responsibility for the content of the publication.

ORCID

JEFFREY B. JACKSON http://orcid.org/0000-0002-9756-7841
RYTIS PAKROSNIS http://orcid.org/0000-0001-7817-1502
GRÉGOIRE VITRY http://orcid.org/0000-0002-1375-8860
ZOIE J. BARTLING http://orcid.org/0009-0006-4442-3365
LAURENT COLLONGUES http://orcid.org/0009-0001-5791-6170
PADRAIC GIBSON http://orcid.org/0009-0003-8474-4153
CLAUDETTE PORTELLI http://orcid.org/0009-0008-4865-0471
MATTEO PAPANTUONO http://orcid.org/0009-0001-7984-0453

References

Aafjes-van Doorn, K., Békés, V., & Luo, X. (2021). COVID-19 related traumatic distress in psychotherapy patients during the pandemic: The role of attachment, working alliance, and therapeutic agency. *Brain Sciences*, *11*(10), 1288. https://doi.org/10.3390/brainsci11101288

Aafjes-van Doorn, K., Békés, V., Stukenberg, K. W., & Roberts, K. E. (2023). Therapists' views and recommendations on the ins and outs of practicing teletherapy. *Counselling Psychology Quarterly*, *37*(3), 1–21. https://doi.org/10.1080/09515070.2023.2227133

Adam, J., Goletz, H., Viefhaus, P., Woitecki, K., & Doepfner, M. (2022). Webcam-based online coaching with children and adolescents with obsessive-compulsive disorders–A single-case study. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, *51*(3), 207–221. https://doi.org/10.1024/1422-4917/a000904

Babiano-Espinosa, L., Wolters, L. H., Wieidle, B., Compton, S. N., Lydersen, S., & Skokauskas, N. (2021). Acceptability and feasibility of enhanced cognitive behavioral therapy (ECBT) for children and adolescents with obsessive–compulsive disorder. *Child and Adolescent Psychiatry and Mental Health*, *15*(1), 1–11. https://doi.org/10.1186/s13034-021-00400-7

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. A. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services*, *26*(2-4), 109–160. https://doi.org/10.1080/15228830802094429

Bartle-Haring, S., Walsh, L., Blalock, J., & Bryant, A. (2021). Using therapeutic alliance to predict treatment attendance among couples. *Contemporary Family Therapy*, *43*(4), 370–381. https://doi.org/10.1007/s10591-021-09602-9

Benz, C., Norman, R., Hendrie, D., Welsh, M., Huntley, M., & Robinson, S. (2022). Use of teletherapy for allied health interventions in community-based disability services: A scoping review of user perspectives. *Health & Social Care in the Community*, *30*(6), e3934–e3948. https://doi.org/10.1111/hsc.14105

Bradford, A. B., Johnson, L. N., Anderson, S. R., Banford-Witting, A., Hunt, Q. A., Miller, R. B., & Bean, R. A. (2023). Call me maybe? In-person vs. Teletherapy outcomes among married couples. *Psychotherapy Research*, *34*(5), 611–625. https://doi.org/10.1080/10503307.2023.2256465

Bringhurst, D. L., Watson, C. S., Miller, S. D., & Duncan, B. L. (2006). The reliability and validity of the outcome rating scale: A replication study of a brief clinical measure. *Journal of Brief Therapy*, *5*(1), 23–29.

Burgoyne, N., & Cohn, A. S. (2020). Lessons from the transition to relational teletherapy during COVID-19. *Family Process*, *59*(3), 974–988. https://doi.org/10.1111/famp.12589

Campbell, A., & Hemsley, S. (2009). Outcome rating scale and session rating scale in psychological practice: Clinical utility of ultra-brief measures. *Clinical Psychologist*, *13*(1), 1–9. https://doi.org/10.1080/13284200802676391

Cochrane, A. L. (1972). *Effectiveness and efficiency: Random reflections on health services*. Nuffield Provincial Hospitals Trust, p. 92.

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

Connolly, S. L., Miller, C. J., Lindsay, J. A., & Bauer, M. S. (2020). A systematic review of providers' attitudes toward telemental health via videoconferencing. *Clinical Psychology: Science and Practice*, *27*(2), e12311. https://doi.org/10.1111/cpsp.12311

Conroy, K., Kehrer, S. M., Georgiadis, C., Hare, M., Ringle, V. M., & Shaw, A. M. (2024). Learning from adolescents and caregivers to enhance acceptability and engagement within virtual dialectical behavior therapy for adolescents skills groups: A qualitative study. *Child Psychiatry & Human Development*, *56*(5):1418–1432. https://doi.org/10.1007/s10578-023-01641-7

Cronin, I., Judson, A., Ekdawi, I., Verma, G., Baum, S., Grant, P., Harrison-Rowe, H., Parker, J., Ramsey, B., & Nicholson, J. (2021). Holding onto the 'mystery' within online family and systemic therapy. *Journal of Family Therapy*, *43*(2), 295–313. https://doi.org/10.1111/1467-6427.12330

Davis, K. A., Zhao, F., Janis, R. A., Castonguay, L. G., Hayes, J. A., & Scofield, B. E. (2023). Therapeutic alliance and clinical outcomes in teletherapy and in-person psychotherapy: A non-inferiority study during the COVID-19 pandemic. *Psychotherapy Research*, *34*(5), 589–600. https://doi.org/10.1080/10503307.2023.2229505

Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D. (2003). The session rating scale: Preliminary psychometric properties of a "working" alliance measure. *Journal of Brief Therapy*, *3*(1), 3–12.

Dunn, K., & Wilson, J. (2021). When online and face to face counseling work together: Assessing the impact of blended or hybrid approaches, where clients move between face-to-face and online meetings. *Person-Centered & Experiential Psychotherapies*, *20*(4), 312–326. https://doi.org/10.1080/14779757.2021.1993970

Erasmus, P., Borrmann, M., Becker, J., Kuchinke, L., & Meinlschmidt, G. (2025). Comparing digital versus face-to-face delivery of systemic psychotherapy interventions: Systematic review and meta-analysis of randomized controlled trials. *Interactive Journal of Medical Research*, *14*, e46441. https://doi.org/10.2196/46441

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*(1), 43–53. https://doi.org/10.1037/a0014699

Giovanetti, A. K., Punt, S. E., Nelson, E., & Ilardi, S. S. (2022). Teletherapy versus in-person psychotherapy for depression: A meta-analysis of randomized controlled trials. *Telemedicine and E-Health*, *28*(8), 1077–1089. https://doi.org/10.1089/tmj.2021.0294

Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire*. Oxford University Press.

Goldberg, D., & Williams, P. (1991). *A user's guide to the general health questionnaire*. NFER-Nelson.

Greenwood, H., Krzyzaniak, N., Peiris, R., Clark, J., Scott, A. M., Cardona, M., Griffith, R., & Glasziou, P. (2022). Telehealth versus face-to-face psychotherapy for less common mental health conditions: Systematic review and meta-analysis of randomized controlled trials. *JMIR Mental Health*, *9*(3), e31780. https://doi.org/10.2196/31780

Griffith, B., Archbold, H., Sáez Berruga, I., Smith, S., Deakin, K., Cogan, N., Tanner, G., & Flowers, P. (2023). Frontline experiences of delivering remote mental health supports during the COVID-19 pandemic in Scotland: Innovations, insights and lessons learned from mental health workers. *Psychology, Health & Medicine*, *28*(4), 964–979. https://doi.org/10.1080/13548506.2022.2148698

Heiden-Rootes, K., Ferber, M., Meyer, D., Zubatsky, M., & Wittenborn, A. (2021). Relational teletherapy experiences of couple and family therapy trainees: "Reading the room," exhaustion, and the comforts of home. *Journal of Marital and Family Therapy*, *47*(2), 342–358. https://doi.org/10.1111/jmft.12486

Hensel, J. M., Yang, R., Vigod, S. N., & Desveaux, L. (2021). Videoconferencing at home for psychotherapy in the postpartum period: Identifying drivers of successful engagement and important therapeutic conditions for meaningful use. *Counselling and Psychotherapy Research*, *21*(3), 535–544. https://doi.org/10.1002/capr.12372

Hjelle, K. M., Eide, T. O., Thorsen, A. L., Kvale, G., Hagen, K., Snorrason, I., Björgvinsson, T., & Hansen, B. (2023). The bergen 4-day treatment for panic disorder: Adapting to COVID-19 restrictions with a hybrid approach of face-to-face and videoconference modalities. *BMC Psychiatry*, *23*(1), 570. https://doi.org/10.1186/s12888-023-05062-7

International Committee of Medical Journal Editors. (2024). *Clinical trials registration*. https://www.icmje.org/about-icmje/faqs/clinical-trials-registration/

Jackson, J. B., Pietrabissa, G., Rossi, A., Manzoni, G. M., & Castelnuovo, G. (2018). Brief strategic therapy and cognitive behavioral therapy for women with binge eating disorder and comorbid obesity: A randomized clinical trial one-year follow-up. *Journal of Consulting and Clinical Psychology*, *86*(8), 688–701. https://doi.org/10.1037/ccp0000313

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Kafali, N., Cook, B., Canino, G., & Alegria, M. (2014). Cost-effectiveness of a randomized trial to treat depression among Latinos. *The Journal of Mental Health Policy and Economics*, *17*(2), 41–50. https://pubmed.ncbi.nlm.nih.gov/25153092

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576–2590. https://doi.org/10.3758/s13428-021-01587-5

Lesage, F.-X., Martens-Resende, S., Deschamps, F., & Berjot, S. (2011). Validation of the General Health Questionnaire (GHQ-12) adapted to a work-related context. *Open Journal of Preventive Medicine*, *1*(2), 44–48. https://doi.org/10.4236/ojpm.2011.12007

Lin, T., Heckman, T. G., & Anderson, T. (2022). The efficacy of synchronous teletherapy versus in-person therapy: A meta-analysis of randomized clinical trials. *Clinical Psychology: Science and Practice*, *29*(2), 167–178. https://doi.org/10.1037/cps0000056

Little, J. R. (2024). Missing data analysis. *Annual Review of Clinical Psychology*, *20*, 149–173. https://doi.org/10.1146/annurev-clinpsy-080822-051727

McLean, S. A., Booth, A. T., Schnabel, A., Wright, B. J., Painter, F. L., & McIntosh, J. E. (2021). Exploring the efficacy of telehealth

for family therapy through systematic, meta-analytic, and qualitative evidence. *Clinical Child and Family Psychology Review*, *24*(2), 244–266. https://doi.org/10.1007/s10567-020-00340-2

Meier, S., & Feeley, T. (2021). Investigation of causes of ceiling effects on working alliance measures. *Frontiers in Psychology*, *13,* 949326. https://doi.org/10.3389/fpsyg.2022.949326

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Miller, S. D., & Duncan, B. L. (2004). *The outcome and session rating scales: Administration and scoring manual*. Institute for the Study of Therapeutic Change.

Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, *2*(2), 91–100.

Miller, S. D., Hubble, M. A., & Chow, D. (2020). *Better results: Using deliberate practice to improve therapeutic effectiveness*. APA Books.

Miu, A. S., Vo, H. T., Palka, J. M., Glowacki, C. R., & Robinson, R. J. (2023). Teletherapy with serious mental illness populations during COVID-19: Telehealth conversion and engagement. In *How the COVID-19 pandemic transformed the mental health landscape* (pp. 63–80). Routledge.

Murphy, M. G., Rakes, S., & Harris, R. M. (2020). The psychometric properties of the session rating scale: A narrative review. *Journal of evidence-based social work*, *17*(3), 279–299. https://doi.org/10.1080/26408066.2020.1729281

Nardone, G., & Portelli, C. (2005). *Knowing through changing: The evolution of brief strategic therapy*. Crown House Pub Limited.

Nardone, G., & Watzlawick, P. (2005). *Brief strategic therapy: Philosophy, techniques, and research*. Jason Aronson.

Piccinelli, M., Bisoffi, G., Bon, M., Cunico, L., & Tansella, M. (1993). Validity and test-retest reliability of the Italian version of the 12-item general health questionnaire in general practice: A comparison between three scoring methods. *Comprehensive Psychiatry*, *34*(3), 198–205. https://doi.org/10.1016/0010-440X(93)90048-9

Sánchez-López, M. d. P., & Dresch, V. (2008). The 12-item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. *Psicothema*, *20*(4), 839–843.

Sawilowsky, S. S. (2009). Effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Seidel, J., & Miller, S. D. (2012). Manual 4: Documenting change: A primer on measurement, analysis, and reporting. In B. Bertolino & S. D. Miller (Eds.), *International Center for Clinical Excellence manuals on feedback-informed treatment* (pp. 1–41). International Center for Clinical Excellence.

Smith, A. K., Ayanian, J. Z., Covinsky, K. E., Landon, B. E., McCarthy, E. P., Wee, C. C., & Steinman, M. A. (2014). Secondary analysis of existing data: Opportunities and implementation. *Swiss Medical Weekly*, *144*, w14025. https://doi.org/10.4414/smw.2014.14025

Sucala, M., Schnur, J. B., Constantino, M. J., Miller, S. J., Brackman, E. H., & Montgomery, G. H. (2012). The therapeutic relationship in e-therapy for mental health: A systematic review. *Journal of Medical Internet Research*, *14*(4), e110. https://doi.org/10.2196/jmir.2084

Verster, J. C., van Rossum, C. J., & Scholey, A. (2019). Advantages and limitations of naturalistic study designs and their implementation in alcohol hangover research. *Journal of Clinical Medicine*, *8*(12), 2160. https://doi.org/10.3390/jcm8122160

Vitry, G., Duriez, N., Lartilleux-Suberville, S., Pakrosnis, R., Beau, A., Garcia-Rivera, T., Brosseau, O., Vargas Avalos, P., Bardot, E., & Ray, W. A. (2020). Introducing SYPRENE: An international practice research network for strategic and systemic therapists and researchers. *Family Process*, *59*(4), 1628–1639. https://doi.org/10.1111/famp.12520

Vitry, G., Pakrosnis, R., Brosseau, O. G., & Duriez, N. (2021). Effectiveness and efficiency of strategic and systemic therapy in naturalistic settings: Preliminary results from a Systemic Practice Research Network (SYPRENE). *Journal of Family Therapy*, *43*(4), 516–537. https://doi.org/10.1111/1467-6427.12343

Vitry, G., Pakrosnis, R., Jackson, J. B., Gallin, E., & Hoyt, M. F. (2024). Problem resolution scale: A single-item instrument for easily assessing clinical improvement. *Journal of Marital and Family Therapy*, *50*(2), 477–494. https://doi.org/10.1111/jmft.12690

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). Routledge.

Werneke, U., Goldberg, D. P., Yalcin, I., & Ustün, B. T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine*, *30*(4), 823–829. https://doi.org/10.1017/S0033291799002287

Wootton, B. M. (2016). Remote cognitive–behavior therapy for obsessive–compulsive symptoms: A meta-analysis. *Clinical Psychology Review*, *43*, 103–113. https://doi.org/10.1016/j.cpr.2015.10.001