Supplementary Online Content

Brenes GA, Danhauer SC, Lyles MF, Hogan PE, Miller ME. Telephone-delivered cognitive behavioral therapy and telephone-delivered nondirective supportive therapy for rural older adults with generalized anxiety disorder: a randomized clinical trial. *JAMA Psychiatry*. Published August 12, 2015. doi:10.1001/jamapsychiatry.2015.1154.

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Details of Monitoring Approach Presented to the DSMB

Recruitment was anticipated to be completed by August, 2013. Although considerable effort (targeted mailings and presentations at community sites) and monetary resources were devoted to recruitment during 2013, an average of only 1.75 participants/month were recruited from January 1, 2013 until the end of August, 2013. Prior to that time (March 2011 through December 2012), the study randomized an average of 5.5 participants/month. There was no preplanned decision to monitor for efficacy or futility in the study, but over the eight month period beginning in 2013, it became clear that meeting the goal of 160 evaluable participants at the 4 month visit within the funding period was going to be challenging given that all potentially eligible participants in the selected counties had already been sent study mailings.

For the November 1, 2013 DSMB meeting, the study statistician (completely blinded until that time) thought that it would be prudent to perform an ad hoc monitoring analysis to aid the DSMB in their recommendations as to whether to continue recruitment or focus the remaining study resources on enhancing follow-up activities. This information was presented in a "closed" session between the study statistician and the DSMB members. The study PI and other staff remained blinded to the results until all follow-up data were collected at the four month visit (the visit where the co-primary outcomes were tested). Conditional power (Proschan, M., Lan, K.K.G., Wittes, J.T. 2006. Statistical Monitoring of Clinical Trials. Springer. New York.) was used to evaluate futility and O'Brien-Fleming boundaries were specified at that time for monitoring efficacy.

The information fraction (i.e. available amount of follow-up information based on the original study sample size) at the time when futility monitoring was conducted was approximately 68% for each outcome (108 HAMA and 110 PSWQA participants of a targeted randomization of 160 participants). The boundaries for the interim evaluation based on O'Brien-Fleming boundaries were determined to be z-statistics of ±2.744 for the PSWQA and ±2.766 for the HAM-A. Given the interim evaluation, the final boundary for each outcome, had study recruitment continued and attained 160 evaluable participants, would be ±2.275 for the PSWQA and ±2.273 for the HAM-A, with an associated total alpha of 0.0125 in each tail across the interim and final evaluations of each hypotheses. Given this interim look, the two-sided p-value to declare significance at the final analysis would have been required to be less than 0.0230 for the PSWQA and 0.0229 for the HAM-A. Because our interim analysis did not include all randomized participants (some had not reached the 4-month visit at that time), we have used these levels to declare significance for our final analysis. Without this interim look, the final boundary to be used for a two-sided test would have been ±2.245 and p-values less than 0.025 (i.e., a Bonferroni correction) would be declared significant.

At the time of the interim evaluation, the pre-specified models were fit to the available data and the z-scores for the intervention effect on each outcome were calculated. The z-statistics for the PSWQ-A outcome at Month 4 was z=3.23 and the z-statistic for the HAM-A at Month 4 was z=0.84. Given that the z-score for the PSWQ-A had already crossed the boundary, conditional power was not calculated for this outcome. For HAM-A, conditional power was computed under an assumption that the study would be able to successfully measure 160 total participants and that the difference between intervention groups for the remaining observations would be equal to the difference in means specified under the alternative hypothesis in the protocol. The conditional power of rejecting the null hypothesis at the 4-month visit under these assumptions was approximately 24%. Given eight months of slow recruitment, the DSMB made a recommendation that the investigators stop recruitment and focus resources on full follow-up of participants already in the study. The Study PI was kept blinded to all results until all 4-month outcomes were collected.

eAppendix 2. Sensitivity Analyses, Primary Outcomes and Secondary Outcomes

Sensitivity Analyses

A sensitivity analysis exploring the effect of missing outcomes on the overall conclusions was conducted for each outcome (HAMA, PSWQ-A, GAD-7, BDI). This approach used Markov chain Monte Carlo multiple imputation to initially fill in non-monotone (i.e., where an outcome may be missing at an early assessment visit but is obtained at a later visit) missing outcomes and subsequently used regression-based multiple imputation to fill in monotone (i.e. from drop-out) missing outcomes based on information from prior visits. Imputation was performed under two assumptions: 1) within treatment; and 2) based on regression relationships in the opposite treatment. This latter approach would potentially be conservative, assuming that the missing data after drop-out is governed by regression-based outcome relationships in the opposite treatment.

Primary Outcomes

HAMA: Conclusions from analyses using multiple imputation were consistent with the results obtained not using imputation. Within group imputation estimated a difference in improvement of -1.73 (95% CI -4.32 to 0.86; p=0.189) and opposite group imputation estimated a difference of -1.33 (95% CI -3.79 to 1.14; p=0.291).

Secondary Outcomes

PSWQ-A: Conclusions from analyses using multiple imputation were consistent with the results obtained not using imputation. Within group imputation estimated a difference in improvement of -3.85 (95% CI -5.92 to -1.78; p=0.0003) and opposite group imputation estimated a difference of -3.14 (95% CI -5.28 to -1.00; p=0.004).

Secondary Outcomes

GAD-7: Conclusions from analyses using multiple imputation were consistent with the results obtained not using imputation. Within group imputation estimated a difference in improvement of -2.34 (95% CI -4.09 to -0.59 ; p=0.009) and opposite group imputation estimated a difference of -2.06 (95% CI -3.73 to -0.40; p=0.015).

BDI: Estimates from analyses using multiple imputation were similar to the results obtained not using imputation. Within group imputation estimated a difference in improvement of -2.74 (95%CI -5.36 to -0.12; p=0.040) and opposite group imputation estimated a difference of -2.46 (95% CI –5.10 to -0.19; p=0.068).