



Published in final edited form as:

Clin Psychol (New York). 2016 June ; 23(2): 165–176. doi:10.1111/cpsp.12153.

A Meta-Analysis of Personalized Treatment Goals in Psychotherapy: A Preliminary Report and Call for More Studies

Oliver Lindhiem¹, Charles B. Bennett², Trina E. Orimoto³, and David J. Kolko¹

¹University of Pittsburgh

²University of Pittsburgh Medical Center

³University of Hawai'i at Mānoa

Abstract

The aim of this study was to explore the hypothesis that psychotherapy has larger effect sizes for personalized treatment goals than for symptom checklists. We conducted a meta-analysis of clinical trials that measured treatment success *both* in terms of symptom checklists and personalized treatment goals. Our search of the literature yielded 12 studies that met our inclusion criteria. Effect sizes were substantially larger for personalized treatment goals ($ES = .86, p < .0001$) than for symptom checklists ($ES = .32, p = .003$). The magnitude of this difference was significant ($p < .05$). Our results suggest that psychotherapy is perhaps more effective in helping patients with individual goals than reducing scores on broad measures of symptoms. Estimates of the effectiveness of psychotherapy that are based on symptom checklists perhaps underestimate the true benefit of psychotherapy. We discuss the implications for research and clinical practice.

Keywords

personalized treatment goals; clinical trials; psychotherapy; treatment outcome; meta-analysis

Quantifying the effectiveness of psychotherapy is methodologically challenging. More recent meta-analyses generally indicate effect sizes in the medium range ($d = .4$ to $.6$) when psychotherapy is compared to active placebo conditions (e.g., Lambert & Ogles, 2004). When compared to “well-designed” placebos (conditions that are “structurally equivalent” to the psychotherapy), effect sizes may even be quite small (below $d = .3$; e.g., Baskin, Tierney, Minami, & Wampold, 2003). These and other attempts at quantifying the effectiveness of psychotherapy, however, have relied largely on omnibus measures of treatment outcome such as symptom checklists. Symptom checklists inevitably include items that are not relevant to all patients, perhaps resulting in diluted effect sizes (Kiresuk & Sherman, 1968). Estimates of the effectiveness of psychotherapy that are based on symptom checklists, therefore, may underestimate the true benefit of psychotherapy. Recognizing the limitations of symptom checklists for measuring the effectiveness of psychotherapy, researchers are

increasingly turning to standardized measures of personalized treatment goals (e.g., Kolko, Campo, Kelleher, Cheng, 2010; Kolko, Lindhiem, Hart, & Bukstein, 2014; Weisz et al., 2011). This approach has a long history, dating back almost 50 years to the development of Goal Attainment Scaling (GAS; Kiresuk & Sherman, 1968).

Goal Attainment Scaling

Goal Attainment Scaling (GAS; Kiresuk & Sherman, 1968) remains perhaps the most widely used approach to the measurement of personalized treatment goals. The GAS was developed due to the need for standardized measures that could be used across a variety of mental health programs with different outcomes of interest. The GAS approach involves three primary steps: 1) setting personalized treatment goals, 2) establishing a measurable scale for each goal, and 3) transforming post-treatment goal attainment into a standardized T-score. The GAS is typically used to supplement global measures of treatment outcome. From our review of the literature, it continues to be the most widely used measure of personalized treatment goals for psychological interventions. In one study, the GAS was used to measure clinical improvement in a sample of men and women receiving an internet treatment for social phobia (Berger, Hohl, & Caspar, 2009). Results indicated that the GAS effect sizes were larger than the effect sizes of the global measures. The GAS is particularly useful in assessing personalized improvement in studies featuring samples with a variety of diagnoses. In one study, the GAS was used with a sample of patients with diagnoses of anxiety disorders, depressive disorders, and adjustment disorders (Shefler, Dasberg, & Ben-Shakhar, 1995). Findings showed that GAS effect sizes were larger than global measure effect sizes.

Personalized Treatment Goals for Psychotherapy Research

Recognizing the limitations of symptom checklists for psychological research, measures of personalized treatment goals for psychotherapy such as the Individualized Goal Achievement Rating (IGAR) scale (e.g., Kolko et al., 2010; Kolko et al., 2014) and the Top Problems Assessment (TPA; e.g., Weisz et al., 2011) have been developed. These measures can trace their roots back to the GAS. The IGAR was developed to measure goal attainment related to child problem behaviors. Like the GAS, it is used to measure improvement according to the specific needs of the patient. Each point on the IGAR represents a level of severity (e.g., 1 = not a problem at all; 7 = very serious problem; Kolko et al., 2009). Unlike the GAS, scores on the IGAR are not converted to T scores. In one study using the IGAR, effect sizes for IGAR scores were substantially larger than the effect sizes for all other treatment outcomes (Kolko et al., 2009). The TPA is also primarily used in the treatment of childhood behavior problems. Using the TPA, the top three problems of both the child and parent are identified and evaluated on a scale from 0 (“not at all a problem”) to 10 (“very, very much a problem”). Like the IGAR, the TPA was designed to be administered to parents and youth multiple times over the course of treatment to measure the trajectory of improvement over time. The measure has demonstrated test–retest reliability, convergent and discriminant validity, sensitivity to change, slope reliability, and an association with slopes on other standardized measures including the Child Behavior Checklist (CBC;; Achenbach & Edelbrock, 1983) and Youth Self-Report (YSR; Achenbach & Rescorla, 2001).

Current Study

In the current study, we explored the hypothesis that psychotherapy has larger effect sizes for personalized treatment goals than for symptom checklists. Specifically, we conducted a meta-analysis of psychotherapy trials that included both symptom checklists and measures of individualized treatment goals. In order to examine the robustness of the results, we examined four potential moderators, namely, treatment duration (brief versus long), participant type (children versus adults), type of control condition (no treatment, waitlist, treatment as usual), and whether or not participants were randomized to the experimental and control conditions.

Method

Procedure

Eligibility criteria—All articles in this meta-analysis reported on personalized treatment goals in the context of a clinical trial. Inclusion criteria included, 1) a psychotherapy outcome study, 2) a control condition, 3) a measure of both personalized treatment goals and symptom checklists, 4) written in English, and 5) peer review or dissertation.

Study selection—Studies were identified through a search of all articles indexed in the database PsycINFO as of June 24, 2014 using the ProQuest search engine. An initial search for articles that cited Kiresuk & Sherman (1968) was conducted to identify keywords and authors utilizing measures of personalized treatment goals. After the initial search, a search of *goal attainment* and a search that crossed *idiographic, personalized, or individualized* with *treatment goals* were conducted. The names of authors identified through these methods were also searched for additional articles featuring an individualized measure of treatment outcome. These searches yielded 5,718 articles. Of these articles, 5,668 articles were excluded for one or more of the following reasons: 1) did not include a control group, 2) treatment was not psychotherapy, 3) did not include an individualized measure of treatment outcome, 4) not written in English, or 5) books or chapters (not peer reviewed). From these searches, 50 articles were obtained for search in greater detail. The reference sections of these 50 articles were also searched for eligible studies, however, no additional articles were found. Of these 50 articles, twelve were excluded due to the lack of a comparable control group (includes studies comparing two active treatments, healthy controls, and within-group comparisons), twelve did not feature a psychotherapeutic intervention, four did not include an individualized measure of clinical outcome, three featured medication as part of the treatment, two did not evaluate treatment outcome, and one did not include both individualized goals and symptom checklists. The current meta-analysis includes 12 studies (see Figure 1).

Coding—The following variables were coded:

1. Sample descriptors: age of sample (adult vs. children).
2. Research design descriptors: Total sample size, treatment group sample size, control group sample size, treatment duration (number of sessions), unit of

randomization (randomization vs. no randomization), and the number of anchors on each individualized measure.

Standardized mean difference effect sizes (ES_d) were calculated using the equations and guidelines as explained by Lipsey and Wilson (2001). Means and standard deviations were the primary method used for calculating effect sizes for individualized and global measures of treatment outcome. We used the following formulas:

$$ES = \frac{\bar{X}_1 - \bar{X}_2}{s_p}; s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}},$$

where \bar{X} represents the mean reported for each group, s_p is the pooled standard deviation, and n is the number of participants in each condition. To calculate one effect size in which means and standard deviations were not available, the proportion of participants in each condition who successfully obtained their treatment goals was used to calculate the effect size. Using the proportions, an odds-ratio was first calculated. The odds-ratio was then used to calculate the standardized mean

difference effect size using the formula, $ES = \ln(OR) \times \frac{\sqrt{3}}{\pi}$. Three studies featured more than two treatment conditions. For these studies, the stated aim of the study was used to determine the appropriated treatment conditions with which to estimate an effect size. Specifically, the means and standard deviations of the cognitive behavioral therapy (CBT) groups (Kolko 1996; Kolko 2012) and the combined cognitive counseling and skills training group (Gormally, Varvil-Weld, Raphael, & Sipps, 1981) were used in the effect size calculations. Data collected at the termination of treatment were used to calculate effect sizes. All articles were double-coded by the second and third authors. The inter-rater agreement was 82%. Discrepancies were resolved through conferencing among the first, second, and third authors.

Personalized treatment goals—Table 1 summarizes the individualized measures of treatment outcome. The meta-analysis included four studies using the GAS, six studies using the IGAR (including two early versions of the IGAR; the Individualized Treatment Problems Rating, Kolko, 1996; Individualized Child Fire Problems, Kolko, 2001), and two studies that used single items to measure personalized treatment goals. Of these measures, ten were completed by the patients or parents (if the patients were children) and two were completed by blinded research staff. All studies assessed between one and five personalized goals. In studies with multiple goals, the attainment of all goals was averaged and included in the analysis.

As shown in Table 1, the reliabilities and validities of each measure were reported in the articles included in this meta-analysis. The IGAR (including early versions) has shown temporal stability, face validity, and construct validity. The GAS has shown test-retest, temporal stability, and inter-rater reliabilities, face validity, and content validity. The inter-rater reliability (Marson, Wei, & Wasserman, 2009) and content validity (Garwick & Lampman, 1972) of the GAS were obtained from published reviews of the measure. Both single item assessments of goal attainment showed face validity, and one provided evidence of temporal stability (Swildens et al., 2011).

Symptom checklists—As shown in Table 2, psychometrically established global measures were used to calculate the aggregate effect size for symptom checklists. Two studies included only one measure. These studies featured the Social Functioning Scale (SFS; Birchwood, Smith, Cochrane, Wetton, & Copestake, 1990) and the Survey of Heterosexual Interactions (SHI; Twentyman & McFall, 1975). In one study (Berger, Hohl, & Caspar, 2009), the primary method for assessing treatment outcome comprised an aggregate of three measures: the Liebowitz Social Anxiety Scale (LSAS-SR; Baker, Heinrichs, Kim, & Hofmann, 2002), the Social Phobia Scale (SPS; Mattick & Clarke, 1998) and the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998). The effect size for each measure was used to calculate an average effect size which was included in the meta-analysis. For studies with multiple measures that did not identify a primary measure for treatment outcome, measures that were common across studies were used to calculate effect sizes. Two studies used the Child Behavior Checklist (CBCL; Gardner et al., 1999), three studies used the Global Severity Index of the Brief Symptom Inventory (GSI; Derogatis & Melisaratos, 1983), and two studies used the Vanderbilt Attention-Deficit/Hyperactivity Disorder Diagnostic Parent Rating Scale (VADPRS; Wolraich, Lambert, Doffing, Bickman, Simmons, & Worley, 2003). For studies in which these methods for identifying a global measure were unavailable, measures that closely aligned with the aims of the study were selected (i.e., the Fire History Screen, FHS; Kolko & Kazdin, 1989; the Pediatric Symptom Checklist-17, PSC-17; Gardner, et al., 1999). When studies analyzed data according to the subscales of measures, the means and standard deviations of the subscale that most closely represented improvement for the sample were chosen to calculate an effect size (e.g., the externalizing subscale of the CBCL for children receiving treatment for behavioral problems; Kolko, Lindhiem, Hart, Bukstein, 2014).

Data Analyses

Analyses were conducted using Lipsey and Wilson's (2001) macros for SPSS. Separate analyses were conducted to estimate effect sizes for personalized treatment goals ($k = 12$ effect sizes) and symptom checklists ($k = 12$ effect sizes). Aggregate effect sizes were estimated using a random effects models and the Q homogeneity statistic was utilized to determine whether heterogeneity in effect sizes supported examination of moderators of effect size. The Q statistic has a chi-square distribution based on $k - 1$ degrees of freedom where k represents the number of effect sizes. A homogeneity Q statistic that is statistically significant suggests that the distribution of effect sizes is heterogeneous. When the distribution of effect sizes was heterogeneous, we individually evaluated potential moderator variables with the $Q_{between}$ statistic using maximum likelihood estimation. A statistically significant $Q_{between}$ supports the variable as a moderator of effect size heterogeneity.

Results

Global Treatment Outcomes

The meta-analytic results for symptom checklists are summarized in Table 3 and Figure 2. Across the 12 independent effect sizes, the aggregate effect size was $ES = .32$, $p = .003$. This falls within the small to medium range based on Cohen's definitions of "small" ($d = .20$) and "medium" ($d = .50$) effect sizes. The homogeneity Q statistic ($Q = 24.67$, $p = .01$) indicated

statistically significant variability in the effect sizes between the 12 studies. Of the study-level variables that were examined, only type of control condition moderated the effect sizes (between $Q = 13.94$, $p = .0002$). Effect sizes were smaller for studies with “treatment as usual” control conditions ($ES = .16$, $p = .01$) than for studies with “waitlist” control conditions ($ES = .97$, $p < .0001$).

Personalized Treatment Goals

The meta-analytic results for personalized treatment outcomes are summarized in Table 4 and Figure 2. Across the 12 independent effect sizes, the aggregate effect size was $ES = .86$, $p < .0001$. This is a large effect size based on Cohen's definition of a “large” ($d = .80$) effect size. The homogeneity Q statistic ($Q = 66.38$, $p < .0001$) indicated statistically significant variability in the effect sizes between the 12 studies. As with the effect sizes for global treatment outcomes, only type of control condition was a significant study-level moderator (between $Q = 5.39$, $p = .02$). Again, effect sizes were smaller for studies with “treatment as usual” control conditions ($ES = .65$, $p = .002$) than for studies with “waitlist” control conditions ($ES = 1.72$, $p < .0001$).

Personalized Goals Versus Symptom Checklists

The magnitude of the difference between the aggregate effect size for personalized treatment goals and the aggregate effect size for symptom checklists was statistically significant, paired $t(11) = 2.43$, $p < .05$. Figure 3 contrasts the two effect sizes. Furthermore, the effect sizes for personalized treatment goals were larger than the effect sizes for symptom checklists for all of the subgroups based on the moderators we examined, without exception. The fail-safe N using Orwin's (1983) formula indicated that 21 additional studies with null findings would be necessary to reduce the aggregate effect size for personalized treatment goals (.86) down to the level of the aggregate effect size for global outcome measures (.32). In other words, the ratio of unpublished to published studies would need to be almost 2:1.

Discussion

Consistent with our hypothesis, effect sizes were substantially larger for personalized treatment goals than for symptom checklists. The overall difference was statistically significant and the pattern was consistent across all moderators that were examined. Effect sizes for symptom checklists were comparable to those reported in extant meta-analyses of psychotherapy effectiveness (e.g., Baskin et al., 2003; Lipsey & Wilson, 1993). Effect sizes for personalized treatment goals, however, were quite large in comparison. These results support the hypothesis that estimates of psychotherapy effectiveness that are based on symptom checklists perhaps underestimate the true benefit of psychotherapy.

We were careful to select studies with rigorous enough research designs to rule out plausible alternative explanations for differences in effect sizes. Most importantly, all 12 studies had control conditions allowing us to estimate standardized mean differences at post-treatment for the experimental and control conditions. This effectively rules out potential confounds including self-report bias and social desirability effects. In addition, we selected studies that included *both* symptom checklists and individualized treatment goals. For the direct

comparison of effect sizes, we were therefore able to use a paired t-test whereby each treatment acted as its own control condition. This rules out the possibility that the larger aggregate effect size for personalized treatment goals is due to more effective treatments. One difference we were not able to control for was the number of items. Effect sizes for personalized treatment goals were based on far fewer items (1 to 5) than symptom checklists (average around 30 items). However, fewer items should only result in more random variability or greater “noise,” not systematically larger effect sizes. Therefore, the fewer number of items is not a plausible explanation for the substantially larger, and statistically significant, aggregate effect size for personalized treatment goals. A similar critique might be that the measures of personalized treatment goals are less reliable and/or valid than symptom checklists. Again, this would only have the effect of generating more random variability and cannot plausibly explain systematically larger, and statistically significant, aggregate effect sizes for personalized treatment goals. If anything, less reliable and/or valid measures would be expected to result in diluted and therefore smaller effect sizes. A more plausible explanation for the results is that personalized treatment goals are more specific measures of treatment effectiveness than symptom checklists.

Implications for Practice, Research, & Policy

The results highlight the importance of setting personalized treatment goals as part of routine clinical practice. Often the most pressing treatment goals for clients and families are those that have a direct impact on their own improved functioning and quality of life. Weisz and colleagues (2011) have argued that the utilization of team-identified and client-identified problems assists in both prioritizing and identifying targets that might not arise in standardized measures. One measure of personalized treatment goals with substantial clinical utility is the Individualized Goal Assessment Rating (IGAR) scale (e.g., Kolko, Campo, Kelleher, Cheng, 2010; Kolko, Lindhiem, Hart, & Bukstein, 2014). Recent versions of the IGAR assess up to four individualized child or parent target behaviors identified by caregivers at intake. For each target, the caregiver identifies behavioral anchors for each point on the scale to facilitate more precise assessment of the level of improvement per behavior during treatment and at follow-up. With its ease of collection and inclusion of a graphic interface showing both progress and care processes delivered, this feedback tool may encourage more careful documentation of targeted problems in clinical trials targeting different types of problems and conducted in different treatment or care delivery settings (Kolko et al., 2010; Kolko et al., 2012; Kolko, Campo, et al., 2014). Another straightforward and practical method for setting personalized therapy goals and tracking progress is the “Planning and Assessment in Clinical Care” (PACC) approach (Woody, Deteiler-Bedell, Teachman, & O’Hearn, 2004). The PACC approach involves creating a patient “problem list,” prioritizing the problem list, and measuring progress using a numerical scale. Like the IGAR, the PACC approach emphasizes tracking patient progress session by session and graphing the patient’s progression through treatment.

The results of the current study also have several important implications for research and policy. First, the use of personalized treatment goals might result in more *specific* tests of treatment effectiveness. The use of standardized methods that allow for the assessment of unique or individualized content could be used to supplement more traditional measures of

treatment effectiveness, such as broad symptom reduction (see Bond, Drake, Rapp, McHugo, & Xie, 2009). It will be important that the methods selected, like those included in the current meta-analysis, are standardized and calibrated with an interpretable scoring system (rating scales). Also, because symptom checklists inevitably include items that are not relevant to all patients, the use of personalized goals may also shorten the time needed for assessment and monitoring. As seen by comparing tables 1 and 2, measures of personalized treatment goals are much more brief (1 to 5 items) than symptom checklists (average around 30 items). Setting personalized goals also allows clinicians to focus treatment more narrowly on those problems of concern to clients. This may have the added benefit of enhancing the treatment's specificity and effectiveness. Finally, the monitoring of patient progress itself has the potential to enhance treatment outcome (e.g., Bickman, 2008; Bickman, Kelley, Breda, de Andrade, & Riemer, 2011; Kelley, de Andrade, Sheffer, & Bickman, 2010).

Limitations and Future Directions

A clear limitation of the study was the relatively small number of studies that met our inclusion criteria. In addition, the studies varied considerably on important parameters including age of sample, clinical diagnoses, and assessment strategy. Our hope is that future clinical trials will increasingly include measures of individualized treatment goals. This will allow us to conduct an updated meta-analysis in the future, providing for a much more robust test of our hypothesis. Another limitation is that half of the studies included in the current meta-analyses were conducted by the same research team and lead author. Again, this limitation will be mitigated in the future as measures of individualized treatment goals are routinely included in clinical trials. Finally, although several options exist, we still need to explore the most reliable, valid, and efficient way for identifying, rating, and monitoring client progress towards personalized treatment goals.

Conclusions

Our results suggest that psychotherapy is perhaps more effective in helping patients with their individual goals than reducing scores on omnibus measures of symptoms. Estimates of the effectiveness of psychotherapy that are based on measures of global outcome perhaps underestimate the true benefit of psychotherapy. These conclusions are very tentative given the small number of studies included in this meta-analysis. Routine measurement of personalized treatment goals in the context of clinical trials for psychotherapy will allow for a more robust test of this hypothesis.

Acknowledgments

This study was supported by a grant from the National Institute of Mental Health (NIMH) to the first author (MH 093508).

References

* denotes articles included in the meta-analysis

- Achenbach, TM.; Edelbrock, C. Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University Associates in Psychiatry; 1983.
- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2001.
- Baker SL, Heinrichs N, Kim H, Hofmann SG. The Liebowitz Social Anxiety Scale as a self-report instrument: A preliminary psychometric analysis. *Behaviour Research and Therapy*. 2002; 40(6): 701–715. DOI: 10.1016/S0005-7967(01)00060-2 [PubMed: 12051488]
- Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: A meta-analysis of structural equivalence in placebo controls. *Journal of Consulting and Clinical Psychology*. 2003; 71:973–979. [PubMed: 14622072]
- *. Berger T, Hohl E, Caspar F. Internet-based treatment for social phobia: A randomized controlled trial. *Journal of Clinical Psychology*. 2009; 65(10):1021–1035. DOI: 10.1002/jclp.20603 [PubMed: 19437505]
- Bickman L. A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2008; 47(10):1114–1119. DOI: 10.1097/CHI.0b013e3181825af8 [PubMed: 20566188]
- Bickman L, Kelley SD, Breda C, de Andrade AR, Riemer M. Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services*. 2011; 62(12):1423–1429. DOI: 10.1176/appi.ps.002052011 [PubMed: 22193788]
- Birchwood M, Smith J, Cochrane R, Wetton S, Copestake S. The Social Functioning Scale: The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. *The British Journal of Psychiatry*. 1990; 157:853–859. [PubMed: 2289094]
- Bond GR, Drake RE, Rapp CA, McHugo GJ, Xie H. Individualization and quality improvement: Two new scales to complement measurement of program fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*. 2009; 36(5):349–357. DOI: 10.1007/s10488-009-0226-y [PubMed: 19499322]
- Derogatis LR, Melisaratos N. The Brief Symptom Inventory: An introductory report. *Psychological Medicine*. 1983; 13(3):595–605. [PubMed: 6622612]
- *. Flückiger C, Holtforth MG. Focusing the therapist's attention on the patient's strengths: A preliminary study to foster a mechanism of change in outpatient psychotherapy. *Journal of Clinical Psychology*. 2008; 64(7):876–890. DOI: 10.1002/jclp.20493 [PubMed: 18484601]
- Gardner W, Murphy JM, Childs G, Kelleher K, Pagano ME, Jellinek MS, Chiapetta L. The PSC-17: A brief pediatric symptom checklist with psychosocial problem subscales. A report from PROS and ASPN. *Ambulatory Child Health*. 1999; 5:225–236.
- Garwick G, Lampman S. Typical problems bringing patients to a community mental health center. *Community Mental Health Journal*. 1972; 8:271–280. [PubMed: 4669435]
- *. Gormally J, Varvil-Weld D, Raphael R, Sipps G. Treatment of socially anxious college men using cognitive counseling and skills training. *Journal of Counseling Psychology*. 1981; 28(2):147–157. DOI: 10.1037/h0077967
- Kelley SD, de Andrade ARV, Sheffer E, Bickman L. Exploring the black box: Measuring youth treatment process and progress in usual care. *Administration and Policy in Mental Health and Mental Health Services Research*. 2010; 37(3):287–300. DOI: 10.1007/s10488-010-0298-8 [PubMed: 20238155]
- Kiresuk TJ, Sherman RE. Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal*. 1968; 4(6):443–453. [PubMed: 24185570]
- *. Kolko DJ. Individual cognitive behavioral treatment and family therapy for physically abused children and their offending parents: A comparison of clinical outcomes. *Child Maltreatment*. 1996; 1(4):322–342. DOI: 10.1177/1077559596001004004
- *. Kolko DJ. Efficacy of cognitive-behavioral treatment and fire safety education for children who set fires: Initial and follow-up outcomes. *Journal of Child Psychology and Psychiatry*. 2001; 42(3): 359–369. DOI: 10.1111/1469-7610.00729 [PubMed: 11321205]

- *. Kolko DJ, Campo JV, Kelleher K, Cheng Y. Improving access to care and clinical outcome for pediatric behavioral problems: A randomized trial of a nurse-administered intervention in primary care. *Journal of Developmental and Behavioral Pediatrics*. 2010; 31(5):393–404. DOI: 10.1097/DBP.0b013e3181dff307 [PubMed: 20495474]
- *. Kolko DJ, Campo J, Kilbourne AM, Hart J, Sakolsky D, Wisniewski S. Collaborative care outcomes for pediatric behavioral health problems: A cluster randomized trial. *Pediatrics*. 2014; 133(4):e981–e992. [PubMed: 24664093]
- *. Kolko DJ, Campo JV, Kilbourne AM, Kelleher K. Doctor-office collaborative care for pediatric behavioral problems: a preliminary clinical trial. *Archives of Pediatric Adolescent Medicine*. 2012; 166(3):224–231. DOI: 10.1001/archpediatrics.2011.201
- Kolko DJ, Dorn LD, Bukstein OG, Pardini D, Holden EA, Hart J. Community vs. clinic-based modular treatment of children with early-onset ODD or CD: A clinical trial with 3-year follow-up. *Journal of Abnormal Child Psychology*. 2009; 37(5):591–609. DOI: 10.1007/s10802-009-9303-7 [PubMed: 19221871]
- Kolko DJ, Kazdin AE. The Children's Firesetting Interview with psychiatrically referred and nonreferred children. *Journal of Abnormal Child Psychology*. 1989; 17(6):609–624. [PubMed: 2607053]
- *. Kolko DJ, Lindhiem O, Hart J, Bukstein OG. Evaluation of a booster intervention three years after acute treatment for early-onset disruptive behavior disorders. *Journal of Abnormal Child Psychology*. 2014; 42(3):383–398. DOI: 10.1007/s10802-013-9724-1 [PubMed: 23494526]
- Lambert, MJ.; Ogles, BM. The efficacy and effectiveness of psychotherapy. In: Lambert, MJ., editor. *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. New York, NY: John Wiley & Sons; 2004. p. 139-193.
- Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*. 1993; 48(12):1181–1209. DOI: 10.1037/0003-066X.48.12.1181 [PubMed: 8297057]
- Lipsey, MW.; Wilson, DB. *Practical meta-analysis*. Thousand Oaks, CA: Sage; 2001.
- Marson SM, Wei G, Wasserman D. A reliability analysis of Goal Attainment Scaling (GAS) weights. *American Journal of Evaluation*. 2009; 30(2):203–216. DOI: 10.1177/1098214009334676
- Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*. 1998; 36(4):455–470. [PubMed: 9670605]
- Orwin RG. A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*. 1983; 8:157–159.
- *. Shefler G, Dasberg H, Ben-Shakhar G. A randomized controlled outcome and follow-up study of mann's time-limited psychotherapy. *Journal of Consulting and Clinical Psychology*. 1995; 63(4): 585–593. DOI: 10.1037/0022-006X.63.4.585 [PubMed: 7673536]
- *. Swildens W, van Busschbach JT, Michon H, Kroon H, Koeter MWJ, Wiersma D, van Os J. Effectively working on rehabilitation goals: 24-month outcome of a randomized controlled trial of the Boston psychiatric rehabilitation approach. *The Canadian Journal of Psychiatry / La Revue Canadienne De Psychiatrie*. 2011; 56(12):751–760.
- Twentyman CT, McFall RM. Behavioral training of social skills in shy males. *Journal of Consulting and Clinical Psychology*. 1975; 43(3):384–395. DOI: 10.1037/h0076743 [PubMed: 1159128]
- *. Weiss M, Nordlie JW, Siegel EP. Mindfulness-based stress reduction as an adjunct to outpatient psychotherapy. *Psychotherapy and Psychosomatics*. 2005; 74(2):108–112. DOI: 10.1159/000083169 [PubMed: 15741760]
- Weisz JR, Chorpita BF, Frye A, Ng MY, Lau N, Bearman SK, Hoagwood KE. Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology*. 2011; 79(3):369–380. DOI: 10.1037/a0023307 [PubMed: 21500888]
- Wolraich ML, Lambert W, Doffing MA, Bickman L, Simmons T, Worley K. Psychometric properties of the Vanderbilt ADHD Diagnostic Parent Rating Scale in a referred population. *Journal of Pediatric Psychology*. 2003; 28(8):559–568. DOI: 10.1093/jpepsy/jsg046 [PubMed: 14602846]

Woody, SR.; Detweiler-Bedell, J.; Teachman, BA.; O'Hearn, T. Treatment and planning in psychotherapy: Taking the guesswork out of clinical care. New York: The Guilford Press; 2004.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

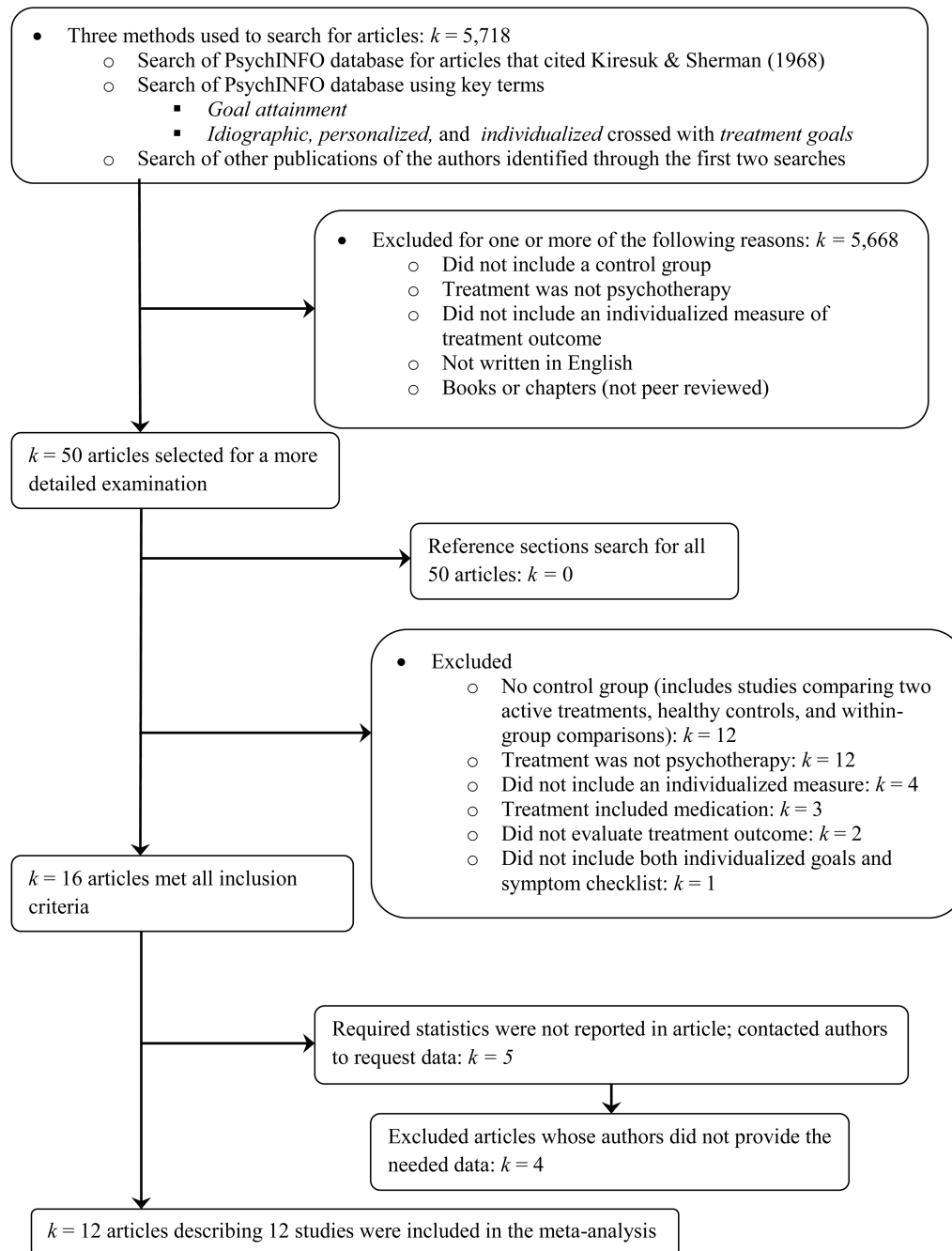


Figure 1. Study selection flow chart

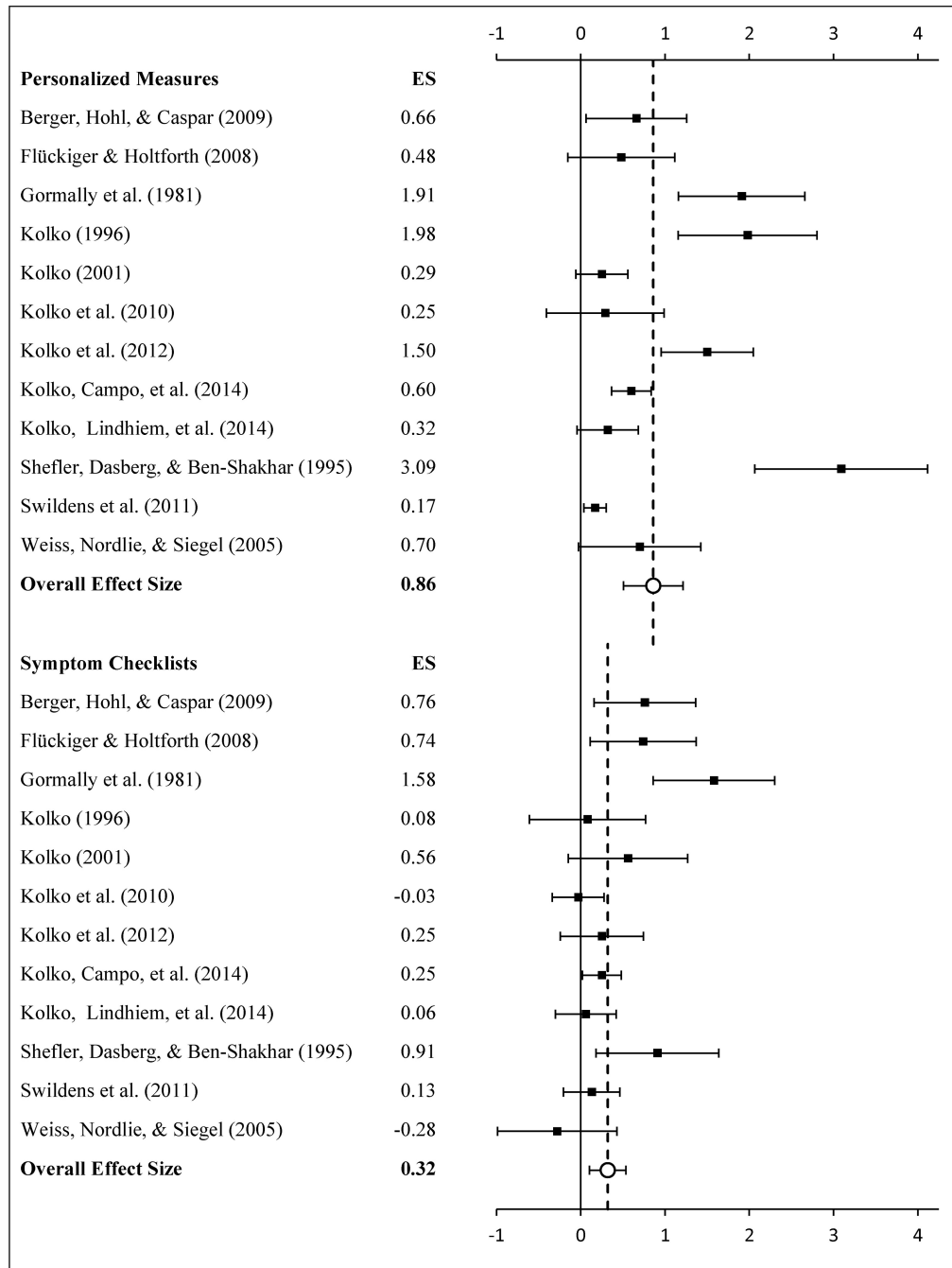


Figure 2. Effect sizes for symptom checklists and personalized measures of treatment outcome. Error bars represent 95% CIs

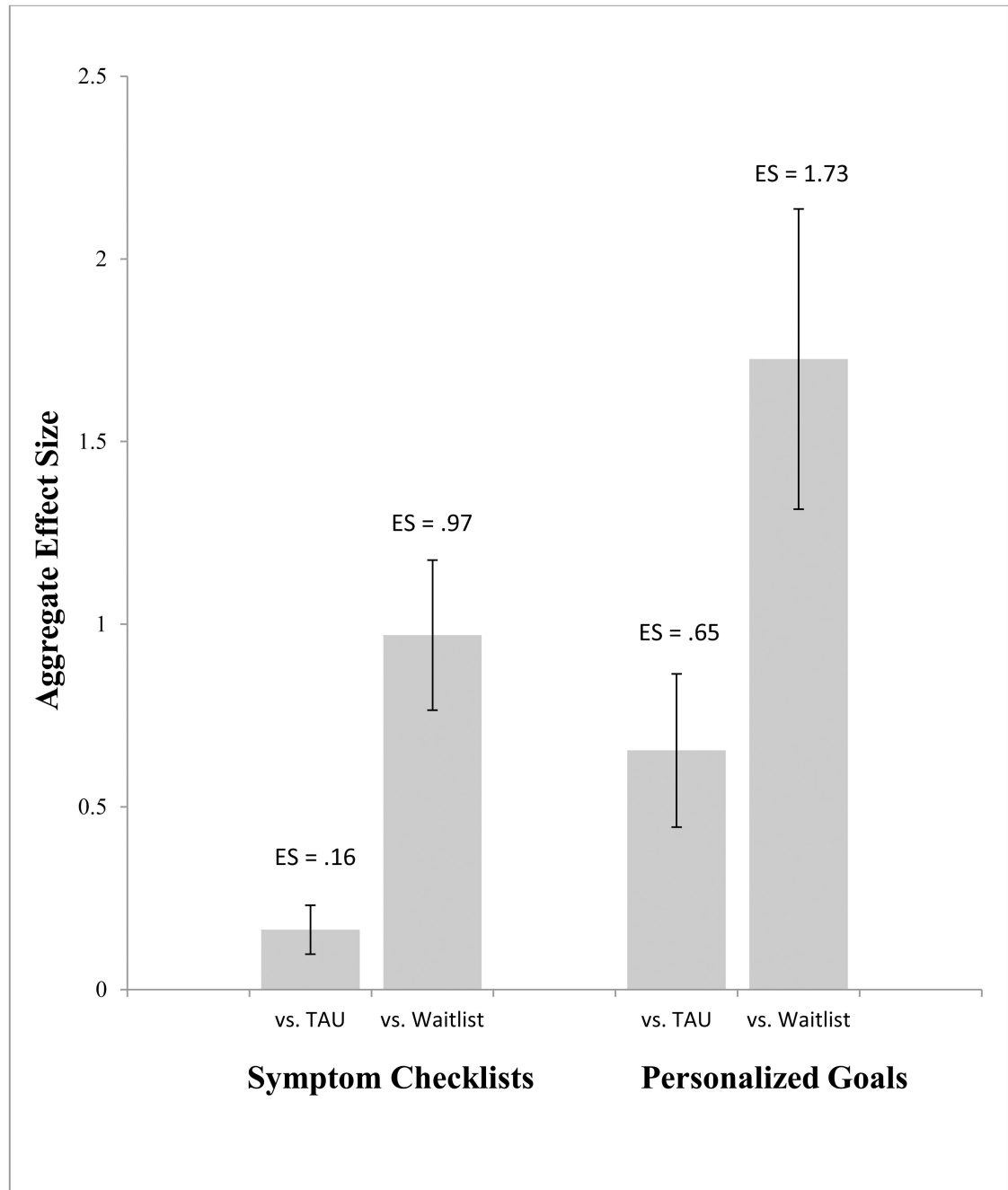


Figure 3. Aggregate effect sizes for symptom checklists and personalized treatment goals with standard error bars

Table 1

Measures of Individualized Treatment Goals

Author	Measure Name	Measure completed by	Number of Goals	Scale	Reliability	Validity
Berger, Hohl, & Caspar (2009)	GAS	Patient	3	-2 to 4	Test-retest reliability, temporal stability, inter-rater reliability	Face validity, content validity
Kolko (1996)	Individualized Treatment Problems Rating (IGAR)	Parent of Patient	3	1 to 10	Temporal stability	Face validity, construct validity (moderately correlates with symptom reduction)
Kolko (2001)	Individualized Child Fire Problems (IGAR)	Patient	1	1 to 3	Temporal stability	Face validity, construct validity (moderately correlates with symptom reduction)
Kolko et al. (2010)	IGAR	Parent	1 to 4	1 to 5	Temporal stability	Face validity, construct validity (moderately correlates with symptom reduction)
Kolko et al. (2012)	IGAR	Parent	1 to 4	1 to 5	Temporal stability	Face validity, construct validity (moderately correlates with symptom reduction)
Kolko, Campo, et al. (2014)	IGAR	Parent	1 to 4	1 to 5	Temporal stability	Face validity, construct validity (moderately correlates with symptom reduction)
Kolko, Lindhiem, et al. (2014)	IGAR	Parent	1 to 5	1 to 7	Temporal stability	Face validity, construct validity (moderately correlates with symptom reduction)
Flückiger & Holtforth (2008)	GAS	Patient	Unspecified	-2 to 4	Test-retest reliability, temporal stability, inter-rater reliability	Face validity, content validity
Gormally et al. (1981)	GAS	Patient	3	1 to 5	Test-retest reliability, temporal stability, inter-rater reliability	Face validity, content validity
Shefner, Dasberg, & Ben-Shakhar (1995)	GAS	Blinded researchers	5	-2 to 2	Test-retest reliability, temporal stability, inter-rater reliability	Face validity, content validity
Swildens et al. (2011)	Single item	Blinded researchers	1	1 to 3	Temporal stability	Face validity
Weiss, Nordlie, & Siegel (2005)	Single item	Patient	1	1 to 100	-	Face validity

Note. GAS = Goal Attainment Scale; IGAR = Individualized Goal Achievement Rating

Table 2
Symptom Checklists

Author	Measure Name	Subscale	Number of items
Berger, Hohl, & Caspar (2009)	Liebowitz Social Anxiety Scale	Total Score	12
	Social Phobia Scale	Total Score	20
	Social Interaction Anxiety Scale	Total Score	19
Kolko (1996)	Child Behavior Checklist	Externalizing	33
Kolko (2001)	Fire History Screen	Total Score	46
Kolko et al. (2010)	Pediatric Symptom Checklist-17	Externalizing	7
Kolko et al. (2012)	Vanderbilt Attention-Deficit/Hyperactivity Disorder Diagnostic Parent Rating Scale	Disruptive Behavior	22
Kolko, Campo, et al. (2014)	Vanderbilt Attention-Deficit/Hyperactivity Disorder Diagnostic Parent Rating Scale	Oppositional defiant/conduct disorder	22
Kolko, Lindhiem, et al. (2014)	Child Behavior Checklist	Externalizing	33
Flückiger & Holtforth (2008)	Brief Symptom Inventory	Total Score (Global Severity Index)	53
Gormally et al. (1981)	Survey of Heterosexual Interactions	Total Score	20
Shefler, Dasberg, & Ben-Shakhar (1995)	Brief Symptom Inventory	Total Score (Global Severity Index)	53
Swildens et al. (2011)	Social Functioning Scale	Total Score (except the Employment subscale)	79
Weiss, Nordlie, & Siegel (2005)	Brief Symptom Inventory	Total Score (Global Severity Index)	53

Table 3

Meta-Analysis of Symptom Checklists

	<i>k</i>	<i>N</i>	<i>d</i>	95% CI	<i>Q</i>
Total	12	1,043	0.32 ^{**}	0.11 to 0.53	24.67 [*]
Age of Sample					3.83
Adults	6	327	0.49 ^{***}	0.22 to 0.76	
Children	6	716	0.16	-0.04 to 0.36	
Type of Control					13.94 ^{***}
Treatment as usual	9	925	0.16 [*]	0.03 to 0.30	
Waitlist	3	118	0.97 ^{***}	0.57 to 1.37	
Treatment Duration					0.15
< 12 sessions	5	312	0.36 [*]	0.03 to 0.70	
12 sessions	7	731	0.28 [*]	0.04 to 0.53	
Randomization					0.27
Randomized	10	972	0.32	0.11 to 0.53	
Not randomized	2	71	0.26	-0.31 to 0.84	

* $p < .05$;** $p < .01$;*** $p < .001$

Table 4

Meta-Analysis of Personalized Measures of Treatment Outcome

	<i>k</i>	<i>N</i>	<i>d</i>	95% CI	<i>Q</i>
Total	12	1,043	0.86 ^{***}	0.52 to 1.21	66.38 ^{***}
Age of Sample					0.35
Adults	6	327	1.05 ^{**}	0.41 to 1.68	
Children	6	716	0.78 [*]	0.18 to 1.38	
Type of Control					5.39 [*]
Treatment as usual	9	925	0.65 ^{**}	0.24 to 1.07	
Waitlist	3	118	1.73 ^{***}	0.92 to 2.53	
Treatment Duration					0.55
< 12 sessions	5	312	0.71 [*]	0.03 to 1.39	
12 sessions	7	731	1.04 ^{***}	0.48 to 1.60	
Randomization					0.40
Randomized	10	972	0.97	0.49 to 1.45	
Not randomized	2	71	0.59	-0.49 to 1.66	

* $p < .05$;** $p < .01$;*** $p < .001$