Psychotherapy and Psychosomatics

Standard Research Article

Psychother Psychosom
DOI: 10.1159/000521951

Received: September 14, 2021
Accepted: January 6, 2022
Published online: February 14, 2022

Intensity of Treatment as Usual and Its Impact on the Effects of Face-to-Face and Internet-Based Psychotherapy for Depression: A Preregistered Meta-Analysis of Randomized Controlled Trials

Thomas Munder[a] Alessia Geisshüsler[a] Tobias Krieger[b]
Johannes Zimmermann[c] Markus Wolf[a] Thomas Berger[b] Birgit Watzke[a]

[a]Department of Psychology, University of Zurich, Zurich, Switzerland; [b]Institute of Psychology, University of Bern, Bern, Switzerland; [c]Department of Psychology, University of Kassel, Kassel, Germany

Keywords
Depression · Psychotherapy · Randomized controlled trial · Treatment as usual · Meta-analysis

Abstract
Introduction: Treatment as usual (TAU) is the most frequently used control group in randomized trials of psychotherapy for depression. Concerns have been raised that the heterogeneity of treatments in TAU leads to biased estimates of psychotherapy efficacy and to an unclear difference between TAU and control groups like waiting list (WL). ***Objective:*** We investigated the impact of control group intensity (i.e., amount and degree to which elements of common depression treatments are provided) on the effects of face-to-face and internet-based psychotherapy for depression. ***Methods:*** We conducted a preregistered meta-analysis (www.osf.io/4mzyd). We included trials comparing psychotherapy with TAU or WL in patients with symptoms of unipolar depression. Six indicators were used to assess control group intensity. Primary outcome: Standardized mean difference (SMD) of psychotherapy and control in depressive symptoms at treatment termination. ***Results:*** We included 89 trials randomizing 14,474 patients to 113 psychotherapy conditions and 89 control groups (TAU in 42 trials, WL in 47 trials). Control group intensity predicted trial results in preregistered (one-sided *p*s < 0.042) and exploratory analyses. Psychotherapy effects were significantly smaller (one-sided $p = 0.002$) in trials with higher intensity TAU (SMD = 0.324, CI 0.209 to 0.439) than in trials with lower intensity TAU (SMD = 0.628, CI 0.455 to 0.801). Psychotherapy effects against lower intensity TAU did not differ from effects against WL (two-sided $p = 0.663$). ***Conclusions:*** Our results suggest that variation in TAU intensity impacts the outcome of trials. More scrutiny in the design of control groups for clinical trials is recommended.

Introduction

Psychotherapy is a first-line specialized treatment for depressive disorders according to treatment guidelines [1, 2], and has a large database that documents its efficacy [3, 4]. However, findings of a substantial proportion of nonresponders to psychotherapy [5], of a high risk for relapse for some subgroups [6], and of limited access to psychotherapy [7, 8] necessitate further research efforts

Karger@karger.com
www.karger.com/pps

Correspondence to:
Thomas Munder, thomas.munder@uzh.ch

on how existing treatment approaches can be improved and supplemented. New or improved treatments have to demonstrate their efficacy, typically by being tested in a randomized controlled trial, where they are compared to one of several control groups and/or an alternative treatment [9]. According to a recent systematic review [4], treatment as usual (TAU) is the most frequently used control/comparison group in trials of psychotherapy for depression. In TAU – sometimes referred to as care as usual, standard care, or routine care – participants receive existing treatments accessible in their community or health care system [10–12]. While control groups like waiting list (WL) or psychological placebo control for spontaneous remission and/or nonspecific effects, the concept of TAU is to additionally indicate if a new treatment is as good or better than current standard treatments (i.e., the *comparative* efficacy of new treatments) [13, 14].

Despite the widespread use of TAU, several methodological commentators have voiced concerns that their use is problematic [11, 13, 15–20]. The most important issue is that the treatments provided in TAU vary substantially across trials, ranging from current best practice treatments to minimal services [13], to substantial shares of participants receiving no treatment at all [21]. In addition, the nature of treatments received in TAU is often unclear because TAU is neither standardized nor monitored [18]. One resulting concern is that the results of TAU-controlled trials are difficult to interpret because it remains unclear whether benefits of new treatments are due to "specific treatment ingredients […] or to nonspecific factors" [18, p. 278] or "simply to the fact that participants in that condition actually got treatment" [16, p. 534]. Another concern is that this variability in TAU introduces bias in meta-analyses which use TAU to indirectly compare different treatment approaches [17].

Some meta-analyses have investigated potential bias from TAU heterogeneity. Watts and colleagues [22] found that the effects in 48 trials of cognitive-behavior therapy for anxiety or depression differed across five categories of TAU referring to different treatments or service providers. These results were partially replicated by a recent meta-analysis of 140 trials of psychotherapy for depression [14]: while psychotherapy effects did not differ significantly across five TAU categories referring to different care settings (e.g., primary care or specialized care), effects differed within TAU categories depending on country, suggesting variations in care among health care systems. Using a continuous approach, Spielmans and colleagues [23] related a score of quantitative and qualitative treatment aspects in TAU to the results of 34 TAU-controlled trials of evidence-based psychotherapy for youths. Psychotherapy effects were found to decrease, the more treatment aspects were delivered in TAU. Together, these meta-analyses provide tentative support for an impact of TAU heterogeneity on trial results, but robust evidence is still lacking.

These previous meta-analyses suggest that further research should focus on the *intensity* of TAU, rather than on the setting in which TAU is provided. The intensity of TAU (and other control groups) in depression research can be conceptualized as the amount and degree to which specific or nonspecific elements of common depression treatments are provided to participants in control groups [12, 24]. Consequently, higher control group intensity should, on average, lead to better clinical responses of participants in the control group. Also, higher control group intensity should increase the stringency of a test against this control group (i.e., smaller benefits of treatment over control are expected) [12]. Table 1 summarizes crucial aspects of TAU intensity as discussed in the literature [11–13, 16, 17, 23].

We aimed to investigate (a) the possible effects of TAU intensity on the results of randomized trials of psychotherapy for adult depression. We further aimed to investigate (b) if TAU and WL control groups differed in their intensity, and (c) if trials of face-to-face psychotherapy (F2F) and internet-based psychotherapy (INT) differed in the intensity of their control groups. We conducted a prospective and preregistered systematic review and meta-analysis.

Materials and Methods

A study protocol with preregistration of variable definitions, hypotheses, and analytic strategy was registered with OSF (www.osf.io/4mzyd). Our reporting follows the PRISMA guideline [25] (for checklist see online suppl. material, available at www.karger.com/doi/10.1159/000521951).

Inclusion Criteria

We included randomized trials that contrasted (a) outpatient F2F (individual or group) or INT (guided or self-guided) psychotherapy with (b) a TAU or WL, in (c) adults with elevated symptoms of depression (including subthreshold depression, first-time and recurrent major depressive disorder, and dysthymia/persistent depressive disorder). Although the focus of this study is on TAU control groups, we included WL-controlled trials to be able to investigate the empirical differentiation and overlap of TAU and WL. We excluded (a) trials focused on relapse prevention, and (b) trials in which all patients shared a comorbid mental disorder (e.g., substance use disorder), somatic disorder (e.g., heart surgery), or social

Table 1. Tentative criteria for the intensity of TAU control groups in trials of psychotherapy for depression

Criterion	Higher intensity of TAU according to criterion
Any depression treatment	All participants in TAU receive some treatment for depression
No restrictions on health care	Participants in TAU are allowed unrestricted access to all common treatments for depression
Evidence base	Treatments in TAU are evidence-based or in accordance with treatment guidelines
Treatment dose	Treatments in TAU are delivered in typical doses
Professional background of providers	Treatments in TAU are provided by mental health professionals
Training and supervision of providers	Providers delivering TAU have access to training or supervision

TAU, treatment as usual.

problem (e.g., unemployment). We excluded the latter trials to ensure that TAU targeted depression and participants were seeking help for depression. We further excluded (c) trials investigating effects of a treatment component rather than a stand-alone treatment, (d) trials which focused on bipolar affective disorders, and (e) trials published before the year 2000. We excluded earlier studies because INT had not been introduced in earlier decades and the comparability of psychotherapies across decades has been questioned [26, 27]. We imposed no language restrictions on reports.

Literature Search

We (a) searched PsycINFO database for meta-analyses on psychotherapy for depression published from 2015 to January 3, 2020 to screen reference lists for relevant trials. Furthermore, (b) we searched PsycINFO and the Cochrane's Central Register of Controlled Trials (CENTRAL) from 2018 to January 3, 2020 to identify newer trials. We used search terms relating to psychotherapy, depression, and randomized trial methodology. Search in CENTRAL was limited to entries from EMBASE and PubMed. Search in PsycINFO was limited to entries from peer-reviewed journals. See online supplementary Table S1 for the search terms used in both searches.

Variable Definitions and Coding

Coding rules for all variables were defined in a preregistered codebook (www.osf.io/4mzyd). Two independent coders extracted all information. Disagreements were resolved through consensus. Moderating/confounding variables (risk of bias, depression severity, treatment rationale, and number of sessions/modules) are described in online supplementary note S1.

Control Group Intensity

Intensity of treatments used in TAU and WL control groups was assessed with six independent dichotomous indicators derived from previous works [11–13, 16, 17, 23]. Indicators were summed to create a continuous index (range 0 to 6) to express control group intensity, with higher scores reflecting higher intensity. Specifically, we extracted if all patients in control groups (a) received some form of treatment for depression, and (b) had unconstrained access to usual care. We further extracted if control group participants who received treatment (c) received specialized care for depression, (d) were treated by mental health professionals, (e) were treated by providers with access to training and/or supervision, and (f) received a minimum treatment dose. If information regarding any of these indicators was missing or unclear, we treated the indicator as not fulfilled. Inter-rater reliability for the index was 0.75 (two-way mixed-effects intra-class correlation, absolute agreement, average for two coders).

Treatment Outcome

Primary outcome was depressive symptoms at termination expressed as standardized mean difference (SMD). Positive SMDs indicate more change in the psychotherapy group compared to the control group. We extracted all reported patient- and observer-rated depression instruments. Secondary outcome was depressive symptoms at 6-month follow-up.

Statistical Analyses

We used non-parametric methods to test for differences in control group intensity between F2F, INT, and their four subtypes (i.e., individual F2F, group F2F, self-guided INT, and guided INT). We calculated Hedges' g as a measure of SMD. M and SD were preferred for calculations; if unavailable, we approximated M from change scores, and SD from SE or CI, or extracted SMDs as reported in trial reports. We preferred to calculate SMDs based on intention-to-treat analyses. We used inverse-variance weighted, random-effects meta-analytic methods, as implemented in STATA's "meta" command. We aggregated multiple depression instruments within trials using R package "MAd" assuming a correlation of $r = 0.50$ between instruments [28]. In studies with more than one eligible intervention we adjusted standard errors [29]. We used "meta" default settings which included restricted maximum likelihood estimation and significance tests based on the normal distribution. We used meta-regression to estimate the effect of control group intensity on outcome and introduced type of control group as the first predictor to adjust for possible differences between TAU and WL. Significance tests for the effect of control group intensity were one-sided at alpha = 0.05, all other tests were two-sided at alpha = 0.05.

Results

Included Trials

We retrieved 473 references from 87 meta-analyses and 4,345 references from bibliographic databases. We screened 3,666 unique references, 530 full texts, and finally included 89 eligible trials (see online suppl. Fig. S1

for PRISMA flowchart and online suppl. note S2 for references of trial reports). In these trials, 14,474 patients were randomized to 113 eligible psychotherapy arms (8,284 patients) and 89 control arms (6,190 patients). TAU was used as control group in 42 trials (47.19%) and WL in 47 trials (52.81%). Five trial reports (5.62%) described the control group employed both as TAU and WL (see online suppl. Table S2). We classified these 5 control groups as TAU. F2F was investigated in 37 trials (42.05%; individual F2F: 20 trials, 22.47%; group F2F: 18 trials, 20.22%) and INT in 51 trials (57.95%; self-guided INT: 29 trials, 32.58%; guided INT: 28 trials, 31.46%). No trial contrasted F2F with INT. One trial investigated a combination of F2F and INT (see online suppl. note S2 for reference) and was excluded from modality-specific analyses. Only 16 trials (17.98%) reported outcomes at 6-month follow-up. Descriptive information for all trials is provided in online supplementary Table S2.

Differences in Control Group Intensity

Control groups had an average intensity of 0.58 (SD = 1.05, range 0 to 5), i.e., the average control group fulfilled 0.58 of the six intensity indicators. Fifty-nine trials (66.29%) fulfilled none of the intensity indicators, 17 trials (19.10%) fulfilled 1 indicator, 9 trials (10.11%) fulfilled 2 indicators, and 4 trials (4.71%) fulfilled more than 2 indicators. Intensity of TAU (M = 1.00, SD = 1.23) and WL (M = 0.21, SD = 0.69) differed significantly (SMD = 0.56, $p < 0.001$ from Mann-Whitney U test), with TAU being more intense than WL. Control group intensity did not differ significantly between F2F trials (M = 0.86, SD = 1.42) and INT trials (M = 0.39, SD = 0.63, SMD = 0.30, $p = 0.255$ from Mann-Whitney U test). No significant difference in intensity was found between the control groups of 80 trials investigating only one of four modality subtypes ($p = 0.108$ from Kruskal-Wallis test). See online supplementary Tables S3 and S4 for descriptive information on control group intensity and intensity indicators.

Preregistered Meta-Regressions

Control Group Intensity as Predictor of Outcome

Across all 89 trials, the effect of psychotherapy vis-à-vis control groups was SMD = 0.701 (95% CI 0.563 to 0.839, $p < 0.001$, $I^2 = 93.77\%$). Psychotherapy effects were similar in trials with TAU and WL control groups ($p = 0.354$, see Table 2). For preregistered meta-regressions, we transformed control group intensity scores ≥2 to 2 (possibility of transformation was preregistered and type of transformation was determined before any analysis based on the observed distribution described in the previous section). Control group intensity was a significant predictor of psychotherapy effects (one-sided $p = 0.042$), with effects being smaller in trials with higher control group intensity compared to trials with lower control group intensity (see Table 2). Control group intensity was no significant predictor of psychotherapy effects in 16 trials reporting 6-months follow-up assessments (see online suppl. Table S5).

Sensitivity Analyses

We tested the robustness of control group intensity effects in three preregistered sensitivity analyses. Excluding four outlying trials (see online suppl. note S2 for references) with SMDs larger than 2.00, significant effects for control group intensity were found (see Table 2), with higher control group intensity being related to smaller psychotherapy effects. Heterogeneity was reduced ($I^2 = 75.37\%$). Furthermore, we tested for the presence of small-study effects and Egger's test found smaller trials to show larger effects for psychotherapy ($p < 0.001$; see online suppl. Fig. S2 for a funnel plot). To test for a potential impact of small-study effects, we recalculated the meta-regression in 40 trials with total sample sizes ≥100 and again found a significant effect for control group intensity (see Table 2). Heterogeneity was high with $I^2 = 95.54\%$. We further explored the possible influence of risk of bias, as defined by Cochrane's revised risk of bias tool [30, 31] (see online suppl. note S1). Twelve trials (13.48%) were judged to have low risk of bias, while for 77 trials (86.52%) there was either some concern or high risk of bias. Risk of bias was not found to predict treatment outcome (see Table 2), and again higher control group intensity predicted smaller psychotherapy effects. Heterogeneity remained high ($I^2 = 93.04\%$).

Exploratory Analyses

To reduce heterogeneity, we excluded outliers, as defined above, from all exploratory analyses.

Stratification for Type of Control Group

We analyzed the effects of control group intensity separately in trials with TAU or WL control groups and found stable effects of intensity in TAU trials ($N = 40$, 2 outliers excluded, $B = -0.158$, SE = 0.061, one-sided $p = 0.005$, $I^2 = 76.99\%$) but not in WL trials ($N = 45$, 2 outliers excluded, $B = 0.120$, SE = 0.150, $p = 0.426$, $I^2 = 71.60\%$). Variability of intensity scores in WL trials was small (41 trials, 87.23%, had a score of 0, 4 trials, 8.51%, of 1, and 2 trials, 4.26%, of 2 or more). Because of our primary interest in TAU, we excluded WL trials from further analyses.

Table 2. Preregistered meta-regressions of control group intensity as a predictor of the effect of psychotherapy for depression

Block	Predictor variable	*B*	SE	*p*	I^2
Main analysis (89 trials)					
1					93.68%
	Intercept	0.699	0.070	0.000	
	Control group	–0.065	0.070	0.354	
2					93.41%
	Intercept	0.790	0.088	0.000	
	Control group	–0.003	0.078	0.973	
	Intensity	–0.185	0.107	0.042[a]	
Outliers excluded (85 trials)					
1					76.74%
	Intercept	0.559	0.038	0.000	
	Control group	–0.113	0.038	0.003	
2					75.31%
	Intercept	0.615	0.048	0.000	
	Control group	–0.071	0.044	0.106	
	Intensity	–0.114	0.060	0.029[a]	
Only trials with total N ≥ 100 (40 trials)					
1					95.97%
	Intercept	0.516	0.098	0.000	
	Control group	–0.011	0.098	0.912	
2					95.54%
	Intercept	0.654	0.115	0.000	
	Control group	0.092	0.106	0.386	
	Intensity	–0.293	0.141	0.019[a]	
Adjustment for risk of bias (89 trials)					
1					93.37%
	Intercept	0.749	0.076	0.000	
	Control group	–0.054	0.070	0.436	
	Risk of bias	–0.338	0.196	0.085	
2					93.04%
	Intercept	0.845	0.092	0.000	
	Control group	0.011	0.078	0.887	
	Risk of bias	–0.350	0.193	0.070	
	Intensity	–0.192	0.105	0.035[a]	

[a] One-sided test. Control group = treatment as usual (1), waiting list (–1). Intensity = transformed control group intensity score ranging from 0 (lowest intensity) to 2 (highest intensity). Risk of bias = low risk of bias (1), some concerns or high risk of bias (0).

Meta-Regression with a Dichotomous Intensity Score

We tested the assumed linearity of intensity effects by entering dummy variables for intensities of 1 and ≥2 in a meta-regression of 40 TAU trials (2 outliers excluded) and found both predictors to be significant with similar effect sizes (TAU intensity = 1: $B = –0.278$, SE = 0.116, $p = 0.017$; TAU intensity ≥2: $B = –0.300$, SE = 0.124, $p = 0.015$), which suggested no linear effect. We therefore deemed a dichotomous intensity score (TAU intensity = 0 versus TAU intensity ≥1) to be more adequate and used this in all further analyses, thereby distinguishing trials with lower and higher intensity TAU.

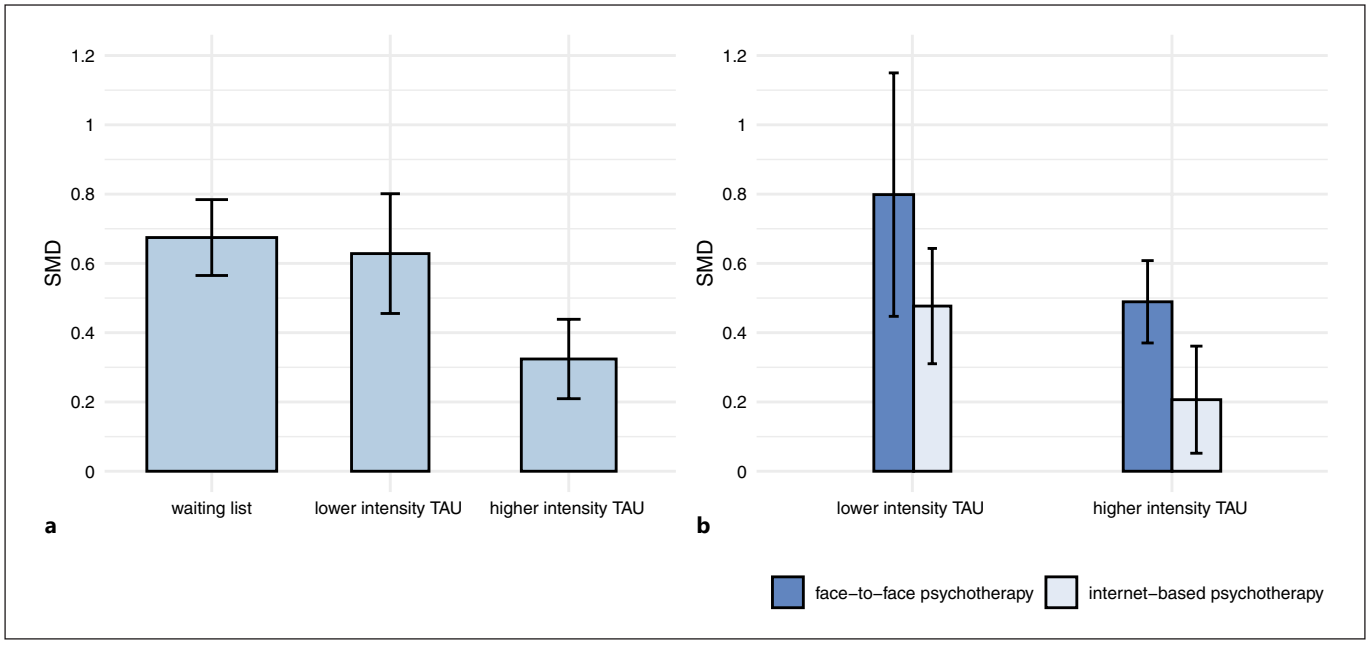

Fig. 1. Efficacy of psychotherapy for depression in randomized trials with lower intensity treatment as usual (TAU) or higher intensity TAU as control group. Standardized mean differences (SMD) are derived from random-effects meta-analyses and express the relative efficacy of psychotherapy and control at post-assessment. Error bars represent 95% confidence intervals. Analyses are based on 85 trials (four outliers excluded). Bar width reflects number of trials. **a** Psychotherapy efficacy in trials with lower intensity TAU, higher intensity TAU, and waiting list. **b** Separate efficacy estimates for face-to-face psychotherapy and internet-based psychotherapy.

Psychotherapy Effects in Trials with Higher Intensity TAU, Lower Intensity TAU, or WL

Psychotherapy effects were significantly smaller in 24 trials with higher intensity TAU (SMD = 0.324, 95% CI 0.209 to 0.439, $p < 0.001$, $I^2 = 72.43\%$) than in 16 trials with lower intensity TAU (SMD = 0.628, 95% CI 0.455 to 0.801, $p < 0.001$, $I^2 = 81.67\%$, 2 outliers excluded), corresponding to a difference of SMD = –0.287 (SE = 0.100, one-sided $p = 0.002$, $I^2 = 76.13\%$) (see online suppl. Fig. S3 for a forest plot). Psychotherapy effects in trials with lower intensity TAU were similar to effects in 45 trials with WL (SMD = 0.675, 95% CI 0.565 to 0.785, $p < 0.001$, $I^2 = 71.21\%$, 2 outliers excluded) (p for difference = 0.663). Figure 1a shows psychotherapy effects in trials with higher intensity TAU, lower intensity TAU, or WL. To determine if TAU intensity could be confounded with risk of bias, we examined the distribution of risk of bias and found the 7 TAU trials (17.50%) with low risk of bias to be equally distributed across trials with lower intensity and higher intensity TAU (2 outliers excluded, p from Fisher's exact test = 0.999).

Robustness across Treatment Modalities and Further Sensitivity Analyses

We further explored the robustness of TAU intensity effects across psychotherapy modalities and their subtypes by introducing them as predictors in two further meta-regressions. TAU intensity emerged as a significant predictor both in a meta-regression that included F2F vs. INT (39 TAU trials, 2 outliers excluded, one trial of blended psychotherapy excluded; TAU intensity: $B = –0.249$, SE = 0.097, one-sided $p = 0.010$; $I^2 = 72.97\%$) and in a meta-regression with four modality subtypes (37 TAU trials, 2 outliers excluded, 1 trial of blended psychotherapy excluded, 2 trials with more than one modality subtype excluded; intensity: $B = –0.229$, SE = 0.099, one-sided $p = 0.011$, $I^2 = 71.77\%$) (see online suppl. Table S6 and Table S7). Figure 1b shows the effects of F2F and INT in trials with lower intensity and higher intensity TAU. The effect of TAU intensity remained robust in further sensitivity analyses including depression severity, treatment rationale, and number of sessions/modules as covariates, and using a different meta-analytic model (see online suppl. Tables S8–S10 and online suppl. note S3).

Discussion

This study set out to investigate the intensity of TAU and WL control groups (i.e., the degree to which quantitative and qualitative aspects of relevant treatments were provided) and its possible impact on the results of randomized trials of psychotherapy for adult depression. In 89 trials comparing F2F or INT psychotherapy with TAU or WL, we found control group intensity to predict psychotherapy effects, with psychotherapy effects decreasing with increasing control group intensity. The effect was strongest for TAU control groups: effects of psychotherapy were 0.287 standard deviation units smaller in trials with higher intensity TAU than in trials with lower intensity TAU, which corresponds to a small to moderate, and thus, clinically relevant effect size [32]. The effect for control group intensity was robust across a series of preregistered and exploratory sensitivity analyses, including tests for the impact of outliers, small-study effects, risk of bias, psychotherapy modality (i.e., F2F vs. INT, and their subtypes), depression severity, type of treatment, number of sessions/modules, and meta-analytic model. The finding is in line with a meta-analysis in which aspects of quality and quantity of care in TAU were related to effects of psychotherapy for youths [23], and compatible with previous research on the impact of TAU heterogeneity [13, 14, 22, 33]. In a wider context, our finding corresponds with meta-analytic findings from behavioral medicine in which higher intensity TAU predicted better health benefits for TAU participants in trials on smoking cessation and adherence to HIV medication [34–36].

We found the intensity of TAU control groups to be low for the majority of trials. For example, only 9 out of 42 TAU trials documented that all participants in TAU received some kind of active depression treatment, and only 12 out of 42 TAU trials stated that participants in TAU had unconstrained access to depression care. In fact, in many trials the intensity of treatments provided in TAU appeared to be as low as in WL control groups. Consistent with this, the effects of psychotherapy in trials with low intensity TAU were closely similar to the effects of psychotherapy compared to WL. Similar effects of psychotherapy for depression in TAU and WL trials have also been reported in a previous meta-analysis [33].

We found no significant difference in control group intensity between F2F and INT trials, which possibly suggests that the efficacy estimates for these modalities are only weakly confounded. However, the statistical power to detect differences for this comparison was low; therefore, more research into this issue is needed. Furthermore, small differences in control group intensity between psychotherapy modalities do not rule out the more general issue that intensity differences in TAU groups might be correlated with intervention categories.

Our meta-analysis has a number of strengths. All key methods and analytic decisions were preregistered, thereby following recommendations for ensuring replicability of research [37]. Furthermore, we employed stringent inclusion criteria and excluded trials in which psychotherapy was specialized for specific target groups or in which psychotherapy or TAU focused on comorbid conditions. By including trials published in the year 2000 or later our results inform on contemporary treatments and trial methods. Additionally, our study included prominent psychotherapy modalities used today, including digital interventions and group therapy. A number of limitations and possible sources of imprecision need to be discussed. In many included trials, the reporting on treatments received in TAU and WL was incomplete, which led to a substantial number of unclear codes that we conservatively interpreted as non-utilization of treatments. This might have led to the misclassification of some trials with higher control group intensity. However, most likely this would have led to an underestimation of the effect of control groups intensity and, thus, does not pose a severe threat to the validity of our findings. Also, risk of bias was highly prevalent in the included trials and there was evidence of publication bias. However, we found no indication that our results were biased by one of these factors. Residual heterogeneity in all analyses was substantial, even after accounting for sources of bias and possible moderator variables. Thus, it is possible that factors not included in this investigation, such as conflict of interest [38], account for additional variation. In addition, the generalizability of our findings is complicated by the clinical heterogeneity of the included samples, treatments, and settings, which is a recurring problem in meta-analyses [39]. We tried to account for this by extensive moderator analyses but we cannot rule out that further relevant moderators have not been captured. Furthermore, it should be added that the value of average effect estimates from meta-analyses for clinical decision making (i.e., treatment selection for the individual patient) is limited because estimates do not take into account the complex and cumulative interplay of specific treatments with characteristics of patients (e.g., comorbidities, treatment history, iatrogenic effects), of settings and of providers, in the determination of clinical outcomes [40, 41].

Our results, together with previous conceptual and empirical work, suggest that TAU, as currently used in

trials of psychotherapy for depression, is an inadequate control group. While conceptually often understood as a comparison with current clinical practice [14, 17], our and others' findings [14, 21] suggest that trials often operationalize TAU as a minimal to no treatment control group, signaling low construct validity and substantial threats to the internal validity of TAU-controlled trials. Consequently, in contrast to their promise, TAU-controlled trials often do not allow conclusions regarding the potential of new treatments to expand or optimize current standard treatments.

While there is a need for further research to replicate our findings in trials of other treatments and populations, the problems of TAU seem evident enough to give recommendations for future trials and meta-analyses, many of which echo previous recommendations [11, 18, 19]. (1) As recommended by many, including the CONSORT statement [42, 43], trialists using TAU control groups need to gather and report specific information on the actual treatments delivered in TAU. In efficacy trials, investigators need to exert more control over TAU, which includes use of standardization and assessment of treatment integrity [11, 19]. Usually, this will lead to an increase of TAU intensity (see Table 1) and also raise the costs of trials [18]. Unclear, uncontrolled or lower intensity TAU control groups should be regarded as similar to no treatment and WL control groups. If at all, they should be used in feasibility and preliminary trials (phases I and II). Consequently, these types of TAU-controlled trials should be regarded as insufficient for new treatments to lead to strong recommendations in guidelines. (2) For most efficacy trials (phases II and III) other control and comparison groups should be preferred over lower intensity TAU. A decision framework proposed by several scholars [11, 18] offers helpful guidance: for earlier phase trials, which aim to determine if a new treatment has potential for a larger efficacy trial and which typically are limited in resources, lower intensity control groups such as nonspecific factor control groups [11] or WL seem appropriate. In later phase trials, which aim to test the efficacy of a new treatment in a larger trial, nonspecific factor control groups or comparisons with established treatments seem appropriate. Different types of nonspecific factor control groups varying in intensity and standardization are available and need to be distinguished [11, 18, 19]. In this regard, Guidi et al. [19] have pointed to the utility of clinical management, a type of nonspecific factor control group in which factors like treatment time and therapist contact are held constant. (3) For effectiveness trials (phase IV), which test how a new treatment fares under clinically representative conditions, TAU is the natural control group [16]. In this case, specific information on treatments actually delivered in TAU is essential for trial interpretation, and a minimum of standardization of TAU is required to assure internal validity. There is evidence that TAU control groups used in trials tend to be less intense than TAU as practiced in the real world [17]. To allow valid conclusions it is essential for effectiveness trials that both, the experimental treatment and TAU, are delivered under clinically representative conditions [11]. (4) Meta-analysts should avoid pooling the results of TAU-controlled psychotherapy trials without taking into account intensity differences between TAU groups. This seems especially important for network meta-analysis which requires the similarity of treatment and control categories across trials [44]. Possible options are to differentiate TAU groups with lower and higher intensity, or to categorize TAU into more meaningful subgroups [21, 39]. Doing so will reduce both the risk of biased estimates of psychotherapy efficacy and the risk of confounding in indirect comparisons of different psychotherapeutic approaches.

Acknowledgements

We thank Anna Leana Hostettler and Cveta Mitrovic, both from University of Bern, Switzerland, for their help with data extraction.

Statement of Ethics

No ethical approval was needed for this systematic review and meta-analysis.

Conflict of Interest Statement

None of the authors have any conflicts of interest to declare.

Funding Sources

None.

Author Contributions

T.M. and B.W. conceived the study. All authors contributed to the design of the study. T.M., T.K., M.W., T.B., and B.W. submitted the study protocol. T.M., A.G., and T.K. selected trials. T.M., A.G., T.K., A.L.H., and C.M. extracted data. T.M., A.G., T.K., and

J.Z. verified the data. T.M. and J.Z. analyzed the data. J.Z. and T.M. created figures. All authors interpreted the results. T.M. and B.W. wrote the first draft of the manuscript. All authors had access to all data. T.M., T.K., J.Z., M.W., T.B., and B.W. provided critical input and revisions to draft manuscripts. All authors approved the final manuscript.

Data Availability Statement

Following publication, the full matrix of data used in this meta-analysis (including explanations) will be made available, for anyone who wishes to access, at OSF (https://osf.io/yspr6).

References

1 American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder. 3rd ed.; 2009. Available from: http://psychiatryonlineorg/guidelinesaspx.
2 National Collaborating Centre for Mental Health; National Institute for Health and Clinical Excellence: Guidance. Depression: the treatment and management of depression in adults (Updated edition). Leicester, UK: British Psychological Society; 2010.
3 Cuijpers P, Quero S, Noma H, Ciharova M, Miguel C, Karyotaki E, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry. 2021 Jun;20(2):283–93.
4 Munder T, Flückiger C, Leichsenring F, Abbass A, Hilsenroth MJ, Luyten P, et al. Is psychotherapy effective? A re-analysis of treatments for depression. Epidemiol Psychiatr Sci. 2019 Jun;28(3):268–74.
5 Leichsenring F, Steinert C, Ioannidis JP. Toward a paradigm shift in treatment and research of mental disorders. Psychol Med. 2019 Sep;49(13):2111–7.
6 Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman AT. Prevalence and predictors of recurrence of major depressive disorder in the adult population. Acta Psychiatr Scand. 2010 Sep;122(3):184–91.
7 Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci. 2011 Feb;6(1):21–37.
8 Vasiliadis H-M, Dezetter A, Latimer E, Drapeau M, Lesage A. Assessing the costs and benefits of insuring psychological services as part of Medicare for depression in Canada. Psychiatr Serv. 2017 Sep;68(9):899–906.
9 Onken LS, Blaine JD, Battjes RJ. Behavioral therapy research: a conceptualization of a process. In: Henggeler SW, Santos AB, editors. Innovative approaches for difficult-to-treat populations. Arlington, VA: American Psychiatric Association; 1997. p. 477–85.
10 Borkovec TD. Control groups and comparison groups in psychotherapy outcome research. In: Onken LS, Blaine JD, editors. Psychotherapy and counseling in the treatment of drug abuse. NIDA Res Monogr 104. Rockville, MD: National Institute on Drug Abuse; 1990. p. 50–65.
11 Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychother Psychosom. 2009;78(5):275–84.
12 Freedland KE, Mohr DC, Davidson KW, Schwartz JE. Usual and unusual care: existing practice control groups in randomized controlled trials of behavioral interventions. Psychosom Med. 2011 May;73(4):323–35.
13 Wampold BE, Budge SL, Laska KM, Del Re AC, Baardseth TP, Flückiger C, et al. Evidence-based treatments for depression and anxiety versus treatment-as-usual: a meta-analysis of direct comparisons. Clin Psychol Rev. 2011 Dec;31(8):1304–12.
14 Cuijpers P, Quero S, Papola D, Cristea IA, Karyotaki E. Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: a meta-analysis. Psychol Med. 2021 Mar;51(4):634–44.
15 Barkham M, Lambert MJ. The efficacy and effectiveness of psychological therapies. In: Barkham M, Lutz W, Castonguay LG, editors. Bergin and Garfield's handbook of psychotherapy and behavior change. Hoboken, NJ: Wiley; 2021. p. 135–89.
16 Chambless DL, Hollon SD. Treatment validity for intervention studies. In: Chambless DL, Hollon SD, editors. APA handbook of research methods in psychology. Washington, DC: American Psychological Association; 2012. Vol. 2; p. 529–52.
17 Kazdin AE. Treatment as usual and routine care in research and clinical practice. Clin Psychol Rev. 2015 Dec;42:168–78.
18 Gold SM, Enck P, Hasselmann H, Friede T, Hegerl U, Mohr DC, et al. Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. Lancet Psychiatry. 2017 Apr;4(9):725–32.
19 Guidi J, Brakemeier E-L, Bockting CL, Cosci F, Cuijpers P, Jarrett RB, et al. Methodological recommendations for trials of psychological interventions. Psychother Psychosom. 2018 Sep;87(5):276–84.
20 Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, et al. The Lancet Psychiatry commission on psychological treatments research in tomorrow's science. Lancet Psychiatry. 2018 Mar;5(3):237–86.
21 Michopoulos I, Furukawa TA, Noma H, Kishimoto S, Onishi A, Ostinelli EG, et al. Different control conditions can produce different effect estimates in psychotherapy trials for depression. J Clin Epidemiol. 2021 Apr;132:59–70.
22 Watts SE, Turnell A, Kladnitski N, Newby JM, Andrews G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. J Affect Disord. 2015 Apr;175:152–67.
23 Spielmans GI, Gatlin ET, McFall JP. The efficacy of evidence-based psychotherapies versus usual care for youths: controlling confounds in a meta-reanalysis. Psychother Res. 2010 Mar;20(2):234–46.
24 Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. J Consult Clin Psychol. 2003 Dec;71(6):973–79.
25 Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009 Jul;6(7):e1000097.
26 Johnsen TJ, Friborg O. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: a meta-analysis. Psychol Bull. 2015 May;141(4):747–68.
27 Baldwin SA, Imel ZE. Studying specificity in psychotherapy with meta-analysis is hard. Psychother Res. 2020 Mar;30(3):294–6.
28 Del Re A, Hoyt WT. MAd: meta-analysis with mean differences. R package version 0.8-2; 2014.
29 Higgins JPT, Eldridge S, Li T. Chapter 23 including variants on randomized trials. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane handbook for systematic reviews of interventions. Version 6.2 (updated 2021 Feb): Cochrane; 2021.
30 Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019 Aug;366:l4898.
31 Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. Psychotherapy Res. 2018 May;28(3):347–55.
32 Cuijpers P, Turner EH, Koole SL, Van Dijke A, Smit F. What is the threshold for a clinically relevant effect? The case of major depressive disorders. Depress Anxiety. 2014 May;31(5):374–8.
33 Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, et al. Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. Transl Behav Med. 2014 Dec;4(4):407–23.

34 Black N, Eisma MC, Viechtbauer W, Johnston M, West R, Hartmann-Boyce J, et al. Variability and effectiveness of comparator group interventions in smoking cessation trials: a systematic review and meta-analysis. Addiction. 2020 Sep;115(9):1607–17.
35 de Bruin M, Viechtbauer W, Hospers HJ, Schaalma HP, Kok G. Standard care quality determines treatment outcomes in control groups of HAART-adherence intervention studies: implications for the interpretation and comparison of intervention effects. Health Psychol. 2009 Nov;28(6):668–74.
36 de Bruin M, Viechtbauer W, Schaalma HP, Kok G, Abraham C, Hospers HJ. Standard care impact on effects of highly active antiretroviral therapy adherence interventions: a meta-analysis of randomized controlled trials. Arch Intern Med. 2010 Feb 8;170(3):240–50.
37 Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Nat Acad Sci U S A. 2018 Mar;115(11):2600–6.
38 Munder T, Brütsch O, Leonhart R, Gerger H, Barth J. Researcher allegiance in psychotherapy outcome research: an overview of reviews. Clin Psychol Rev. 2013 Jun;33(4):501–11.
39 Concato J, Horwitz RI. Limited usefulness of meta-analysis for informing patient care. Psychother Psychosom. 2019 Sep;88(5):257–62.
40 Cohen ZD, DeRubeis RJ. Treatment selection in depression. Annu Rev Clin Psychol. 2018 May;14:209–36.
41 Fava GA, Cosci F, Guidi J, Rafanelli C. The deceptive manifestations of treatment resistance in depression: a new look at the problem. Psychother Psychosom. 2020 Apr;89(5):265–73.
42 Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux P, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010 Aug;63(8):e1–37.
43 Grant S, Mayo-Wilson E, Montgomery P, Macdonald G, Michie S, Hopewell S, et al. CONSORT-SPI 2018 explanation and elaboration: guidance for reporting social and psychological intervention trials. Trials. 2018 Jul;19(1):406.
44 Chaimani A, Caldwell D, Li T, Higgins JPT, Salanti G. Chapter 11: undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane handbook for systematic reviews of interventions. Version 6.2 (updated 2021 Feb): Cochrane; 2021.