NCT04559698

Title of Research Project: Assessing mental health providers' clinical knowledge and skills via an online training on LGBTQ-affirmative cognitive-behavioral therapy

Date: 12.7.21

Assessing Mental Health Providers Clinical Knowledge and Skills Via an Online Training on

LGBTQ-affirmative Cognitive-behavioral Therapy

Statistical Analysis Plan

Feasibility and Acceptability

Feasibility was assessed in terms of training attendance. Acceptability was assessed in terms of a 7-item measure we created for the present study presented to participants immediately following the intervention (immediate $n = 52$, waitlist control $n = 44$). Sum scores were taken for each participant, and means and standard deviations were recorded for each condition to determine training acceptability.

Intervention Efficacy

Intervention efficacy was assessed using an intent-to-treat analysis and included all participants ($n = 121$). First, to test the effectiveness of randomization, we examined differences in baseline participant demographic and professional characteristics between the immediate intervention condition ($n = 61$) and waitlist control ($n = 60$) using t-tests for continuous measures and chi-square tests for categorical measures. Age, fulltime employment status, and number of years working in LGBTQ mental healthcare differed between conditions ($p < .10$), with the waitlist control group containing participants with a younger mean age, who were less likely to be working fulltime, and who reported fewer years working in LGBTQ mental healthcare. Thus, we included fulltime employment status and number of years working in LGBTQ mental health care as covariates in subsequent analyses. We did not include age as a covariate because of its high correlation ($r = .50$) with number of years working in LGBTQ mental healthcare. Dependent variables were assessed for normality using skewness thresholds of ± 2 and kurtosis

thresholds of ±7 (Byrne, 2010; Hair et al., 2010), and all dependent variables were considered normal except simulated practice skills for general CBT. To meet model assumptions of normal distribution, scores on the simulated practice skills for general CBT scale were square-root transformed prior to efficacy analyses.

To examine intervention efficacy, we used linear mixed models with maximum likelihood estimation and an unstructured covariance matrix to test the Condition × Time interaction for all intervention outcomes including LGBTQ cultural competence, LGBTQ cultural humility, minority stress knowledge, LGBTQ-affirmative CBT knowledge, LGBTQ-affirmative CBT skills familiarity, LGBTQ-affirmative CBT skills use, LGBTQ-affirmative simulated practice skills, and general CBT simulated practice skills. Primary analyses between conditions were limited to baseline (time = 0) and 4-months post-baseline assessment (time =1; changed from 3-months post-baseline prior to the beginning of the study) and examined the Condition × Time effect of receiving immediate intervention (condition = 1) versus receiving the 4-month waitlist (condition = 0). Thus, these models compared pre-intervention and immediate post-intervention outcomes in the immediate intervention group with 4-months pre-intervention and pre-intervention outcomes in the waitlist control group. Effect sizes (*d*) for linear mixed models were calculated as mean pre-post change in the immediate intervention group minus the pre-post change in the waitlist control group, divided by the pooled baseline standard deviation (Morris, 2008).

To examine longer-term persistence of intervention efficacy, we secondarily limited analyses to the immediate intervention group and compared 8-months post-baseline assessment (time = 2; changed from 6-months post-baseline prior to the beginning of the study) to both baseline (time = 0) and 4-months post-baseline assessment (time = 1). Differences between

baseline and 4-month follow-up would indicate delayed efficacy if differences were detected at 4-month follow-up but not post-intervention. Differences between 4-months post-baseline assessment and 8-months post-baseline assessment would reveal whether efficacy was maintained, increased, or decreased from 4-months post-baseline assessment. Effect sizes (*d*) were calculated as mean change from baseline- and 4-months post-baseline to 6-months post-baseline divided by the baseline immediate intervention standard deviation accounting for correlation between time points (Morris & DeShon, 2002).

All results were evaluated at $p < .05$. We report means, standard errors, 95% confidence intervals, and effect sizes.

For the simulated practice videos, two bachelors-level research assistants independently coded the 587 responses (baseline = 235; 4-months post-baseline = 189; 8-months post-baseline = 163). Two post-doctoral researchers trained the two coders in identifying LGBTQ-affirmative CBT principles and intervention strategies based on the content of the training. All responses were given unique identification codes so that coders were masked to the assessment time-point and condition of each response. Inter-rater reliability was assessed via intra-class correlation (ICC) estimates based on a two-way mixed effects model. A post-doctoral researcher calculated ICCs for the average total of LGBTQ-affirmative CBT skills items and the average total of general CBT skills items. The coders initially met with the post-doctoral researchers after every 20 coded responses. The post-doctoral researchers led coding meetings to establish consensus on any inconsistently applied ratings. In cases where the coders could not reach consensus, the post-doctoral researchers met with both coders to resolve the discrepancy. ICC estimates were deemed indicative of poor reliability for values less than .5, moderate reliability between .5 and .75, good reliability between .75 and .9, and excellent reliability greater than .9 (Koo & Li, 2016;

Portney & Watkins, 2000). Once excellent inter-rater agreement was reached, one of the research assistants coded the remaining responses in intervals of 100 responses, while the other research assistant coded 20 of those 100 responses to continue assessing inter-rater reliability. Overall, the average ICC for the LGBTQ-affirmative CBT skills items was .87 and the average ICC for the general CBT skills items was .81.

References

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge/Taylor & Francis Group.

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate data analysis: A global perspective*. Pearson.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364-386. https://doi.org/10.1177/1094428106291059

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105-125. https://doi.org/10.1037/1082-989X.7.1.105

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice.* (2nd ed.). Prentice Hall Health.