First impressions count: Therapists' impression on patients' motivation and helping alliance predicts psychotherapy dropout

Kristin Jankowsky, Johannes Zimmermann, Ulrich Jaeger, Robert Mestel & Ulrich Schroeders

View supplementary material ⏴

Published online: 09 Oct 2024.

Submit your article to this journal ⏴

Article views: 1350

View related articles ⏴

View Crossmark data ⏴

Routledge
Taylor & Francis Group

Check for updates

RESEARCH ARTICLE

First impressions count: Therapists' impression on patients' motivation and helping alliance predicts psychotherapy dropout

KRISTIN JANKOWSKY[1], JOHANNES ZIMMERMANN[1], ULRICH JAEGER[2], ROBERT MESTEL[3], & ULRICH SCHROEDERS[1]

[1]*Department of Psychology, University of Kassel, Kassel, Germany;* [2]*Asklepios Clinic Tiefenbrunn, Rosdorf, Germany &* [3]*Vamed Clinic, Bad Grönenbach, Germany*

Abstract
Objective: : With meta-analytically estimated rates of about 25%, dropout in psychotherapies is a major concern for individuals, clinicians, and the healthcare system at large. To be able to counteract dropout in psychotherapy, accurate insights about its predictors are needed. **Method**: We compared logistic regression models with two machine learning algorithms (elastic net regressions and gradient boosting machines) in the prediction of therapy dropout in two large inpatient samples ($N = 1{,}691$ and $N = 12{,}473$) using baseline and initial process variables reported by patients and therapists. **Results**: Predictive accuracies of the two machine learning algorithms were similar and higher than for logistic regressions: Therapy dropout could be predicted with an AUC of .73 and .83 for Sample 1 and 2, respectively. The initial evaluation of patients' motivation and the therapeutic alliance rated by the respective therapist were the most important predictors of dropout. **Conclusions**: Therapy dropout in naturalistic inpatient settings can be predicted to a considerable degree by using baseline indicators and therapists' first impressions. Feature selection via regularization leads to higher predictive performances whereas non-linear or interaction effects are dispensable. The most promising point of intervention to reduce therapy dropouts seems to be patients' motivation and the therapeutic alliance.

Keywords: therapy dropout; predictive modeling; machine learning; inpatients; helping alliance

Clinical or methodological significance of this article: Therapy dropout is a highly relevant issue for clinicians since prematurely terminated therapies imply worse treatment response and thus, non-optimal spent health resources. We found that inpatients' therapy dropout could be predicted to a large degree employing regularized logistic regressions using baseline indicators and initial process variables. The therapists' initial impression of patients' motivation or helping alliance was the most important variable for the prediction of therapy dropout irrespective of modeling algorithm. This finding provides clinicians with a potential point of reference to address patients at risk of dropout. On a more methodological stance, we discuss the requirements of reasonable (machine learning based) predictive modeling for clinical classification tasks.

More than one in four people are affected by mental illness at some point in their life (e.g., Steel et al., 2014). To reduce individual suffering and problematic consequences for society as a whole (e.g., due to a high number of sick leaves), it is particularly important to offer the best possible treatment. To this end, psychotherapy has been shown to be effective (Barth et al., 2013; Kamenov et al., 2017; Leichsenring et al., 2022), under the condition of its regular completion. However, with meta-analytically

estimated rates of about 20-25%, dropout (usually defined as the termination of psychotherapy initiated by the patient against the therapist's advice) is a major issue in psychotherapies (Fernandez et al., 2015; Hans & Hiller, 2013). On an individual level, dropout is problematic because patients respond worse to premature discontinued treatments (e.g., Barrett et al., 2008), which in turn might lead to disease chronification. On a societal level, dropout implies that limited health care resources are not optimally spent.

Previous studies found several patient characteristics to be associated with therapy dropout such as clinical variables (e.g., substance abuse, comorbidities, baseline symptom severity), demographics (e.g., lower education, younger age, being male, unemployment), personality traits, and attitudes or motivation toward therapy (Bucher et al., 2019; Fernandez et al., 2015; Karterud et al., 2003; Schmidt et al., 2019; Swift & Greenberg, 2012; Zimmermann et al., 2017). To reduce dropout rates, more recent studies proposed to predict patients' risk of eventual dropout using the oftentimes large amount of patient data usually gathered within baseline documentations of psychotherapies (e.g., Bennemann et al., 2022; Gonzalez Salas Duhne et al., 2022). Bennemann et al. (2022) recently compared 21 different machine learning algorithms for the prediction of therapy dropout of cognitive–behavioral therapy in a German outpatient sample and found that an ensemble of a Random Forest and Nearest Neighbor Model achieved the overall moderate best predictive accuracy with an Area Under the Curve (AUC) of .66. Important predictors were education, age, and subscales of the Personality Style and Disorder Inventory (PSDI; Kuhl & Kazén, 2009) as well as the Brief Symptom Inventory (BSI; Derogatis, 1982). However, outpatients differ from inpatients regarding the frequency of therapy sessions, overall treatment timeframe, and dropout rates (e.g., Fernandez et al., 2015). Hence, it is unclear whether these results can be generalized across treatment settings.

Regarding therapy dropout in German inpatient samples, Reuter and Scheidt (2014) summarized 15 studies with 18,147 patients and found that dropout rates varied between 7 and 23% across studies (weighted mean of 12.5%). Most studies focused on demographic, socio-medical, and clinical variables when comparing dropouts to patients who completed their therapy regularly. On average, dropouts were younger, unemployed or incapable to work, more likely diagnosed with personality disorders, and reported lower symptom severity (which was a finding that did not generalize to therapist-reported symptom severity). Other potentially relevant variables such as a patients' greater

motivation or ability to engage in therapy were only assessed in single studies but showed a negative association with dropout. Contrary to the aforementioned findings for outpatient samples, sex or highest level of education did not show an association with therapy dropout.

Hitherto, there were only few attempts to predict therapy dropout in inpatient samples using baseline variables and none of them used modeling approaches that are capable of capturing complex relationships between predictor variables and dropout. In one of the few studies that used a comprehensive set of predictors including demographic, socio-medical, clinical, and motivational characteristics of patients (Lang et al., 2006) to predict psychotherapy dropout in German inpatients ($N = 3,632$), a stepwise logistic regression achieved a Nagelkerke's R^2 of .04 and classified all patients into the group of regular terminations (which corresponded to 93.1% of all patients). Beyond baseline information and process variables, the literature review by Reuter and Scheidt (2014) also summarized the reasons given by patients when asked directly why they terminated their inpatient treatment prematurely. Reasons related to the inpatient setting were, for example, homesickness or dissatisfaction with the accommodation, food, or specific clinic rules, again emphasizing the need to conceptualize reasons for dropout differently in the inpatient setting than in the outpatient setting.

The Present Study

In this study, we aim to answer two main questions on the prediction of therapy dropout in inpatients of two German psychotherapy clinics: First, we will investigate how accurately therapy dropout can be predicted for inpatient samples using variables collected within the standard baseline documentation and after an initial therapy session. Complex machine learning algorithms might help to enhance the accuracy of such predictions because they handle multicollinearity between predictor variables and allow to integrate non-linear and (higher-order) interaction effects without the need to a priori specify the associations between variables and the respective outcome (James et al., 2017). Thus, we will compare two machine learning algorithms (a regularized logistic regression and a more complex tree-based algorithm) to an unregularized logistic regression analysis at predicting therapy dropout. Second, we examine which self- or therapist-reported variables are particularly indicative of subsequent therapy dropout. We will compare our results across two inpatient samples and discuss them regarding previous research on outpatients to be able to evaluate the generalizability of

predictive accuracies and variable importances across samples and therapy settings. This study was not pre-registered. All analyses are exploratory; we had no specific a priori hypotheses regarding the two afore-mentioned goals or research questions apart from the respective rationales we presented (see also Wagenmakers et al., 2012).

Method

Samples

We reanalyzed anonymized routine outcome moni-toring data assessed at admission from two psy-chotherapy clinics in Germany. All patients gave written informed consent. We assert that all pro-cedures contributing to this work comply with the Helsinki Declaration of 1975, as revised in 2013. In Germany, patients are usually referred to inpatient psychotherapy clinics either by physicians or practi-cing psychotherapists in cases in which outpatient therapies were not successful or there is a critical worsening of symptoms. Although the therapeutic concept may differ in some details between clinics, there are some overarching principles that apply to inpatient therapy in Germany. First, patients are offered a multi-modal treatment including individual therapy sessions and several group therapy sessions per week, but also therapy offers such as art therapy or physiotherapy, which are tailored to the individual patient's needs. Following this logic, there is also a multi-professional team involved which creates indi-vidual treatment plans (e.g., therapy frequency or programs differ across patients). Finally, patients meet other patients, which adds to the unique dynamic of inpatient stays where patients ideally benefit from each other's perspective.

Sample 1 of our study comprised 1,691 patients treated between 2007 and 2011 (62.12% women, age ranging from 17 to 71 years; $M = 35.49$, $SD = 11.88$). Length of stay was between 1–252 days ($M = 79.81$, $SD = 35.75$). A total of 131 (7.8%) termi-nated their therapy prematurely and against their therapists' recommendation. Most common diag-noses were depressive or major depressive disorder, anxiety or stress-related disorders, and personality disorders. Patients suffering from acute psychosis, acute suicidality, dementia or withdrawal symptoms were generally not admitted. Sample 2 comprised 12,473 patients treated between 1995 and 2010 (72.18% women, age ranging from 17 to 80 years; $M = 38.73$, $SD = 11.04$). Length of stay was between 1–224 days ($M = 55.27$, $SD = 23.61$). A total of 640 (5.1%) patients terminated their therapy prematurely and against their therapists' rec-ommendation. The clinic mainly focused on depression and anxiety disorders, personality dis-orders (especially borderline personality disorders), psychosomatic diseases, eating disorders, and sub-stance use disorders. Patients suffering from severe internal or brain-organic diseases or with acute psy-chosis were not admitted. For a more detailed over-view of patient characteristics, please see Table S1.

Measures

Outcome Variable

For our analyses, we defined all patients within the category "regular completion" as patients who suc-cessfully completed their therapy and all cases in which the premature termination was initiated by the patient as dropouts. By only using patients of these categories, the resulting sample sizes were $n = 1,691$ and $n = 12,473$ for Sample 1 and 2, respect-ively (for the different reasons of discharge and their prevalence, see Table S2). Thus, dropout was used as a dichotomous outcome: Patients either com-pleted their therapy regularly (coded as 0) or patients prematurely terminated the therapy (coded as 1).

Predictor Variables

We used a broad range of variables to predict therapy dropout: 168 for Sample 1 and 147 for Sample 2 after dummy-coding. All variables were assessed at baseline, that is, on the day of the initial interview within a short time after admission. The baseline documentation in both clinics included a large set of variables that have been shown to be relevant in other studies on psy-chotherapy dropout in inpatients (see Reuter & Scheidt, 2014), allowing comparisons of our results with previous findings. We summarize the available predictors in Table I (see Table S1 for the exact response options and descriptive statistics including information on the amount of missingness for the full datasets). Generally, there was a high overlap of avail-able predictor variables or constructs across samples (especially for demographic variables), but some infor-mation, such as substance abuse or traumatic experi-ences, were only available in one of the samples.

Statistical Analyses

For all models, we employed a nested cross-vali-dation approach combining an outer and inner vali-dation loop (e.g., Pargent et al., 2023) to avoid any information leakage between the training and the testing data. In each iteration of the outer validation loop, the full data were split into training data (80% of the full data set) and testing data (the remaining

Table I. Predictor variables across samples

	Variable/Construct	Sample 1	Sample 2
Demographics	Age	X	X
	Gender	X	X
	Marital or relationship status	X	X
	Education	X	X
	Occupation	X	X
	Nationality	X	X
	Source of income	X	X
	Living situation	X	X
	Number of children	X	X
Socio-medical	Pension due to reduced earning capacities	X	X
	Working capability at admission	X	X
	Caring for relatives		X
Treatment history	Type of referral / admission	X	X
	Waiting time		X
	Symptom onset		X
	Previous treatments	X	X
	Previous stays in the same clinic		X
Clinical/personality	Suicidality	X	X
	Previous psychological conditions	X	
	Traumatic experiences	X	
	Drug (ab)use		X
	ICD-10 Diagnosis[a]	X	X
	Impairment Severity Score (BSS; Schepank, 1995)	X	X
	Beck Depression Inventory (BDI; Beck et al., 1961)		X
	Structural Analysis of Social Behavior (SASB; Benjamin, 1974)		X
	Symptom Checklist-90 (SCL-90; Franke, 1995)	X	X
	Operationalized Psychodynamic Diagnosis (OPD; Arbeitskreis OPD, 1996)		X
	Interpersonal Problems (IPP; Alden et al., 1990)	X	X
	Relationship Questionnaire (Griffin & Bartholomew, 1994)		X
Other	Life satisfaction	X	X
Rated by the therapists after an initial session	Helping alliance[b]	X	
	Patients' motivation[c]		X

Note. [a]Diagnoses were assigned based on ICD-10 criteria/checklists and additional available information on previous symptoms by a therapist who conducted the individual therapy sessions. [b]The helping alliance measure was based on a helping alliance concept proposed by Luborsky et al. (1983). We used the scale sum of the 9 items (agreement is rated on a 6-point scale) that the therapists rated after an initial session of the patients' stay. The measure assesses the therapist's perception of their ability to empathize with the patient, the patient's motivation and introspection capabilities, the quality of the therapeutic alliance, shared treatment goals, the patient's emotional receptiveness to interventions, the personal impact of the patient's difficulties on the therapist, the therapist's effectiveness in helping with the patient's issues, and the overall satisfaction with the therapeutic work. [c]Motivation was rated by the therapists who conducted a first individual therapy session; they were asked to rate the patients' motivation for therapy (at the beginning of the stay) on a five-point scale ranging from 1 (*not motivated*) to 5 (*highly motivated*).

20%). In the inner loop, the models were trained using 10-fold cross-validation with separately imputed training datasets. We relied on up-sampling to address imbalanced datasets, where persons from the minority group (i.e., dropouts) were up-sampled to match the size of completers (e.g., Jacobucci & Li, 2022). Finally, we assessed the predictive performance of these models using the independent testing data. We performed 300 iterations of the outer validation loop and averaged the results to obtain a robust estimate of the expected prediction performance when presented with unseen testing data.

All our analyses were conducted using the R package *caret* (Kuhn, 2008) as an interface for prediction and model evaluation. We evaluated three incremental complex modeling approaches to predict therapy dropout. We started with a logistic regression model as a baseline and then compared it to logistic elastic net regressions (Enet; using the R package *glmnet*; Friedman et al., 2001), and gradient boosting machines (GBM; using the package *gbm*; Greenwell et al., 2019). Elastic net regressions are regularized regressions that balance between ridge and least absolute shrinkage and selection operator (LASSO) regressions to create parsimonious models and to maximize predictive performance (e.g., Zou & Hastie, 2005). GBM are tree-based algorithms that take into account non-linear and higher-order interaction effects into the modeling

process without requiring any prior specifications of the functions between predictor variables and the outcome (James et al., 2017).

The comparison of these three modeling approaches roughly corresponds to the "deductive data mining" approach proposed by Hong et al. (2020) which represents a stepwise model comparison to gain more information about the likely nature of relationships in the data. If the elastic net regressions perform more accurately in predicting therapy dropout than the logistic regressions, it suggests that overfitting (i.e., the difference between training and testing performance) hampers the performance of unregularized logistic regressions. If the GBM models have an advantage over the elastic net regressions, it indicates the presence of non-linear effects or interactions. We have not included additional ML algorithms because the functionality of other algorithms is similar and the results have been shown to be essentially the same in comparable cases (see Bennemann et al., 2022, for the prediction of therapy dropout in an outpatient setting). Annotated syntax for all analyses including the specific tuning parameter grids of the machine learning algorithms is available in an open project repository at https://osf.io/dr54h/.

Model Evaluation

To evaluate the classification performance, we calculated several metrics that were averaged across 300 iterations of the outer loop. In more detail, we calculated the balanced accuracy (average of sensitivity and specificity), sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), and the Area Under Curve (AUC). For the best performing models, we also show ROC curves for each testing split and an averaged ROC curve across all 300 testing splits. Additionally, to focus not only on predictive accuracy, but also to gain insight into which variables were most predictive of therapy dropout, we report variable importance measures scaled from 0 to 100. The *varImp* function which is included in the R package caret provides a measure of contribution for each predictor in the model. For the elastic net regressions, these variable importances represent a transformation of the absolute standardized regression coefficients of the tuned training model. For the GBM, they represent the relative influence of a predictor averaged over all generated trees (Friedman, 2001).

Additional Analyses and Sensitivity Checks

The following section discusses additional analyses that were conducted after an initial set of results was established. First, as a general sensitivity check,

we examined whether the models predominantly classified early dropouts correctly, that is, patients who dropped out during the first week of treatment, indicating that we would only be able to detect patients who probably were less inclined to participate from the onset. To do so, we excluded all early dropouts (28 patients or 21% in Sample 1 and 211 patients or 33% in Sample 2), reran the elastic net regressions using all predictor variables, and compared the performance of these models with our initial modeling approach using all data.

Second, given the superior predictive performance in Sample 2 compared to in Sample 1, we explored whether the observed pattern of results was merely a function of sample size or whether other factors might also play a role. To shed some light on this issue, we investigated the effect of sample size, number of dropout events, and events fraction (i.e., the number of dropouts relative to the full sample size) on predictive accuracy using elastic net regressions and all available predictor variables in Sample 2. The setup was as follows: In a Sample Size condition, we examined the effect of a stepwise sample size reduction for Sample 2 (full sample, 10,500, 8,500, 6,500, 5,500, 4,500, 3,500, 2,500, 2,000, 1,700, 1,400, 1,100, 800, 500) and kept the events fraction fixed at 5.13% (as the prevalence in the full sample). In a Dropouts condition, we examined the effect of a stepwise reduction of the event fraction (5.13, 4.36. 3.55, 2.75, 2.33, 1.91, 1.50, 1.28, 1.07, .0.86, 0.73, 0.60, 0.47, 0.35, 0.22) by keeping all completers in the sample and reducing the number of events at the same rate as they were reduced in the Sample Size condition.

Third, given the outstanding importance of the initial impression regarding patients' motivation or helping alliance as assessed by the therapist—irrespective of sample or algorithm—we additionally tested the predictive performance of elastic net regressions using all predictors except of helping alliance for Sample 1 and patients' motivation for Sample (see the "all but one predictor" model in Table II). In addition, we used logistic regressions with these most important variables as the only predictors (see the "single predictor" model in Table II).

Results

In Table II, we summarize common metrics of predictive performance across 300 iterations of logistic regressions, elastic net regressions, and GBM for both samples. Considering only the averaged balanced accuracy (ABA) within the testing datasets in Sample 1, elastic net regressions and GBM achieved identical values (ABA = .67), whereas the

Table II. Averaged performance metrics across 300 testing iterations

	Sample 1						Sample 2					
	ABA	Sensitivity	Specificity	PPV	NPV	AUC	ABA	Sensitivity	Specificity	PPV	NPV	AUC
Logistic Regression	.58 [.05]	.36 [.10]	.81 [.03]	.13 [.03]	.94 [.01]	.64 [.06]	.74 [.02]	.68 [.04]	.80 [.01]	.15 [.01]	.98 [.00]	.81 [.02]
Elastic Net Regression	.67 [.04]	.64 [.09]	.71 [.02]	.15 [.02]	.96 [.01]	.74 [.04]	.75 [.02]	.70 [.04]	.80 [.01]	.16 [.01]	.98 [.00]	.82 [.02]
Gradient Boosting	.67 [.05]	.64 [.11]	.70 [.04]	.15 [.02]	.96 [.01]	.72 [.04]	.74 [.03]	.73 [.07]	.75 [.10]	.15 [.03]	.98 [.01]	.83 [.02]
Single Predictor	.66 [.04]	.63 [.08]	.70 [.02]	.15 [.02]	.96 [.01]	.73 [.04]	.73 [.02]	.63 [.04]	.83 [.01]	.17 [.01]	.98 [.00]	.77 [.02]
All But One Predictor	.59 [.04]	.50 [.09]	.68 [.04]	.11 [.02]	.94 [.01]	.62 [.05]	.67 [.02]	.61 [.04]	.73 [.01]	.11 [.01]	.97 [.00]	.74 [.02]

Note. ABA = Average Balanced Accuracy; Sensitivity = Ratio of correctly classified dropouts to all dropouts; Specificity = Ratio of correctly classified completers to all completers; PPV = Positive Predictive Value (Proportion of true dropouts out of all patients who were flagged as dropouts); NPV = Negative Predictive Value (Proportion of true completers out of all patients who were flagged as completers); AUC = Area Under the Curve. Standard deviations are given in brackets. In the single predictor model, only helping alliance (Sample 1) or patients' motivation (Sample 2) as rated by the therapists are included in a logistic regression model. In the all but one predictor model, those variables were excluded of an elastic net regression model, respectively.

predictions with logistic regressions were clearly less accurate (ABA = .59). The same pattern across algorithms occurred for AUC, an alternative measure of predictive performance. Compared to Sample 1, the ABA and AUC within Sample 2 were generally higher and the difference between the modeling approaches negligible (logistic regression: ABA = .74; Enet and GBM: ABA = .75). Moreover, when comparing these results to elastic net regression models in which we excluded all early dropouts, ABA slightly decreased in both samples (ΔABA = .01 in Sample 1 and .03 in Sample 2), but this decrease can be explained in part by the inherent reduction in the number of overall dropouts in these analyses. Hence, we did not find any indication for differences in the predictive accuracy dependent on the time of dropout.

To further illustrate the differences in predictive performance across both samples and the two machine learning algorithms, Figure 1 shows ROC curves for each of the 300 testing iterations as well as averaged ROC curves for each sample. Differences between the performance of the algorithms within samples were moderate and unsystematic. However, it is evident that the larger sample size of Sample 2 led to less variation across different training-testing data splits. In other words, the person sampling implemented in the outer validation loop becomes less important.

There are different potential explanations for the superior prediction performance in Sample 2: (a) there might be specific important or more reliable predictor variables in the data set, (b) the larger sample size, and (c) the larger overall number of dropout events. Figure 2 shows the results of an additional analysis aiming to explore the effect of sample size, number of events, and events fraction on predictive accuracy. There were two main findings of this supplementary analyses: First, when reducing the overall sample size (starting at the total sample size of Sample 2) and keeping the events fraction fixed at 5.13% (*Sample Size condition*), balanced accuracies remained stable till $N = 4,500$, decreasing more steeply for $N < 2,000$. When the sample size is reduced to the one of Sample 1, the balanced accuracies of Sample 2 were still higher in comparison, but less pronounced (.70 and .67 compared to .75 and .67 for the full samples). Second, the *Dropouts condition* showed that the effect of reducing the events fraction (i.e., keeping all completers in the sample and only reducing the number of events/dropouts) was even more crucial with an overall steeper decrease and lower accuracies for very small fractions scenarios.

Irrespective of the sample or the algorithm, averaged sensitivities in the testing dataset were generally lower than specificities, meaning that the group of dropouts could be detected less accurately compared
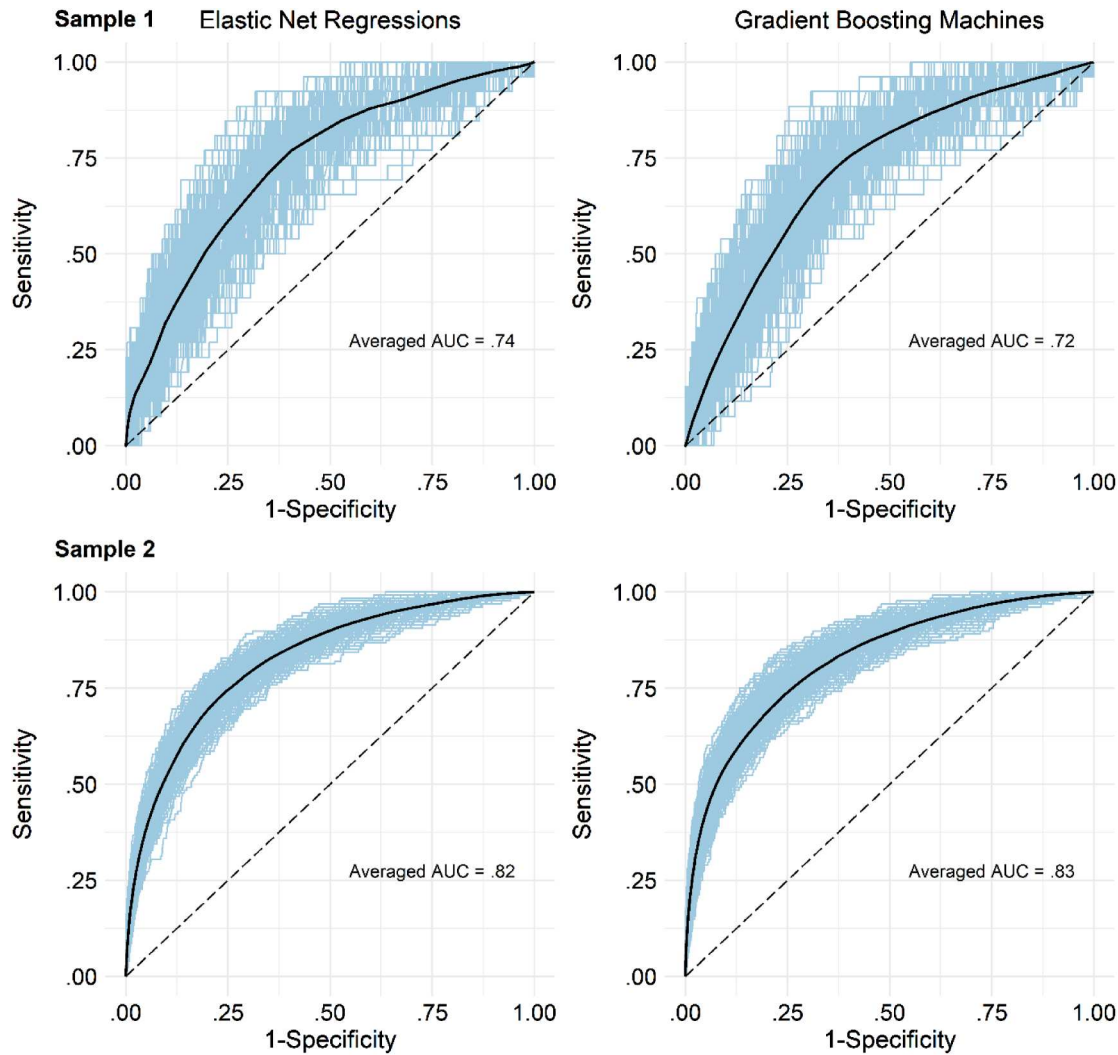
Figure 1. ROC curves for each of the 300 testing iterations including an averaged ROC curve
Note: For the averaged ROC curve (black bold line), ROC curve values were averaged across 50 evenly distributed cutpoints

to those who completed their therapy, which is to be expected given the low percentage of overall dropouts in the samples. Positive Predictive Values (PPV) and Negative Predictive Values (NPV) were overall similar across algorithms: The relatively low PPV for all algorithms indicate that only 15-17% of all flagged patients were actual dropouts, whereas the relatively high NPV of .94-.98 indicate that nearly all patients identified as completers were correctly classified. Given the relatively low prevalence of therapy dropouts in both samples, these patterns were to be expected since PPV and NPV are very sensitive to imbalanced data (see also Belsher et al., 2019).

Which Variables Predict Therapy Dropout?

Figure 3 shows the ten most important predictor variables for the machine learning models (i.e.,

elastic net regressions and GBM). In Sample 1, the mean score of the Helping Alliance Questionnaire (reported by the respective therapist at baseline in the initial interview) was by far the most important predictor across both modeling approaches, followed by the patients' age. Both variables had a negative impact, that is, a good helping alliance and higher age were associated with a lower probability of dropping out. Further, albeit significantly less, important variables of the elastic net models were whether the patients had any affective/mood disorder diagnosis (section F3 in ICD-10), variables concerning their current employment, and satisfaction with their financial situation. Overall, elastic net regressions and GBM had five out of the ten most important predictors in common. However, due to their relatively low and small differences in variable importances, rankings beyond the two most important variables should be taken with a grain of salt.
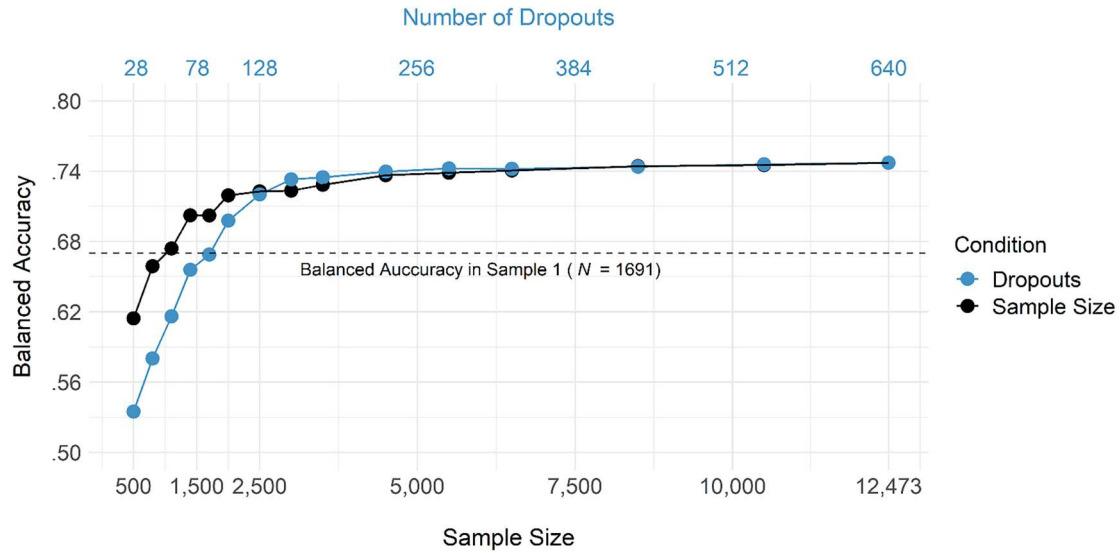
Figure 2. Balanced accuracies in sample 2 depending on the sample size and events (dropouts) fraction

Note: In the Sample Size condition (black line), the overall sample size was reduced stepwise (full sample, 10,500, 8,500, 6,500, 5,500, 4,500, 3,500, 2,500, 2,000, 1,700, 1,400, 1,100, 800, 500) but the events fraction was kept at the same level as in the full sample (N=5.13%). For the Dropouts condition (blue line), we always used all completers (N=11,833) and added 5.13% of different sample size levels as dropouts, thereby reducing only the number of dropouts to that of the Sample Size condition. Ultimately, this reduces the events fraction (5.13, 4.36. 3.55, 2.75, 2.33, 1.91, 1.50, 1.28, 1.07, .0.86, 0.73, 0.60, 0.47, 0.35, 0.22). Importantly, the number of dropouts is identical in both conditions at a given point on x-axis
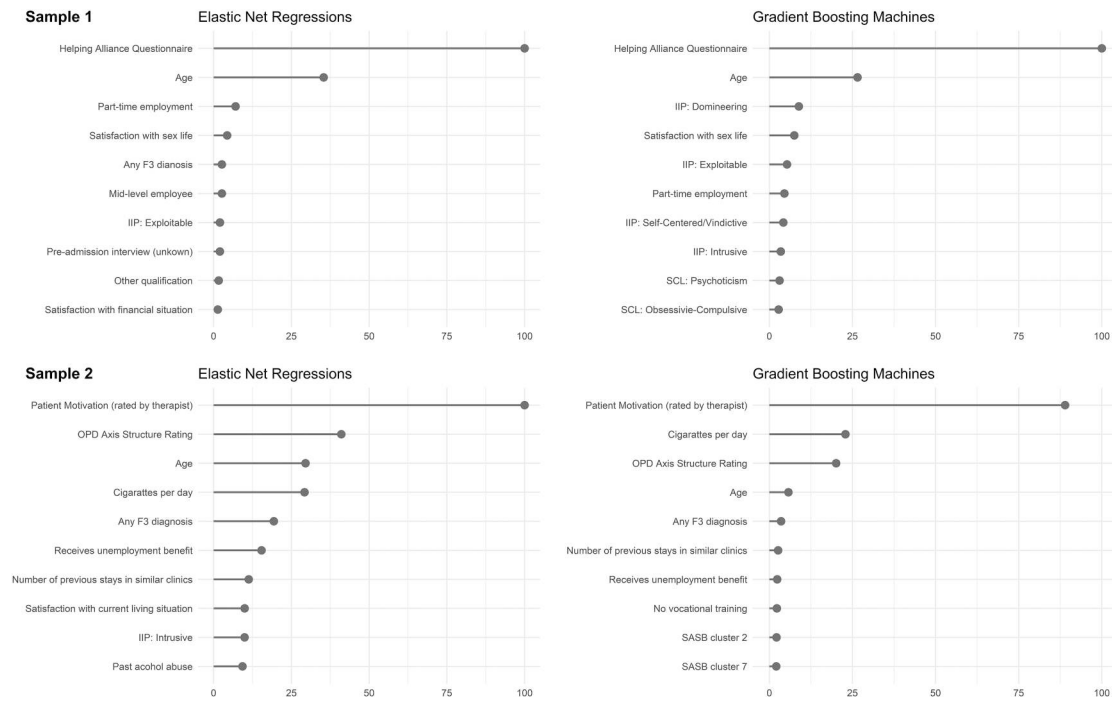

Figure 3. Averaged variable importances for the ten most important predictor variables
Note. N=1,691 (Sample 1) and N=12,473 (Sample 2)

In Sample 2, patients' motivation toward therapy as rated by the therapist was the most important variable across algorithms. Hence, the initial evaluation after the first interview was the most important predictor of eventual therapy dropout across samples and algorithms. Similar to the results of Sample 1, higher age was associated with a lower probability of dropping out for Sample 2 as well.

Apart from that, higher therapist-rated impairments in structural integration (i.e., overall rating on the axis IV of the operationalized psychodynamic diagnostics, e.g., Zimmermann et al., 2012), a higher number of cigarets per day, having any F3 diagnosis, receiving unemployment benefits and more previous stays in similar clinics were among the most important variables for both algorithms. Comparing both samples, it should be noted that there was no information on the level of structural integration rating or on smoking available in Sample 1, so one cannot infer that these variables were not relevant in Sample 1.

Excluding the most important predictor from the full model (the initial impression regarding patients' motivation or helping alliance as assessed by the therapist) led to a substantial drop in overall predictive performance for both samples (ΔABA = .08). Using only the helping alliance as rated by the therapists in Sample 1 in a logistic regression model resulted in model performance that was nearly as good as for the full elastic net regression model (single predictor model in Table II). Notably, this model outperformed even the highly overfitted logistic regressions that included all variables. In Sample 2, a model using solely patients' motivation as a predictor has a slightly lower model performance than a model employing the full predictor set (ΔABA = .02 and ΔAUC = .05), but there were larger changes in the sensitivity = -.07 and specificity = + .30. Because sensitivity is arguably more important in identifying potential dropouts than patients who are regularly continuing their therapy, using all available predictors has incremental value over a simpler model based on the therapists' initial assessment of patients (Sample 2). Comparing the predictor set in both samples, there is an interesting distinction: The therapist-rated motivation variable (Sample 2) seems to detect completers better than dropouts, whereas the helping alliance (Sample 1) seems to discriminate better for higher risk individuals (as indicated by the higher decrease of sensitivity than specificity if the variable is dropped), which suggests that combining both aspects may lead to an improved balanced accuracy.

Discussion

High rates of therapy dropout are disadvantageous for individual patients and the health care system. Thus, it is worthwhile to gain more insights into which patients are at a higher risk of subsequent dropout ultimately aiming to prevent dropout. In the current study, it was possible to correctly classify 67% (Sample 1) and 75% (Sample 2) of all patients into either therapy dropouts or completers using machine learning prediction models including only baseline indicators and therapists' impressions formed directly at the beginning of the therapy of large naturalistic inpatient samples. As it is common for classification tasks with highly unequal group sizes (e.g., Belsher et al., 2019; Jankowsky et al., 2023), we found that members of the majority group, that is therapy completers, could be predicted more accurately. Compared to Bennemann et al. (2022) who predicted therapy dropout in a German outpatient sample, AUC were higher in our analyses (.74/.83 vs. .66) which might be attributed to several reasons such as the different dropout rations, predictor variables, length of therapies, or settings (inpatient vs. outpatients).

The usefulness of any statistical model should ultimately be evaluated with a cost–benefit calculation. Although treatment response is influenced by multiple, time-variant factors emphasizing the need for multi-modal and longitudinal data (e.g., Chekroud et al., 2021), models such as the present ones can easily be implemented because they mostly rely on the baseline assessment that is carried out in German psychotherapy clinics on a routine basis. Hence, using prediction models such as the present ones provides valuable information on the risk of dropout for an individual patient early in the therapeutical process when intervention is still possible without imposing additional burden on therapists or patients.

The Role of Therapists' Impressions for the Prediction of Therapy Dropout

We found that the initial impression regarding patients' helping alliance by the respective therapist was the most important predictor of eventual therapy dropout in Sample 1, that is, a low-quality alliance was associated with a higher probability of dropout. Previous research showed similar effects: In a recent meta-analysis covering 295 studies including more than 30,000 patients, the estimated correlation between alliance and therapy dropout was $r = .18$ (Flückiger et al., 2018). In principle, the importance of alliance is encouraging news, because early intervention would be possible in the event of an unproductive alliance and therapist could act on their initial assessment. In an inpatient setting, a low initial alliance score might indicate some form of disagreement or tension regarding therapy goals or the type of the planned treatment, which could be revealed with brief self-report assessments of so-called alliance ruptures after each session. Repairing such alliance ruptures has been

shown to have a positive effect on treatment response (e.g., $r = .29$ in a meta-analysis by Eubanks et al., 2018).

In Sample 2, patients' baseline motivation as rated by the therapist was the most important predictor of dropout; highly motivated patients had a lower probability of dropping out. The single-item assessment of patients' motivation presumably taps a similar construct as the Helping Alliance Questionnaire used in Sample 1 (which includes questions such as "I believe that the patient is sufficiently motivated for treatment" or "I have the impression that the patient is affectively responsive to my therapeutic interventions"), again underlining the importance of motivational factors in therapy. Low motivation and/or dissatisfaction with the treatment has also been shown to be one of the main reasons for therapy dropout in outpatients (e.g., Bados et al., 2007) and to moderate the relationship between patient-rated therapeutic alliance and treatment outcomes in a cognitive–behavioral treatment (e.g., Rivera et al., 2023). Also, therapists might be adept at assessing which patients are suited for or able to adjust to the particular treatment concept of an inpatient setting, which might guide the impression in their assessment of patients' overall motivation.

Another important variable for the prediction of dropout in both samples was patients' age, that is, older patients were less likely to drop out of therapy. Older patients were also more often diagnosed with an affective disorder and usually have a higher level of structural integration (Zimmermann et al., 2020) which were both negatively associated with therapy dropout. Hence, the effect of age could be a combination of having a better "treatable" diagnosis and demographic factors such as fewer young children to take care of, a more stable financial situation etc. that would make it more accessible to be in treatment continuously for multiple weeks. In Sample 2, the number of cigarets per day and past alcohol abuse were also relevant predictors of therapy dropout, indicating that addictive behaviors can be at odds with successfully following the structured treatment in inpatient clinics.

On the Requirements of Reasonable Classification Models in Clinical Psychology

There is an ongoing debate on the use(fulness) of machine learning prediction models in clinical psychology (e.g., Wilkinson et al., 2020). With the current study, we added to the critical literature arguing that more complex models allowing for interactions and non-linear effects (such as the GBM) did not always necessarily result in higher predictive accuracy when compared to regularized regression models. More critically, initial findings on the superiority of more complex machine learning models in clinical science have serious statistical flaws such as an incorrect cross-validation procedure (Jacobucci et al., 2021; Kapoor & Narayanan, 2022). Furthermore, such models usually require reliable indicators and large sample sizes to be able to correctly identify complex patterns in the data (e.g., Jacobucci & Grimm, 2020). In the context of medical/clinical classification tasks, however, small samples sizes and a low number of events (= the outcome to be predicted) are the rule, not the exception.

Generally, the required sample size and number of events in classification tasks depend on multiple factors and specific recommendations should always be treated as limited to their respective context. For example, Giesemann et al. (2023) recommended at least 300 patients for training samples when using machine learning algorithms to predict therapy dropout. However, they used less predictor variables (7) and had a higher event fraction (30%) as in this study. Three- or even four-hundred patients within the training sample would be similar to the smallest sample size of our simulation (see Figure 2) and thus clearly not advisable for this study. A common (and often criticized) rule of thumb is to use at least ten events per predictor to avoid biased estimates (e.g., Moons et al., 2014). More recent studies suggest that using regularized regression models allows for a relaxation of this rule (e.g., Pavlou et al., 2016) and that the number of predictors, overall sample size, and events fraction are all necessary to consider (Van Smeden et al., 2019). We underline with the present study the notion that a reduction of the events fraction leads to lower predictive performance even if the overall sample size is large ($N > 11,000$).

Limitations

Several limitations pertain to the setup of the study and the validation of the prediction models. First, no information was available whether and how therapists modified their behavior dependent on their initial impression of their patients' motivation or of the helping alliance. To shed more light on this, future research should use assessments of therapists' characteristics, strategies in dealing with non-optimal helping alliance, and the dynamics of the helping alliance (e.g., Flückiger et al., 2020) over the course of the therapy for the prediction of therapy dropout.

Second, we also were only able to use therapists' ratings of the helping alliance and treatment motivation which does not capture the patients'

perspective. Since it has been meta-analytically shown that patient- and therapist-ratings on alliance are only moderately correlated ($r = .36$; Tryon et al., 2007), the inclusion of the patients' view would be important to get a more nuanced picture. Fortunately, this limitation might be less significant according to studies (on outpatients) evidencing that therapists' variability as opposed to patients' variability in alliance relates to treatment outcomes (e.g., Baldwin et al., 2007).

Third, the two data sets varied in many respects, among others in the composition of the patients, the implemented interventions, and the variables gathered upon admission to the clinic. We think that, on the one hand, the two naturalistic samples have high ecological validity and thus the resulting models are practically relevant, but on the other hand, some differences between the samples make a direct comparison difficult. It would therefore be highly beneficial to standardize baseline assessments across clinics—at least to a common core set of measures—to implement a more rigorous form of validation across independent samples in future studies (Dwyer et al., 2018). Concerning generalizability, it should be noted that treatment concepts and patient characteristics differ between in- and outpatients in Germany, but also internationally, so findings based on German inpatients are partly limited to this specific context. However, apart from the question whether the two clinics included in the present study were representative of German inpatient clinics in general, would like to stress that Sample 2 of this study included as may patients as over two thirds of the 15 studies combined that were included in the most extensive review so far on therapy dropout in German inpatients.

Conclusion

Overall, our study demonstrates that therapy dropout in an inpatient setting can be predicted fairly accurately using baseline indicators and therapist-related variables collected at the beginning of a clinic stay. On a methodological stance, we showed that more complex machine learning algorithms are not necessarily superior over simpler regularized algorithms in clinical datasets evaluating therapy success and underlined the importance of data quality for clinical prediction models. Initial therapist assessments of patients' motivation and the therapeutic alliance emerged as the most crucial predictors of dropout. These findings underscore the importance of first impressions in therapeutic contexts and highlight potential intervention points to reduce dropout rates. By focusing on enhancing patient motivation and strengthening the therapeutic alliance early in the treatment process, clinicians may improve retention and outcomes in psychotherapy. We encourage other researchers to conduct similar studies in different contexts to study if the variables found in this study to be predictive do translate to other settings.

Supplemental Data

Supplemental data for this article can be accessed at https://doi.org/10.1080/10503307.2024.2411985.

Data Availability Statement

To maintain patient's privacy, unfortunately, the data cannot be made publicly available. However, the data can be made available upon request using similar data sharing agreements as within our cooperation.

References

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, *55*(3-4), 521–536. https://doi.org/10.1080/00223891.1990.9674088

Arbeitskreis OPD (Hrsg.). (1996). *Operationalisierte Psychodynamische Diagnostik: Grundlagen und Manual*. Huber.

Bados, A., Balaguer, G., & García, C. S. (2007). The efficacy of cognitive–behavioral therapy and the problem of drop-out. *Journal of Clinical Psychology*, *63*(6), 585–592. https://doi.org/10.1002/jclp.20368

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, *75*(6), 842–852. https://doi.org/10.1037/0022-006X.75.6.842

Barrett, M. S., Chua, W.-J., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy: Theory, Research, Practice, Training*, *45*(2), 247–267. https://doi.org/10.1037/0033-3204.45.2.247

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., Jüni, P., & Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine*, *10*(5), e1001454. https://doi.org/10.1371/journal.pmed.1001454

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbauch, J. (1961). *Beck depression inventory (BDI)*. APA PsycTests. https://doi.org/10.1037/t00741-000.

Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., Morgan, R. L., Evatt, D. P., Tucker, J., & Skopp, N. A. (2019). Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*, 76(6), 642–651. https://doi.org/10.1001/jamapsychiatry.2019.0174

Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review*, 81(5), 392–425. https://doi.org/10.1037/h0037024

Bennemann, B., Schwartz, B., Giesemann, J., & Lutz, W. (2022). Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. *The British Journal of Pyschiatry: The Journal of Mental Science*, 220(4), 192–201. https://doi.org/10.1192/bjp.2022.17

Bucher, M. A., Suzuki, T., & Samuel, D. B. (2019). A meta-analytic review of personality traits and their associations with mental health treatment outcomes. *Clinical Psychology Review*, 70, 51–63. https://doi.org/10.1016/j.cpr.2019.04.002

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, 20(2), 154–170. https://doi.org/10.1002/wps.20882

Derogatis, L. R. (1982). *Brief symptom inventory (BSI) [Database record]*. APA PsycTests. https://doi.org/10.1037/t00789-000.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, 14(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. *Psychotherapy*, 55(4), 508–519. https://doi.org/10.1037/pst0000185

Fernandez, E., Salem, D., Swift, J. K., & Ramtahal, N. (2015). Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators. *Journal of Consulting and Clinical Psychology*, 83(6), 1108–1122. https://doi.org/10.1037/ccp0000044

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55(4), 316–340. https://doi.org/10.1037/pst0000172

Flückiger, C., Rubel, J., Del Re, A. C., Horvath, A. O., Wampold, B. E., Crits-Christoph, P., Atzil-Slonim, D., Compare, A., Falkenström, F., Ekeblad, A., Errázuriz, P., Fisher, H., Hoffart, A., Huppert, J. D., Kivity, Y., Kumar, M., Lutz, W., Muran, J. C., Strunk, D. R., … Barber, J. P. (2020). The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis. *Journal of Consulting and Clinical Psychology*, 88(9), 829–843. https://doi.org/10.1037/ccp0000594

Franke, G. H. (1995). *SCL-90-R: Die Symptom-Checkliste von Derogatis – Deutsche Version*. Beltz Test Gesellschaft.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Giesemann, J., Delgadillo, J., Schwartz, B., Bennemann, B., & Lutz, W. (2023). Predicting dropout from psychological treatment using different machine learning algorithms, resampling methods, and sample sizes. *Psychotherapy Research*, 683–695. https://doi.org/10.1080/10503307.2022.2161432

Gonzalez Salas Duhne, P., Delgadillo, J., & Lutz, W. (2022). Predicting early dropout in online versus face-to-face guided self-help: A machine learning approach. *Behaviour Research and Therapy*.

Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers (2019). *gbm: Generalized boosted regression models* (Version 2.1.5) [Computer software]. https://CRAN.R-project.org/package=gbm.

Griffin, D. W., & Bartholomew, K. (1994). *Relationship scales questionnaire (RSQ)*. APA PsycTests. https://doi.org/10.1037/t10182-000.

Hans, E., & Hiller, W. (2013). A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. *Clinical Psychology Review*, 33(8), 954–964. https://doi.org/10.1016/j.cpr.2013.07.003

Hong, M., Jacobucci, R., & Lubke, G. (2020). Deductive data mining. *Psychological Methods*, 25(6), 691–707. https://doi.org/10.1037/met0000252

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, 15(3), 809–816. https://doi.org/10.1177/1745691620902467

Jacobucci, R., & Li, X. (2022). Does minority case sampling improve performance with imbalanced outcomes in psychological research? *Journal of Behavioral Data Science*, 2(1), 59–74. https://doi.org/10.35566/jbds/v2n1/p3

Jacobucci, R., Littlefield, A. K., Millner, A. J., Kleiman, E. M., & Steinley, D. (2021). Evidence of inflated prediction performance: A commentary on machine learning and suicide research. *Clinical Psychological Science*, 9(1), 129–134. https://doi.org/10.1177/2167702620954216

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning*. Springer.

Jankowsky, K., Steger, D., & Schroeders, U. (2023). Predicting lifetime suicide attempts in a community sample of adolescents using machine learning algorithms. *Assessment*, 31(3), 557–573. https://doi.org/10.1177/10731911231167490

Kamenov, K., Twomey, C., Cabello, M., Prina, A. M., & Ayuso-Mateos, J. L. (2017). The efficacy of psychotherapy, pharmacotherapy and their combination on functioning and quality of life in depression: A meta-analysis. *Psychological Medicine*, 47(3), 414–425. https://doi.org/10.1017/S0033291716002774

Kapoor, S., & Narayanan, A. (2022). Leakage and the reproducibility crisis in ML-based science (arXiv:2207.07048). arXiv. http://arxiv.org/abs/2207.07048.

Karterud, S., Pedersen, G., Bjordal, E., Brabrand, J., Friis, S., Haaseth, O., Haavaldsen, G., Irion, T., Leirvåg, H., Tørum, E., & Urnes, O. (2003). Day treatment of patients with personality disorders: Experiences from a Norwegian treatment research network. *Journal of Personality Disorders*, 17(3), 243–262. https://doi.org/10.1521/pedi.17.3.243.22151

Kuhl, J., & Kazén, M. (2009). Persönlichkeits-Stil-und Störungs-Inventar (PSSI). In *Personality styles and disorder inventory (PSDI)*. Hogrefe.

Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5), https://doi.org/10.18637/jss.v028.i0

Lang, K., Koch, U., & Schulz, H. (2006). Abbrüche stationärer Psychotherapien. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 35(4), 267–275. https://doi.org/10.1026/1616-3443.35.4.267

Leichsenring, F., Steinert, C., Rabung, S., & Ioannidis, J. P. A. (2022). The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry*, 21(1), 133–145. https://doi.org/10.1002/wps.20941

Luborsky, L., Crits-Christoph, P., Alexander, L., Margolis, M., & Cohen, M. (1983). Two helping alliance methods for predicting outcomes of psychotherapy. *The Journal of Nervous and Mental Disease*, 171(8), 480–491. https://doi.org/10.1097/00005053-198308000-00005

Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., Reitsma, J. B., & Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Medicine*, *11*(10), e1001744. https://doi.org/10.1371/journal.pmed.1001744

Pargent, F., Schoedel, R., & Stachl, C. (2023). Best practices in supervised machine learning: A tutorial for psychologists. *Advances in Methods and Practices in Psychological Science*, *6* (3), https://doi.org/10.1177/25152459231162559

Pavlou, M., Ambler, G., Seaman, S. R., De Iorio, M., & Omar, R. Z. (2016). Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Statistics in Medicine*, *35*(7), 1159–1177. https://doi.org/10.1002/sim.6782

Reuter, L., & Scheidt, C. (2014). Vorzeitige Beendigung der Therapie in der stationären psychosomatischen Krankenhausbehandlung und Rehabilitation. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *64*(08), 297–308. https://doi.org/10.1055/s-0034-1368730

Rivera, A. P., Maisto, S. A., Connors, G. J., & Schlauch, R. C. (2023). Therapists' first impression of treatment motivation moderates the relationship between the client-rated therapeutic alliance and drinking outcomes during treatment. *Alcoholism: Clinical and Experimental Research*, *47*(4), 806–821. https://doi.org/10.1111/acer.15040

Schepank, H. (1995). *Der Beeinträchtigungs-Schwere-Score (BSS): Ein Instrument zur Bestimmung der Schwere einer psychogenen Erkrankung*. Beltz Test GmbH.

Schmidt, I. D., Forand, N. R., & Strunk, D. R. (2019). Predictors of dropout in internet-based cognitive behavioral therapy for depression. *Cognitive Therapy and Research*, *43*(3), 620–630. https://doi.org/10.1007/s10608-018-9979-5

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. E., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, *43* (2), 476–493. https://doi.org/10.1093/ije/dyu038

Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *80*(4), 547–559. https://doi.org/10.1037/a0028226

Tryon, G. S., Blackwell, S. C., & Hammel, E. F. (2007). A meta-analytic examination of client–therapist perspectives of the working alliance. *Psychotherapy Research*, *17*(6), 629–642. https://doi.org/10.1080/10503300701320611

Van Smeden, M., Moons, K. G., De Groot, J. R., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research*, *28*(8), 2455–2474. https://doi.org/10.1177/0962280218784726

Wagenmakers, E., Wetzels, R., Borsboom, D., Van Der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wilkinson, J., Arnold, K. F., Murray, E. J., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., Lippert, C., Gilthorpe, M. S., & Tennant, P. W. G. (2020). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, *2*(12), e677–e680. https://doi.org/10.1016/S2589-7500(20)30200-4

Zimmermann, J., Ehrenthal, J. C., Cierpka, M., Schauenburg, H., Doering, S., & Benecke, C. (2012). Assessing the level of structural integration using operationalized psychodynamic diagnosis (OPD): implications for DSM-5. *Journal of Personality Assessment*, *94*(5), 522–532. https://doi.org/10.1080/00223891.2012.700664

Zimmermann, J., Müller, S., Bach, B., Hutsebaut, J., Hummelen, B., & Fischer, F. (2020). A common metric for self-reported severity of personality disorder. *Psychopathology*, *53*(3-4), 168–178. https://doi.org/10.1159/000507377

Zimmermann, D., Rubel, J., Page, A. C., & Lutz, W. (2017). Therapist effects on and predictors of non-consensual dropout in psychotherapy. *Clinical Psychology & Psychotherapy*, *24*(2), 312–321. https://doi.org/10.1002/cpp.2022

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x