This is a repository copy of *Using machine learning algorithms to predict the effects of change processes in psychotherapy: Toward process-level treatment personalization*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/207853/

Version: Accepted Version

Article:

Gómez Penedo, J.M. orcid.org/0000-0001-7304-407X, Rubel, J., Meglio, M. et al. (10 more authors) (2023) Using machine learning algorithms to predict the effects of change processes in psychotherapy: Toward process-level treatment personalization. Psychotherapy, 60 (4). pp. 536-547. ISSN 0033-3204

https://doi.org/10.1037/pst0000507

© American Psychological Association, 2023. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: https://doi.org/10.1037/pst0000507

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

© 2023, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/pst0000507

Using machine learning algorithms to predict the effects of change processes in psychotherapy: Towards process-level treatment personalization

Juan Martín Gómez Penedo[1,2], Julian Rubel[3], Manuel Meglio[1], Leo Bornhauser[4], Tobias Krieger[4], Anna Babl[5], Roberto Muiños[1], Andrés Roussos[2,6], Jaime Delgadillo[7], Christoph Flückiger[8,9], Thomas Berger[4], Wolfgang Lutz[10] & Martin grosse Holtforth[4,11]

[1]Universidad de Buenos Aires, Buenos Aires, Argentina

[2]CONICET, Buenos Aires, Argentina

[3]Osnabrück University, Osnabrück, Germany

[4]University of Bern, Bern, Switzerland

[5]Adelphi University, New York, United States of America

[6]IPEHCS/CONICET-UNCo, Neuquén, Argentina

[7]University of Sheffield, Sheffield, United Kingdom

[8]University of Zurich, Zurich, Switzerland

[9]University of Kassel, Kassel, Germany

[10]University of Trier, Trier, Germany

[11]University Hospital Inselspital, Bern, Switzerland

Correspondence concerning this article should be addressed to Juan Martín Gómez Penedo, Address: *Lavalle 2353*, 1052 Buenos Aires, Argentina. E-mail: jmgomezpenedo@gmail.com Phone: +54 9 11 6525 7053.

This study was supported by a Seed grant from the Leading House for the Latin American Region (Universität St.Gallen, Switzerland) (Grants: No. SMG20l6, PIs: Juan Martín Gómez Penedo and Martin grosse Holtforth).

Abstract

This study aimed to develop and test algorithms to determine the individual relevance of two psychotherapeutic change processes (i.e., mastery and clarification) for outcome prediction. We measured process and outcome variables in a naturalistic outpatient sample treated with an integrative treatment for a variety of diagnoses (n=608) during the first ten sessions. We estimated individual within-patient effects of each therapist-evaluated process of change on patient-evaluated subsequent outcomes on a session-by-session basis. Using patients baseline characteristics, we trained machine learning algorithms on a randomly selected subsample (n= 407) to predict the effects of patients' process variables on outcome. We subsequently tested the predictive capacity of the best algorithm for each process on a hold-out subsample (n= 201). We found significant within-patient effects of therapist perceived mastery and clarification on subsequent outcome. In the hold-out subsample, the best-performing algorithms resulted in significant but small-to-medium correlations between the predicted and observed relevance of therapist perceived mastery (r=.18) and clarification (r=.16). Using the algorithms to create criteria for individual recommendations, in the hold-out sample we identified patients for whom mastery (14%) or clarification (18%) were indicated. In the mastery-indicated group, a greater focus on mastery was moderately associated with better outcome (r=.33, d=.70), while in the clarification-indicated group the focus was not related to outcome (r=-.05, d=.10). Results support the feasibility of performing individual predictions regarding mastery process relevance that can be useful for therapist feedback and treatment recommendations. However, results will need to be replicated with prospective experimental designs.

Keywords: machine learning; outcome prediction; processes and mechanisms of change; mastery and clarification, algorithm

Using machine learning algorithms to predict the effects of change processes in psychotherapy: Towards process-level treatment personalization

Although psychotherapy is an effective treatment for a number of mental disorders, not all patients respond successfully (Barkham & Lambert, 2021; Cuijpers et al., 2021). It is estimated that between 60% and 69% of treated patients improve during treatment, while the rest do not benefit or even deteriorate during therapy (Barkham & Lambert, 2021). Thus, there is still considerable room for improvement in psychotherapy outcomes.

Even though *bona fide* psychotherapies have been shown to produce fairly equivalent results across various mental health conditions (i.e., Cuijpers et al., 2021; Podina et al., 2019; Wampold & Imel, 2015), recent empirical studies have observed differential treatment effects based on patient baseline characteristics (e.g., Cohen et al., 2020; Gomez Penedo et al., 2019; Newman et al., 2017). In the field of depression, for example, there is meta-analytic evidence of heterogeneity of treatment response (Kaiser et al., 2022). This means that although - on average - patients will have similar responses to different treatments, individual patients might benefit more from one treatment than another (Friedl et al., 2020).

To enhance outcomes in psychotherapy and taking into account the emerging evidence of treatment response heterogeneity, there has been an increased focus on the development and implementation of evidence-based personalized treatment strategies in the last decade (Cohen et al., 2021; Delgadillo & Lutz, 2020). These approaches aim to optimize psychotherapy by identifying clinically relevant patient characteristics and making individually tailored treatment recommendations based on these characteristics (Lutz et al., 2019). Using feedback systems, this information could be provided to the therapists before therapy starts, supporting them in the decision-making process for each individual treatment (Lutz et al., 2021).

In recent years, the application of machine learning (ML) algorithms as an alternative to more traditional statistical methods (e.g., general linear model) is increasingly being used to develop tools for psychotherapy personalization (Delgadillo & Lutz, 2020). ML is a promising way to achieve more accurate clinical prediction models based on large datasets (Rutledge et al., 2019). This is particularly interesting because it is a statistical analysis methodology that adjusts its predictions adaptively and flexibly to data with complex patterns (e.g., with a lot of between-patient variability) and large numbers of variables with potential multicollinearity (Friedman et al., 2010; B. Schwartz et al., 2021). Some ML learning methods are also capable of modelling nonlinear associations between variables in a data-driven way, without a need for a-priori specification of expected relationships. Furthermore, ML strategies apply methods to minimize overfitting issues (i.e., enhancing generalizability), by using techniques such as resampling and regularization (e.g., Delgadillo, 2021).

In the field of psychotherapy personalization, the use of ML techniques has grown rapidly particularly for treatment selection (Aafjes-van Doorn et al., 2021). In this area, algorithms have been developed to model differential effects of at least two different treatments (e.g., cognitive-behavioral therapy [CBT] versus psychodynamic therapy), identifying the optimal therapy for each individual patient (see e.g., Cohen et al., 2020; Friedl et al., 2020; Schwartz et al., 2021). Although this approach has shown promising findings, pushing forward the field of treatment personalization, it also has some relevant shortcomings when translating this knowledge into clinical practice (Lorenzo-Luaces et al., 2021). One of the main limitations is that, in order to implement these algorithms in practice, it will be necessary to have clinicians trained in different *bona fide* treatments (e.g., CBT and psychodynamic therapy) or outpatient clinics with professionals from the different theoretical orientations.

As an alternative to treatment selection, some studies have used ML to predict the differential effects of trans-theoretical therapeutic change processes instead of ‘treatment packages’. Identifying the more fine-grained processes that lead to individual change would allow clinicians to target the psychotherapy processes that are best suited to the particularities of each patient (e.g., Rubel et al., 2017) independently of the theoretical framework used by the psychotherapist (Rubel et al., 2020). Developing algorithms at the trans-theoretical process level could potentially help to generate methods of personalization that are easier to implement in routine care.

One of the most well-known frameworks to conceptualize trans-theoretical process of change (also known as mechanisms of change) in therapy is the model developed by Klaus Grawe (1997, 2004). Besides the well-established process of the alliance (Flückiger et al., 2018), grounded on a meta-analysis on empirical findings Grawe (1997) identified four additional processes of change: [1] problem actuation (i.e., patients experiencing their problems in session); [2] resource activation (i.e., patients experiencing themselves as someone with resources and strengths to cope with their problems); [3] mastery (i.e., the patients' ability to cope with their problems); and [4] clarification (i.e., the patients' understanding of the sources and consequences of his or her own problematic behaviors and experiences). From these four change processes, mastery and clarification have been the two with strong trans-theoretical (e.g., Allemand & Flückiger, 2017) as well as empirical attention in recent years, with several studies showing that they are significantly associated with psychotherapy outcome (i.e., Gómez Penedo et al., 2022, 2023; Rubel et al., 2017; C. Schwartz et al., 2018).

When predicting process-outcome associations using ML, the first attempts focused on predicting therapeutic alliance effects on outcome (Rubel et al., 2020; Zilcha-Mano et al., 2018). Zilcha-Mano et al. (2018) used ML to identify interpersonal patient characteristics that differentially predicted the effects of alliance on outcome in CBT. An overall greater level of alliance was associated with better outcome in patients with lower problems of being overly cold and higher problems of being exploitable in their relationships. Furthermore, improvements of alliance during treatment were more strongly associated with better outcome when patients had higher problems of being overly cold and lower problems of being intrusive. In another study, Rubel et al. (2020) used ML algorithms to predict alliance effects in CBT based on patients baseline sociodemographic and clinical characteristics. Although the model found 11 reliable predictors (e.g., overall symptom severity, depression severity, interpersonal distress, emotional functioning), in the hold-out sample the correlations between the observed and estimated (i.e., by the algorithm) alliance-outcome effects of the algorithms ranged from r = -.07 to r = .05.

Besides these initial efforts predicting alliance-outcome associations using ML, some studies also tried to predict clarification and mastery effects in psychotherapy (e.g., Gómez Penedo et al., 2022; Lutz et al., 2019, 2022). Considering the high correlations between clarification and mastery when self-reported by patients (r= .65; Flückiger et al., 2010), as one solution proposed in the literature (Rubel et al., 2017), these studies had merged mastery and clarification processes into a broader intrapersonal change process named *problem coping experiences* (i.e., the degree in which the patient works on finding solutions to their problems or understanding their sources). Based on this strategy and using the nearest neighbors methodology on a naturalistic sample, Lutz et al. (2019, 2022) created and prospectively tested an algorithm to determine the most suitable treatment strategy for each patient comparing an alliance building approach, a problem coping focused approach, and a mixed approach. Results showed that when therapists followed the algorithmic recommendations during the first 10 sessions, effect sizes were better than if they did not follow the respective recommendation (~ d = 0.30). Additionally, Gómez Penedo et al. (2022) applied different ML algorithms using patients' baseline characteristics to predict problem coping experience effects on outcome during the first ten sessions of therapy. The best fitting algorithm in the training set presented adequate and stable results in the test set (R^2 =.15, d = 0.84)[1].

Going one step further, developing process-specific algorithms for mastery and clarification effects separately, rather than for problem coping experiences (i.e., a broader

[1] Note that an R^2 = .138 would be the minimum R^2 associated with a large effect size following Cohen's d criteria (i.e., d = .80) (J. Cohen, 1988).

construct that includes both), might provide more fine-grained evidence regarding what processes would be more suitable for each individual patient. This information could enhance personalized psychotherapy evidence-based recommendations, creating practical knowledge more easily to implement within naturalistic settings, compared to treatment selection approaches.

In order to build separate algorithms for mastery and clarification, an alternative to addressing the above-mentioned constructs' collinearity when self-reported by the patient, would be to use the therapists' perceptions of the therapy processes oriented towards mastery and clarification (Gómez Penedo et al., 2023). When analyzing therapist reports, correlations between these processes have been considerably lower (i.e., r=.11). However, using therapist reports, rather than evaluating patient in-session experiences, implies analyzing therapists' self-perceptions regarding the degree to which they have used interventions aiming to foster mastery and clarification in the patient. Besides allowing to deal with the collinearity issue, analyzing and predicting the effects of therapist level actions might provide more actionable clinical evidence than previous studies in this direction. The information based on therapist self-perception of their actions could be implemented in clinical practice in a more direct way than other process-outcome predictions (e.g., algorithms predicting alliance) by providing therapists with feedback about how they might need to intervene to maximize the likelihood of positive outcome for the individual patient.

In this context, the aim of this study was to develop ML algorithms to perform individual predictions regarding the effects of the trans-theoretical therapy change processes of mastery and clarification as measured by therapist reports (i.e., their perception of using interventions that target mastery and clarification). As a first step, we examined if both mastery and clarification therapy change processes during the first ten sessions of treatment significantly predicted subsequent treatment outcomes in terms of patients overall well-being. As a second step, we developed ML algorithms to predict those individual effects using patients' baseline information and tested them on a hold-out sample. Finally, we created and evaluated therapist recommendation criteria derived from the algorithms trained. Our study focused on the first ten sessions of therapy, following previous efforts to enhance the implementation of pre-treatment recommendations (Gómez Penedo et al., 2022; Lutz et al., 2019, 2022).

Methods

This retrospective cohort study was based on the analysis of archival data routinely collected at the outpatient clinic of the University of Bern, Switzerland. The data were gathered from the years 2001 to 2011 as part of the regular procedures of the clinic. The Ethics Committee of the Canton of Bern approved the use of routine assessment data for research (KEK 139/15). All participants signed a written informed consent form. Full methods of the study were documented at the open science framework (OSF; osf.io/mv7kd) (Gómez Penedo, Meglio, et al., 2022).

Participants

To be included in the sample, patients needed to have a maximum of 30% of missing data in the targeted baseline, process, and outcome measures. After excluding cases without enough information, 608 patients were included in the sample. Patients were mostly females (59%) with a mean age of 35.65 years (*SD*= 12.22 years). The most frequent diagnoses in the sample were depressive disorders (37.3%), followed by anxiety disorders (26.8%), and adjustment disorders (9.2%). Full diagnostic information is presented in Table 1SM in the supplemental material.

Treatment

The patients included in the study received psychotherapeutic treatment at the outpatient clinic of the University of Bern. Patients were routinely treated with an integrative cognitive behavioral therapy (CBT) developed by Klaus Grawe (2004) known as

Psychological Therapy. This treatment approach integrates empirically supported interventions from humanistic, systemic, and emotion-focused approaches with cognitive-behavioral interventions following explicit individual case formulations within an overarching theoretical framework regarding human functioning and mechanisms of change in psychotherapy (Caspar et al., 2023; Caspar & grosse Holtforth, 2010; Grawe et al., 1990). The overarching theoretical framework is consistency theory (Grawe, 2004) and case formulations are made using Plan Analysis (Caspar, 2022). According to consistency theory, inconsistency in human functioning (i.e., the tension resulting from discrepancies between needs and reality as well as from internal conflicts) centrally contributes to the development and maintenance of mental disorders and problems (Grawe, 2004). Individual case formulations based on Plan Analysis focus on means-end relationships within intra- and interpersonal functioning, as well as on the therapeutic relationship (Caspar, 2022). In Plan Analysis, therapists analyse the patient's verbal and nonverbal behavior and infer the instrumentality of behavior and experience, i.e., hypothetically underlying unconscious and conscious Plans. Based on Plan Analysis, case formulations explicate hypotheses about the individual etiology of the development and maintenance of patient problems and symptoms, including factors leading to inconsistency, patients' (lacking) resources (e.g., strengths, preferences, favorable circumstances, etc.), as well as problems and potentials for the therapeutic relationship. In the treatment plan, the therapist explicates ways to implement transdiagnostic change factors, i.e., mastery, clarification, resource activation, and problem activation (Grawe, 1997; Probst et al., 2018). In this integrative CBT approach the therapist is free to choose between a whole range of standard CBT interventions as described in manuals as well as select other empirically supported interventions, as long as they are compatible with the requirements of the individual case and the therapeutic relationship (Grawe et al., 1990; grosse Holtforth et al., 2008, 2011).

Effectiveness of this integrative CBT has been confirmed in a randomized controlled trial (Grawe et al., 1990). Another trial that compared this integrative therapy with traditional cognitive behavioral therapy, demonstrated similar results, but superior effects in patients with higher levels of symptom distress (grosse Holtforth et al., 2011).

Measures

Baseline measures

Symptom severity. To measure symptom severity, we used the Brief Symptom Inventory (BSI; Derogatis, 1993). The BSI is a 53-item self-reported measure that assess nine different dimensions of psychological distress (i.e., anxiety, depression, hostility, phobic anxiety, somatization, obsessive-compulsive, interpersonal sensitivity, paranoid ideation, and psychoticism). Items are rated on a 5-point Likert-like scale ranging from 0 (*never*) to 4 (*many times*). Higher scores represent greater severity. The German version of the BSI used in this study presented adequate psychometric properties (Franke, 2000). In the current sample, the BSI showed good internal consistency with αs ranging from .66 (i.e., hostility symptoms and psychoticism symptoms) to .85 (i.e., depressive symptoms).

Interpersonal problems. To measure interpersonal difficulties we used the Inventory of Interpersonal Problems 64 (IIP-64; Horowitz, Alden, Wiggins et al., 2000). The IIP-64 evaluates interpersonal excesses (i.e., “I do too much X”) and inhibitions (i.e., “I have a hard time doing X”). The items are self-reported on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*extremely*). The items of the IIP-64 are organized in eight subscales representing different types of interpersonal problems (i.e., domineering, intrusive, overly nurturant, exploitable, submissive, socially avoidant, cold, and vindictive). The subscales enable the calculation of two meaningful (orthogonal) interpersonal dimensions of communion (i.e., the degree in which someone seeks close relationships, ranging from detached to overly nurturant behaviours) and agency (i.e., the degree in which someone seeks to be dominated by or to dominate others, ranging from submissive to domineering behaviours). Furthermore, an overall interpersonal distress index can be computed based on the sum score of all the items, with higher scores representing greater interpersonal distress. The German version of the IIP-64 is psychometrically valid and reliable (Horowitz et al., 2000). In the sample of the study, the overall interpersonal distress showed adequate internal consistency (α= .93).

Emotion regulation. We used the Emotion Regulation Questionnaire (EMOREG; Znoj & Grawe, 2000), a 24-item self-report scale which measures the adaptive outcomes of well-functioning emotion regulation (i.e., different coping strategies). The items are rated on a 6-point Likert-type rating scale ranging from 1 (very untrue of me) to 6 (very true of me) and are organized into two subscales of adaptive forms of emotion management (emotional expression and emotional self-control; e.g., "When experiencing emotionally overwhelming situations, I often talk in depth about emotionally important topics") and two subscales of maladaptive forms of emotion regulation (emotional avoidance and emotional distort; e.g., "When experiencing emotionally overwhelming situations, often I am a person who moves and acts restlessly as a way of avoiding unpleasant thoughts and feelings"). The German version of the EMOREG showed good psychometric properties (Znoj & Grawe, 2000). In the present sample, the EMOREG showed adequate internal consistency ranging with alphas from .58 (e.g., emotional avoidance) to .82 (e.g., emotional self-control).

Change expectations and fears. To measure expectations of psychotherapy outcomes, we used the Patient Questionnaire on Therapy Expectation and Evaluation (PATHEV; Schulte, 2005). The PATHEV is an 11-item questionnaire rated on 5-point Likert-like scales ranging from 1 (not true at all) to 5 (absolutely true). This self-report scale has three subscales: Hope of recovery (e.g., "I believe my problems can finally be solved"), fear of change (e.g., "From time to time I worry about all the things that will change once my problems have vanished"), and treatment suitability (e.g., "I've found the right therapy").

Positive perception of an adequate treatment, together with the hope for recovery and less fear of change are indications that the patient has a high expectation for his or her treatment. As in this sample we used the PATHEV as a baseline (i.e., pre-treatment) measure, we did not include the items of treatment suitability in the assessment. The German version of the PATHEV is psychometrically valid and reliable (Schulte, 2005). In the current sample, the PATHEV presented adequate internal consistency on Hope for change (α = .81) and Fear of change (α = .69).

Processes measure

Mastery and clarification. To measure therapist perceived mastery and clarification, we used the Bern-Post Session Report – Therapists Form (BPSR-T; Flückiger et al., 2010). The BPSR-T is evaluates psychotherapy processes immediately after a session based on Grawe's model of change (Grawe, 1997). Items are rated on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*exactly right*). The BPSR-T includes subscales for mastery and clarification, assessed by three items each. Sample items are: "Today I worked specifically to try to improve the patient's coping skills" (i.e., mastery) and "Today I have actively worked towards the patient being able to see his/her problems in new contexts" (i.e., clarification). Although there is a patient form of the BPSR, we decided to use the therapist version based on empirical studies showing lack of discriminant validity between mastery and clarification scales, when measured by the patient form (Rubel et al., 2017). Based on a function developed by Zimmermann (2015) to calculate multilevel reliability, the mastery and clarification subscales showed adequate internal consistency both at the between-patient (α mastery= .97, α clarification= .90) and at the within-patient level (α mastery= .89, α clarification= .68).

Outcome measure

Well-being. The primary outcome measure for this study was the Bern Pre-Session Report (Tschacher & Ramseyer, 2009), the only outcome variable with session-by-session repeated measures available in this naturalistic sample. This self-reported measure, completed by the patient immediately before each session, has 11 items rated on a 7-point Likert scale ranging from -3 (*not at all*) to 3 (*yes, exactly right*). Items are distributed in two factors: patients' wellbeing and patients' motivation. The first factor (i.e., patients wellbeing) includes items such as "My symptoms / problems have improved since the last session" and "I feel better overall than I did at the time of the last therapy session", and its considered a measure of micro-outcome (Tschacher & Ramseyer, 2009). For this study we used this subscale of patients' wellbeing as the outcome variable for the models. In the current sample, the well-being measurement showed excellent internal consistency at the between-patient (α= .95) and the within-patient level (α= .84) (Zimmermann, 2015).

Procedures

Patients completed the baseline measures (i.e., BSI, IIP-64, EMOREG, PATHEV) during the initial assessment process. Then, during the first ten sessions of therapy, patients completed the Bern *Pre*-Session Report (i.e., including the well-being subscale) before each therapeutic session, while therapists completed the Bern Post-Session Report after each session (i.e., including the mastery and clarification subscales).

Analytic strategy

The analysis was conducted using a three-step approach, informed by previous literature on process-outcome prediction models (e.g., Gómez Penedo et al., 2022; Rubel et al., 2020). As a first step, we estimated patient-specific effects of each process (i.e., therapist perceived mastery and clarification) on outcome (i.e., well-being) during the first 10 sessions of treatment. As a second step, we trained different ML algorithms to predict those effects based on patients' baseline characteristics. The rationale for this second step is to see if it is possible to predict for each patient before they start therapy, which process may be most relevant to target. In the final step, we evaluated the potential clinical utility of the ML algorithms in a statistically independent test sample (hold-out sample).

Estimating the effect of processes on outcome (step 1)

To estimate the individual processes' effects on outcome, we used Dynamic Structural Equation Models (DSEM; Asparouhov et al., 2018) within the software Mplus 8.8 (Muthén & Muthén, 1998-2022). DSEM is a generalization of time-series strategies to multilevel data. By using structural equation modelling and Bayesian estimation procedures, this method decomposes longitudinal data creating latent models for both between-patient components (i.e., comparing patients' average scores across time) and within-patient components (i.e., modelling each patient's fluctuations from their own average across time) (Hamaker et al., 2018). At the within-patient level, it is possible to estimate lagged relationships between variables, both in terms of auto-regressive and cross-lagged effects (Hamaker et al., 2018).

Thus, to estimate the processes effects on outcome, we followed recent literature that applied DSEM, computing separate models for each process (Gómez Penedo et al., 2022; Gómez Penedo et al., 2021; Rubel et al., 2019). In these models, at the within-patient level, variations in outcome in a given session are predicted by outcome variations in the previous session (i.e., auto-regressive effects) and process variations in the previous session (i.e., cross-lagged effects). On the other hand, process variations in a given session are predicted by process variations in the previous session (i.e., auto-regressive effects) and outcome variations in the same session (i.e., contemporaneous effect). The inclusion of this contemporaneous effect, rather than a cross-lagged effect of outcome on process, is usual when running DSEM in psychotherapy research due to the timing when process (i.e., after the session) and outcome (i.e., before the session) are usually measured (Rubel et al., 2019). Thus, equations of the models at the within-patient level were as follows:

$$\text{Outcome}_{it} = \alpha_1 + \gamma_1 \text{ Outcome}_{it\text{-}1} + \gamma_2 \text{ Process}_{it\text{-}1} + \zeta\text{Outcome}_{it}$$

$$\text{Process}_{it} = \alpha_2 + \gamma_3 \text{ Outcome}_{it} + \gamma_4 \text{ Process}_{it\text{-}1} + \zeta\text{Process}_{it}$$

Outcome variations for patient *i* at time *t* (Outcome_{it}) were predicted by patient *i*'s scores at time *t*-1 in both outcome (γ_1) and mechanism (γ_2), while the mechanism variations for patient *i* at time *t* (Process_{it}) were predicted by patient *i*'s scores at time *t* in outcome (γ_3) and by patient *i*'s scores at time *t* – 1 process (γ_4). α_1 and α_2 represents the intercept from outcome and process equations, respectively.

The inclusion of latent between-patients components in the models as covariates, allowed us to rule-out potential effects of stable patient characteristics as confounding variables when estimating the within-patients effects. Thus, this feature enhances the approximation to causal inferences when estimating the cross-lagged effect of the process on outcome.

To perform Bayesian estimations, we used Mplus default non-informative priors, so that results were exclusively determined by empirical data (Rubel et al., 2019). Following procedures applied by Hamaker et al. (2018), to conduct the DSEM we used 50,000 iterations and two chains (i.e., first half of the chains were used for burn-in process and then discarded). Furthermore, we used a thinning of 10 iterations (i.e., 10% of the iterations were saved), meaning that results reported were based on 5000 iterations. The use of the first 10 sessions of treatment allowed us to reach the minimum of repeated measures required to conduct DSEM (see Schultzberg & Muthén, 2018). Considering that multilevel models reliably accommodate for missing data, mimicking an intent-to-treat approach, we did not need to impute data for DSEM in cases of missingness.

The analysis provides mean estimates for each parameter (i.e., point estimates) and 95% credibility intervals based on the posterior distribution to determine the significance of the parameters (i.e., when zero is not included within the interval). As effect sizes (ES) for the cross-lagged effects, we used within-patient standardized coefficients interpreted based on Orth et al.'s (2022) norms: small ES = 0.03, medium ES = 0.07, and large ES = 0.12. Full Mplus script for the DSEM analyses are presented at OSF (osf.io/mv7kd).

Developing algorithms to predict process effects on outcome (step 2)

Once the DSEM models were performed, we extracted patient-specific estimations of each process's effects on outcome which we refer to as *relevance indices* (i.e., indexing how relevant each process is for each patient's progress), and which were then used as the dependent variable in the subsequent ML analyses.

To develop and test the ML algorithms, we used the caret package (Kuhn, 2020) from the free software package *R* (R Core Team, 2022). These algorithms aimed to predict the relevance indices for therapist perceived mastery and clarification based on baseline patient characteristics. Although more baseline predictors were measured at the clinic, based on the sample size available for the study and a power calculation based on the method proposed by Riley et al., (2019), we set the amount of predictors to 29 (i.e., Age, Sex, Diagnoses [depressive disorder, anxiety disorder, adjustment disorder, eating disorder, posttraumatic-stress disorder, obsessive-compulsive disorder, eating disorder, substance use disorder, personality disorder, other disorders], symptoms severity from BSI subscales [somatization, obsessive-compulsive, depressive, anxiety, hostility, phobic anxiety, paranoid ideation, psychoticism, interpersonal sensitivity, IIP-64 indexes [interpersonal distress, interpersonal agency, interpersonal communion], PATHEV subscales [hope for change, and fear of change], and EMOREG subscales [emotional expression, emotional self-control, emotional avoidance, and emotional distort]). Before running the models, all categorical predictors were dichotomised as -0.5 or 0.5, while all continuous variables were z-standardized. Considering that ML cannot handle missing data, in cases with missingness we applied a Random Forest nonparametric imputation strategy, using the *R* package missForest (Stekhoven & Buhlmann, 2012).

As a common procedure in machine learning, before developing the algorithm we randomly split the sample in two sets: a training set comprising 2/3 of the total sample (*n*= 407) and a test set consisting of the other 1/3 (*n*= 201). Algorithms were developed only in the training set, while the test set was held-out from this process, in order to use it to evaluate the performance of the algorithm in a sample that was not used to develop it (i.e., evaluating the generalizability of the algorithm and protecting against overfitting), the hold-out sample.

To train the algorithms, we applied specific ML algorithm from different families of ML models: linear models (i.e., support vector machines), regularization models (i.e., elastic net), neural networks (i.e., neural networks), bayesian models (i.e., bayesian neural networks), and decision trees (i.e., random forest, model trees and extreme gradient boosting). To train the algorithms we used a resampling method called leave-one-out cross-validation (LOOCV) strategy. In this internal cross-validation strategy, all the cases except one are used to train the algorithm and validate it in the hold-out case. The procedure is then iteratively repeated for all cases of the training set, estimating the average error of the model. To select the final algorithm for both processes, we compared the root mean square error (RMSE) of each model and selected the most parsimonious (i.e., best performing and less complex algorithm). Finally, the selected algorithms were then evaluated on the test set, by calculating Pearson's correlations between the estimated relevance indices predicted by the algorithm and the relevance indices computed with DSEM. The full *R* script to develop the ML algorithms is available in the OSF (osf.io/mv7kd).

Evaluation of clinical utility (step 3)

In order to evaluate the potential clinical utility of the ML algorithms, we developed a targeted prescription rule to establish whether the focus on mastery or clarification processes may be advantageous for each patient. First, the final ML algorithms to predict the mastery-outcome association (*mastery relevance index*) and the clarification-outcome association (*clarification relevance index*) were applied in the test sample cases to output two relevance indices for each patient. Second, a *process relevance index* (therapist perceived mastery vs. clarification) was computed by subtracting the *mastery relevance index* from the *clarification relevance index*. In this index that was standardized, negative values favoured a focus on the process of clarification and positive values favoured a focus on the process of mastery. Third, we applied the 1 standard deviation rule, following prior literature on targeted prescription (Delgadillo & Gonzalez Salas Duhne, 2019), in order to segment the distribution of the *process relevance index* into three groups: [a] a group of outliers where a focus on mastery would be indicated (a *process relevance index* at least 1 SD above the mean level); [b] a group of outliers where a focus on clarification would be indicated (a *process relevance index* at least 1 SD below the mean level); and [c] the group of cases closest to the mean (a *process relevance index* within 1 SD above/below mean), for whom a single process would not be strongly indicated. Finally, to check if the processes indicated were particularly relevant in the groups for which they were recommended (i.e., groups [a] and [b]), we calculated an empirical *treatment focus index* using the scores of therapists in the BPSR-T, across the first 10 sessions of treatment. To calculate this empirical *treatment focus index,* we subtracted the standardized average level of clarification from the standardized average level of mastery (i.e., *treatment focus index* = Average level of mastery used – Average level of clarification used). This index, after standardization, represents the extent to which the focus of therapy mostly emphasized mastery (positive index) or clarification (negative index). Once we created this empirical *treatment focus index*, we tested if patients for whom one process was recommended, the use of that process (vs. the other) was associated with outcome. On the one hand, we tested if in patients for whom mastery was recommended (i.e., group [a]), there was a significant *positive* correlation between *treatment focus index* and outcome (i.e., implying that greater use of mastery-oriented interventions was associated with better outcome in this group). Furthermore, we tested if in patients for whom clarification was recommended (i.e., group [b]), there was a significant *negative* correlation between *treatment focus index* and outcome (i.e., implying that greater use of clarification-oriented interventions were associated with better outcome in this group).To test these hypotheses, we used partial correlations, testing the association between *treatment focus index* with the outcome variable (i.e., well-being) at the end of session 10 and adjusting for baseline levels, separately for groups [a] and [b] which were classified using the ML-based targeted prescription rule.

Results

Sample descriptives

In Table 1, we present means and standard deviations of all the continuous baseline predictors included in the models (with the exception of age, reported above). Regarding the processes of change, across the first ten sessions there was a mean therapist perceived mastery score of 2.10 (*SD*= 0.97) and a mean therapist perceived clarification score of 2.46 (*SD*= 0.81) (theoretical ranges= 0 to 4). Additionally, across the first ten sessions of treatment patients had an average well-being (i.e., outcome) score of 0.40 (SD= 1.02) (theoretical range = -3 to 3). Across the first 10 sessions of therapy studied in this paper, the correlation between the therapy processes of mastery and clarification as measured from therapists' perspective was *r*=.21.

Mastery and clarification effects on outcome

In the DSEM models with *mastery* as a covariate of outcome, we found a significant within-patient effect of therapist perceived mastery on subsequent outcome, $\text{Mastery}_t \rightarrow \text{Outcome}_{t+1}$= 0.20, *SD*= 0.02, 95%CI [0.17, 0.23], $p < .001$. A positive deviation in mastery was associated with positive deviation in subsequent outcome. This magnitude of the effect of mastery on outcome (standardized at the within-patient level) represents a large effect size for cross-lagged estimates (Orth et al., 2022). Furthermore, the contemporaneous effect of outcome on therapist perceived mastery was not statistically significant, Outcome $_t$ → Mastery$_t$= 0.03, *SD*= 0.02, 95%CI [-0.001, 0.06], *p* = .054. The model explained 22% variance in outcome and 23% variance in mastery. Unstandardized posterior means and credible intervals of the fixed and random effects of the DSEM with mastery as a predictor are presented in Table 2SM in the supplemental material.

In the DSEM models with *clarification* as a covariate of outcome, we found a significant within-patient effect of therapist perceived clarification on subsequent outcome, Clarification$_t$ → Outcome$_{t+1}$= 0.04, *SD*= 0.02, 95%CI [0.01, 0.07], *p* = .014. A positive deviation in clarification was associated with positive deviation in subsequent outcome. This magnitude of the effect of clarification, also standardized at the within-patient level, represented a small-to-medium effect size (Orth et al., 2022). Additionally, the contemporaneous effect of outcome on therapist perceived clarification was not statistically significant, Outcome$_t$ → Clarification$_t$ = 0.03, *SD*= 0.02, 95%CI [-0.003, 0.06], *p* = .068. The model explained 20% variance on outcome and 12% variance on clarification. Unstandardized posterior means and credible intervals of the fixed and random effects of the DSEM with clarification as a predictor are presented in Table 3SM in the supplemental material.

Algorithms to predict mastery and clarification effects

In Table 4SM from the supplemental material, we present the performance of the different algorithm-building strategies to predict patient-specific mastery and clarification effects on outcome during the first ten sessions of treatment in the training set. Considering that the difference in errors between the two best performing algorithms was very small, we selected elastic net (i.e., the second best-performing algorithm across both mechanisms) as the final model because it provides a more parsimonious and interpretable solution than Bayesian regularized neural network (i.e., the best-performing algorithm).

When predicting therapist perceived *mastery* effects on outcome, the internal cross-validation resulted on a final elastic net model with parameters $\alpha = 1$ and $\lambda = 0.003$. In elastic net, the parameter α sets the type of penalty while the parameter λ sets the strength of regularization (Friedman et al., 2010). An $\alpha = 1$ implies that a L1 penalization was used consistent with a least absolute shrinkage and selection operator (LASSO) solution (Tibshirani, 2011). LASSO penalization shrinks some coefficients while it sets others to zero. Thus, it provides a parsimonious solution with less predictors than the ones included to develop the model. The final algorithm developed for *mastery* effects, included the following predictors: Anxiety-disorder diagnoses, adjustment disorder diagnoses, obsessive-compulsive disorder diagnoses, substance use disorder diagnoses, depressive symptoms [BSI], somatization symptoms [BSI], hope for change [PATHEV], emotional self-control [EMOREG], and emotional expression [EMOREG].

When predicting therapist perceived *clarification* effects on outcome, the internal cross-validation resulted in a final elastic net model with parameters $\alpha = 1$ and $\lambda = 0.002$, also consistent with LASSO solution. The predictors included in the final model were: Age, sex, eating disorder diagnoses, posttraumatic stress disorder diagnoses, personality disorder diagnoses, depressive disorder diagnoses, obsessive-compulsive disorder diagnoses, substance use disorder diagnoses, depressive symptoms [BSI], anxiety symptoms [BSI], somatization symptoms [BSI], obsessive-compulsive symptoms [BSI], phobic anxiety symptoms [BSI], paranoid ideation symptoms [BSI], emotional self-control [EMOREG], emotional distort [EMOREG], interpersonal distress [IIP-64], interpersonal communion [IIP-64] and interpersonal agency [IIP-64].

When evaluating the performance of these algorithms in the test set, we found significant weak-to-moderate correlations between the predicted and the observed within-patient effects of therapist perceived mastery (r=.18, $p < .001$) and clarification (r= .16, p= .02).

Evaluation of clinical utility of the algorithms

To evaluate the clinical utility of the models, based on the individual *relevance indices* derivate from the algorithms, we developed a targeted prescription rule to establish whether the focus on mastery and/or clarification processes may be advantageous for each patient. Within the test set, in patients for whom the mastery process was indicated (14% of the sample) the *treatment focus index* was positively correlated with adjusted well-being at the end of session 10, with a medium effect size (r = .33, d = .70). As the *treatment focus index* represents the extent to which the focus of therapy mostly emphasized mastery (positive index) or clarification (negative index), this means that in the group where mastery was recommended a great emphasized on mastery (compared to the level of clarification) was associated with better outcome. On the other hand, in the group of patients for whom the algorithm indicated clarification (18% of the sample), there was a small but negative correlation of the *treatment focus index* with outcome (r = -.05, d = .10). Thus, although the association had a small effect size, the results showed that in the group where clarification was indicated, a greater level of clarification (compared to the level of mastery) was related to better outcome.

Discussion

The aim of this study was to develop a targeted prescription rule to recommend a focus on mastery or clarification therapy processes for each individual patient during the first ten sessions of psychotherapy. Results of the study showed significant effects for both, therapist perceived mastery and clarification on outcome during the first ten sessions of therapy. Elastic net algorithms with LASSO-consistent parameters, were selected as the final model, being parsimonious to predict both, the relevance index for mastery and clarification. When the algorithms were applied for treatment recommendations on the test set, the level of the processes used (i.e., mastery versus clarification levels during the first ten sessions) were differentially associated with treatment outcome when comparing the two recommendation groups (i.e., mastery versus clarification indicated), although the best performance was presented in the mastery indicated group.

The significant effects of both mastery and clarification found in this paper are consistent with previous studies reporting effects of these processes on psychotherapy outcome (i.e., Gómez Penedo et al., 2022, 2023; Rubel et al., 2017; C. Schwartz et al., 2018). Nevertheless, previous studies mostly focused on measuring these processes from the perspective of patients (e.g., C. Schwartz et al., 2018), and mainly used a compound construct (e.g., problem coping experiences) that merges mastery and clarification (e.g., Gómez Penedo et al., 2022; Rubel et al., 2017). In contrast, this study focused on mastery and clarification from the perspective of therapists, thus being conceptualized as a therapist-reported process of change. Consequently, this study provided evidence to support both, mastery and clarification-oriented interventions as change processes in psychotherapy. This means that therapists should generally use interventions aiming to enhance patients' ability to cope with their problems (i.e., mastery) and their understanding of the sources and consequences of their problems (i.e., clarification). However, there was considerable between-patient heterogeneity in the effect size attributable to these different processes.

In the development of the ML algorithm for therapist perceived mastery and clarification effects, the final algorithms selected (i.e., elastic net with parameters consistent with a LASSO model; Tibshirani, 2011), has been widely used in psychotherapy research recent years (e.g., Delgadillo et al., 2017; 2022; Furukawa et al., 2020; Kilcullen et al., 2021). As LASSO shrinks the coefficients of some of the predictors to zero, it enables the development of a parsimonious (i.e., less complex) algorithm, enhancing interpretability and further implementation. The variables selected as predictors of mastery relevance included diagnostic information, symptom severity, hope for change, and emotion regulation capacities, while the algorithms for clarification relevance included demographic information, diagnostic information, clinical severity, emotion regulation and interpersonal problems. The different variables included in both algorithms implied that the patient level characteristics that might be relevant to predict mastery and clarification effects varies between processes.

Although the results of the study showed that both algorithms resulted in significant correlations between the observed and estimated effect of the processes, the size of the correlations computed on the test set were weak-to-moderate. This finding raises concerns regarding the performance of the algorithm and generalizability of the results, leaving room for improvement. Future research might need to find assessment, methodological, and analytical alternatives to enhance the predictive precision of these process-level algorithms.

Other than the studies by Lutz, Deisenhofer, et al. (2022) and Gómez Penedo et al. (2022), which developed algorithms to predict problem coping experiences, we are not aware of other studies that predict specifically mastery and clarification effects on psychotherapy outcome. The present study extends beyond earlier research by developing individual algorithms for each process of change. Furthermore, the two processes were measured from the therapists' perspective (i.e., not from the patient's perspective). Thus, the prediction models might provide relevant information for treatment personalization. This strategy might represent an easier to implement alternative compared to other treatment selection models that are conceptualized at the "package" level rather than technique or process level (e.g., Cohen et al., 2020), considering that the therapist could draw from common interventions used in different theoretical frameworks without the need of being extensively trained in diverse treatment packages. This approach could enhance precision in mental health and data-informed psychological therapy (Lutz, Schwartz, et al., 2022).

To create recommendation criteria, we used the two algorithms to compute individual *relevance indices* for the processes, while we computed an empirical *treatment focus index* to test it. Although overall results suggest the clinical utility of the algorithms to provide treatment recommendations at the process level, they seem to be specifically relevant for mastery recommendations. Thus, as mastery-oriented interventions were recommended for a small group of participants of the whole test set (i.e., 14%), the clinical utility of the algorithms should be interpreted cautiously. The small correlation of the *treatment focus index* with outcome in the clarification-indicated group, might also suggest that mastery is a common process across most patients. Thus, in this group clarification-oriented interventions might be recommended but in combination with mastery-oriented interventions.

This study has a number of limitations that would need to be addressed in future research. First, reliability of some of the measures used for baseline (e.g., subscales of BSI, EMOREG and PATHEV) and process assessment (e.g., clarification measure at the within-patient level) were low, increasing the potential noise in the models and undermining their precision. Second, the process variables were measured using self-reports by the therapists, rather than using observer-based measures. Although the therapist version of the BPSR has shown evidence of concurrent validity correlating with patients reports (Flückiger et al., 2010), there are no previous studies determining how it correlates with observer-based measures. Furthermore, the use of therapist reports might raise concerns regarding potential self-perception biases and call for replications of this study using raters to code therapists' activities. However, the information based on therapist perspective might also have advantages in terms of implementation (e.g., enhancing therapists' belief and adherence to recommendations). Third, to establish processes effects we did not use an experimental design but rather naturalistic data routinely collected at an outpatient clinic (see e.g., Allemand & Flückiger, 2020). Although this design might enhance the ecological validity of the study, it does not provide the necessary evidence to support strong causal inferences regarding mastery and clarification effects on outcome. Nevertheless, in this study we used cross-lagged DSEM models that are currently considered the most suitable strategy to approximate causal inferences in the context of observational data (Falkenström et al., 2022; Hamaker et al., 2018). Forth, as so far DSEM does not support three-level models (but only two-level models), we were not able to account for therapist effects when estimating mastery and clarification effects. This might raise concerns for the estimations of the parameters. Nevertheless, a simulation study had also suggested that in models focused on within-patient processes effects, including therapist effects does not improve model performance and might even increase bias of the estimates (Falkenström et al., 2020). Furthermore, the randomization to create the training and test set was done at the patient level, not accounting for the treating therapists. This might also had the risk of biases in the algorithms developed, even though both the predictors (patient characteristics) and the dependent variable (patient individual process-outcome association) in the models were patient-specific, with small to null contributions expected from therapists variables. Fifth, there might be some relevant predictors of mastery versus clarification effects that were not included within the initial screening and, thus, on the algorithms. Finally, a recent simulation study had shown that using individual estimates derived from multilevel models as predictors in a two-step approach, might reduce reliability of the estimates compared to a one-step approach (Liu et al., 2021). Although in this study we used these individual estimates as an outcome variable rather than as a predictor, there might be still concerns about the reliability of these individual coefficients. However, in the current state-of-the-art there are only few ML algorithms that can be conducted in a multilevel context (e.g., Fokkema et al., 2021) to run the models of the study on a single step. A greater development of ML within multilevel approaches might be necessary to develop and test different ML algorithms in a single step method, enhancing the predictive capacity and generalizability of the final algorithms. Nevertheless, the complexity of these models might make them too computational insensitive, thus being unfeasible.

Besides these limitations, the present study provides evidence to support the use of ML algorithms to predict differential processes effects of therapist perceived mastery and clarification in psychotherapy. These findings might help to develop and implement personalization criteria to assist therapists with different backgrounds and theoretical frameworks to base their decision-making regarding the use of mastery and clarification interventions.

References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, *31*(1), 92–116. https://doi.org/10.1080/10503307.2020.1808729

Allemand, M., & Flückiger, C. (2017). Changing personality traits: Some considerations from psychotherapy process-outcome research for intervention efforts on intentional personality change. *Journal of Psychotherapy Integration*, *27*(4), 476–494. https://doi.org/10.1037/int0000094

Allemand, M., & Flückiger, C. (2020). Different Routes, Same Effects. *GeroPsych*, *33*(4), 223–234. https://doi.org/10.1024/1662-9647/a000237

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Barkham, M., & Lambert, M. (2021). The efficacy and effectiveness of psychological therapies. In Michael Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 135–190). Wiley.

Caspar, F. (2022). Optimizing psychotherapy with Plan Analysis. In T. D. Eells (Ed.), *Handbook of Psychotherapy Case Formulation* (3erd ed., pp. 209–251). Guilford.

Caspar, F., Berger, T., Holtforth, M. grosse, Babl, A., Heer, S., Lin, M., Stähli, A., Gomez

Penedo, J. M., Holstein, D., Egenolf, Y., Frischknecht, E., Krieger, T., Ramseyer, F., Regli, D., Schmied, E., Flückiger, C., Brodbeck, J., Greenberg, L., Carver, C. S., … Belz, M. (2023). The impact of integrating emotion focused components into psychological therapy: A randomized controlled trial. *Journal of Clinical Psychology*, *79*(2), 296–315. https://doi.org/10.1002/jclp.23421

Caspar, F., & grosse Holtforth, M. (2010). Klaus Grawe: On a constant quest for a truly integrative and research-based psychotherapy. In L. G. Castonguay, J. C. Muran, L. Angus, J. A. Hayes, N. Ladany, & T. Anderson (Eds.), *Bringing psychotherapy research to life: Understanding change through the work of leading clinical researchers* (pp. 113–123). APA. https://doi.org/https://doi.org/10.1037/12137-000

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edn). Lawrence Erlbaum Associates.

Cohen, Z. D., Delgadillo, J., & DeRubeis, R. (2021). Personalized treatment approaches. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 673–703).

Cohen, Z. D., Kim, T. T., Van, H. L., Dekker, J. J. M., & Driessen, E. (2020). A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychotherapy Research*, *30*, 137–150. https://doi.org/10.1080/10503307.2018.1563312

Cuijpers, P., Quero, S., Noma, H., Ciharova, M., Miguel, C., Karyotaki, E., Cipriani, A., Cristea, I. A., & Furukawa, T. A. (2021). Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry*, *20*(2), 283–293. https://doi.org/10.1002/wps.20860

Delgadillo, J. (2021). Machine learning: A primer for psychotherapy researchers.

Psychotherapy Research, *31*(1), 1–4. https://doi.org/10.1080/10503307.2020.1859638

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., Cohen, Z. D., DeRubeis, R. J., & Barkham, M. (2022). Stratified Care vs Stepped Care for Depression. *JAMA Psychiatry*, *79*(2), 101. https://doi.org/10.1001/jamapsychiatry.2021.3539

Delgadillo, J., & Gonzalez Salas Duhne, P. (2019). Targeted Prescription of Cognitive-Behavioral Therapy Versus Person-Centered Counseling for Depression Using a Machine Learning Approach. *Journal of Consulting and Clinical Psychology*, *88*, 14–24. https://doi.org/10.1037/ccp0000476

Delgadillo, J., Huey, D., Bennett, H., & McMillan, D. (2017). Case complexity as a guide for psychological treatment selection. *Journal of Consulting and Clinical Psychology*, *85*, 835–853.

Delgadillo, J., & Lutz, W. (2020). A development pathway towards precision mental health care. *JAMA Psychiatry*, *77*, 889–890. https://doi.org/10.1136/bmj.e5595

Derogatis, L. (1993). *Brief Symptom Inventory (BSI), administration, scoring, and procedures manual* (N. C. Services (ed.); Third edit). Lawrence Erlbaum Associates.

Falkenström, F., Solomonov, N., & Rubel, J. A. (2020). Do therapist effects really impact estimates of within-patient mechanisms of change? A Monte Carlo simulation study. *Psychotherapy Research*, *30*(7), 885–899. https://doi.org/10.1080/10503307.2020.1769875

Falkenström, F., Solomonov, N., & Rubel, J. A. (2022). How to estimate cross-lagged effects in psychotherapy mechanisms of change research: A comparison of multilevel and structural equation models. *Journal of Consulting and Clinical Psychology*, *90*(5), 446–458. https://doi.org/https://doi.org/10.1037/ccp0000727

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, *55*, 316–340.

Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., & Caspar, F. (2010). Der Berner Patienten- und Therapeutenstundenbogen 2000. Ein Instrument zur Erfassung von Therapieprozessen. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *39*, 71–79. https://doi.org/10.1026/1616-3443/a000015

Fokkema, M., Edbrooke-Childs, J., & Wolpert, M. (2021). Generalized linear mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. *Psychotherapy Research*, *31*(3), 329–341. https://doi.org/10.1080/10503307.2020.1785037

Franke, G. H. (2000). *BSI. Brief Symptom Inventory - Deutsche Version. Manual.* Beltz.

Friedl, N., Berger, T., Krieger, T., Caspar, F., & Grosse Holtforth, M. (2020). Using the Personalized Advantage Index for individual treatment allocation to cognitive behavioral therapy (CBT) or a CBT with integrated exposure and emotion-focused elements (CBT-EE). *Psychotherapy Research*, *30*, 763–775. https://doi.org/10.1080/10503307.2019.1664782

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Furukawa, T. A., Debray, T., Akechi, T., Yamada, M., Kato, T., Seo, M., & Efthimiou, O. (2020). Can personalized treatment prediction improve the outcomes , compared with the group average approach, in a randomized trial? Developing and validating a multivariable prediction model in a pragmatic megatrial of acute treatment for major depression. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2020.05.141

Gomez Penedo, J. M., Constantino, M. J., Coyne, A. E., Bernecker, S. L., & Smith-hansen, L. (2019). Patient baseline interpersonal problems as moderators of outcome in two psychotherapies for bulimia nervosa. *Psychotherapy Research*, *29*, 799–811.

https://doi.org/10.1080/10503307.2018.1425931

Gómez Penedo, J.M., Babl, A., Dyresen, A., Fernández-Álvarez, J., Flückiger, C., & grosse Holtforth, M. (2023). Problem mastery and motivational clarification as mechanisms of change in cognitive-behavioral therapy for depression: Secondary analysis of a randomized controlled trial. *Behaviour Research and Therapy*, *167*, 104343. https://doi.org/10.1016/j.brat.2023.104343

Gómez Penedo, J.M., Hilpert, P., grosse Holtforth, M., & Flückiger, C. (2021). Interpersonal cognitions as a mechanism of change in Cognitive Behavioral Therapy for generalized anxiety disorder? A Multilevel Dynamic Structural Equation Model Approach. *Journal of Consulting and Clinical Psychology*, *89*(11), 898–908. https://doi.org/10.1037/ccp0000690

Gómez Penedo, J.M., Schwartz, B., Giesemann, J., Rubel, J. A., Deisenhofer, A.-K., & Lutz, W. (2022). For whom should psychotherapy focus on problem coping? A machine learning algorithm for treatment personalization. *Psychotherapy Research*, *32*(2), 151–164. https://doi.org/10.1080/10503307.2021.1930242

Gómez Penedo, J. M., Meglio, M., & Bornhauser, L. P. (2022, September 21). Developing a Machine Learning Algorithm to predict the most suitable mechanisms of change (mastery vs. clarification) for each individual patient. Retrieved from osf.io/mv7kd

Grawe, K., Caspar, F., & Ambühl, H. (1990). Die Berner Therapievergleichsstudie: Wirkungsvergleich und differentielle Indikation. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *19*(4), 338–361.

Grawe, K. (1997). Research-informed psychotherapy. *Psychotherapy Research*, *7*(1), 1–19. https://doi.org/10.1080/10503309712331331843

Grawe, K. (2004). *Psychological therapy*. Hogrefe Publishing.

grosse Holtforth, M., Grawe, K., Fries, A., & Znoj, H. (2008). Inkonsistenz als

Differenzielles Indikationskriterium in der Psychotherapie - Eine Randomisierte Kontrollierte Studie. *Zeitschrift Fur Klinische Psychologie Und Psychotherapie*, *37*(2), 103–111. https://doi.org/10.1026/1616-3443.37.2.103

grosse Holtforth, M., Wilm, K., Beyermann, S., Rhode, A., Trost, S., & Steyer, R. (2011). Differential change in integrative psychotherapy: A re-analysis of a change-factor based RCT in a naturalistic setting. *Psychotherapy Research*, *21*(6), 631–643. https://doi.org/10.1080/10503307.2011.602749

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic Structural Equation Models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, *53*, 820–841. https://doi.org/10.1080/00273171.2018.1446819

Horowitz, L. M., Alden, L. E., Kordy, H., & Strauß, B. (2000). *Inventar zur Erfassung interpersonaler Probleme: deutsche Version; IIP-D*. Beltz-Test.

Horowitz, Len M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *IIP: Inventory of Interpersonal Problems manual.* Psychological Corporation.

Kaiser, T., Volkmann, C., Volkmann, A., Karyotaki, E., Cuijpers, P., & Brakemeier, E.-L. (2022). Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clinical Psychology: Science and Practice*, *29*(3), 294–303. https://doi.org/https://doi.org/10.1037/cps0000079

Khazanov, G. K., Xu, C., Dunn, B. D., Cohen, Z. D., DeRubeis, R. J., & Hollon, S. D. (2020). Distress and anhedonia as predictors of depression treatment outcome: A secondary analysis of a randomized clinical trial. *Behaviour Research and Therapy*, *125*(October 2019), 103507. https://doi.org/10.1016/j.brat.2019.103507

Kilcullen, J. R., Castonguay, L. G., Janis, R. A., Hallquist, M. N., Hayes, J. A., & Locke, B. D. (2021). Predicting future courses of psychotherapy within a grouped LASSO framework. *Psychotherapy Research*, *31*(1), 63–77. https://doi.org/10.1080/10503307.2020.1762948

Liu, S., Kuppens, P., & Bringmann, L. (2021). On the use of Empirical Bayes Estimates as measures of individual traits. *Assessment*, *28*(3), 845–857. https://doi.org/10.1177/1073191119885019

Lorenzo-Luaces, L., Peipert, A., De Jesús Romero, R., Rutter, L. A., & Rodriguez-Quintana, N. (2021). Personalized medicine and cognitive-behavioral therapies for depression: Small effects , big problems, and bigger data. *International Journal of Cognitive Therapy*, *14*(1), 59–85. https://doi.org/10.31234/osf.io/du827

Lutz, W., de Jong, K., Rubel, J., & Delgadillo, J. (2021). Measuring, predicting and tracking change in psychotherapy. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield`s Handbook of Psychotherapy and Behavior Change* (7th ed., pp. 89–133). Wiley.

Lutz, W., Deisenhofer, A.-K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, *90*(1), 90–106. https://doi.org/10.1037/ccp0000642

Lutz, W., Rubel, J. A., Schwartz, B., Schilling, V., & Deisenhofer, A. (2019). Towards integrating personalized feedback research into clinical practice: Development of the Trier Treatment Navigator (TTN). *Behaviour Research and Therapy*, *120*, 103438. https://doi.org/10.1016/j.brat.2019.103438

Lutz, W., Schwartz, B., & Delgadillo, J. (2022). Measurement-Based and Data-Informed Psychological Therapy. *Annual Review of Clinical Psychology*, *18*(1), 71–98. https://doi.org/10.1146/annurev-clinpsy-071720-014821

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (eighth edi). Muthen & Muthen.

Newman, M. G., Jacobson, N. C., Erickson, T. M., & Fisher, A. J. (2017). Interpersonal Problems Predict Differential Response to Cognitive Versus Behavioral Treatment in a Randomized Controlled Trial. *Behavior Therapy*, *48*(1), 56–68. https://doi.org/10.1016/j.beth.2016.05.005

Orth, U., Meier, L. L., Bühler, J. L., Dapp, L. C., Krauss, S., Messerli, D., & Robins, R. W. (2022). Effect size guidelines for cross-Lagged effects. *Psychological Methods*. https://doi.org/https://doi.org/10.1037/met0000499

Podina, I. R., Vîslă, A., Fodor, L. A., & Flückiger, C. (2019). Is there a sleeper effect of exposure-based vs. cognitive-only intervention for anxiety disorders? A longitudinal multilevel meta-analysis. *Clinical Psychology Review*, *73*(November 2018), 101774. https://doi.org/10.1016/j.cpr.2019.101774

Probst, T., Jakob, M., Kaufmann, Y. M., Müller-Neng, J. M. B., Bohus, M., & Weck, F. (2018). Patients' and therapists' experiences of general change mechanisms during bug-in-the-eye and delayed video-based supervised cognitive-behavioral therapy. A randomized controlled trial. *Journal of Clinical Psychology*, *74*(4), 509–522. https://doi.org/10.1002/jclp.22519

Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. *Statistics in Medicine*, *38*, 1262–1275. https://doi.org/10.1002/sim.7993

Rubel, J. A., Hilpert, P., Wolfer, C., Held, J., Vîslă, A., & Flückiger, C. (2019). The working alliance in manualized CBT for generalized anxiety disorder: Does it lead to change and does the effect vary depending on manual implementation flexibility? *Journal of Consulting and Clinical Psychology*, *87*, 989–1002. https://doi.org/10.1037/ccp0000433

Rubel, J. A., Rosenbaum, D., & Lutz, W. (2017). Patients' in-session experiences and symptom change: Session-to-session effects on a within- and between-patient level. *Behaviour Research and Therapy*, *90*, 58–66.

Rubel, J. A., Zilcha-mano, S., Giesemann, J., Prinz, J., & Lutz, W. (2020). Predicting personalized process-outcome associations in psychotherapy using machine learning approaches - A demonstration. *Psychotherapy Research*, *30*, 300–309. https://doi.org/10.1080/10503307.2019.1597994

Rutledge, R. B., Chekroud, A. M., & Huys, Q. J. (2019). Machine learning and big data in psychiatry: toward clinical applications. *Current Opinion in Neurobiology*, *55*, 152–159. https://doi.org/10.1016/j.conb.2019.02.006

Schultzberg, M., & Muthén, B. (2018). Number of Subjects and Time Points Needed for Multilevel Time-Series Analysis: A Simulation Study of Dynamic Structural Equation Modeling. *Structural Equation Modeling*, *25*(4), 495–515. https://doi.org/10.1080/10705511.2017.1392862

Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research*, *31*, 33–51. https://doi.org/10.1080/10503307.2020.1769219

Schwartz, C., Hilbert, S., Schlegl, S., Diedrich, A., & Voderholzer, U. (2018). Common change factors and mediation of the alliance – outcome link during treatment of depression. *Journal of Consulting and Clinical Psychology*, *86*, 584–592.

Stekhoven, D. J., & Buhlmann, P. (2012). MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *73*(3), 273–

282. https://doi.org/10.1111/j.1467-9868.2011.00771.x

Tschacher, W., & Ramseyer, F. (2009). Modeling psychotherapy process by time-series panel analysis (TSPA). *Psychotherapy Research*, *19*(4–5), 469–481. https://doi.org/10.1080/10503300802654496

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

Zilcha-Mano, S., Muran, J. C., Eubanks, C. F., Safran, J. D., & Winston, A. (2018). Not just a non-specific factor: Moderators of the effect of within- and between-clients alliance on outcome in CBT. *Cognitive Therapy and Research*, *42*, 146–158. https://doi.org/10.1007/s10608-017-9866-5

Zimmermann, J. (2015). *R script for estimating four different indices of multilevel reliability. R function (Version 1.11).*

Znoj, H. J., & Grawe, K. (2000). The control of unwanted states and psychological health: Consistency safeguards. In A. Grob & W. Perrig (Eds.), *Control of human behaviour, Mental processes and awareness* (pp. 263–282). Erlbaum.

Table 1

Descriptive results of the average scores of the different baseline scales and subscales that were included as predictors (n = 608).

Measures Scales/Subscales	Mean (*SD*)
BSI	
Somatization symptoms	0.72 (0.75)
Obsessive-compulsive symptoms	1.37 (0.88)
Depressive symptoms	1.35 (0.95)
Anxiety symptoms	1.18 (0.88)
Hostility	0.94 (0.70)
Phobic anxiety symptoms	0.70 (0.86)
Paranoid ideation	0.86 (0.77)
Psychoticism	0.90 (0.75)
Interpersonal sensitivity	1.39 (0.95)
IIP-64	
Interpersonal distress	1.50 (0.49)
Interpersonal agency	-0.52 (0.45)
Interpersonal communion	0.24 (0.46)
PATHEV	
Hope for of change	3.92 (0.75)
Fear of change	1.71 (0.78)
EMOREG	
Emotional expression	3.75 (1.28)
Emotional self-control	3.21 (0.86)
Emotional avoidance	3.27 (0.87)
Emotional distort	3.47 (1.05)

Note. SD = Standard Deviation; BSI = Brief Symptom Inventory; IIP64 = Inventory of Interpersonal Problems - 64-item version; PATHEV = Patients' Therapy Expectation and Evaluation Questionnaire; EMOREG = Emotion Regulation Questionnaire.

View publication stats