**Supplementary Statistics**

All analyses were conducted in R 1. Treatment outcome analyses were performed using restricted maximum likelihood mixed models (MLMM) that included all randomized participants. In the between-group comparisons, Internet-CBT was considered to be superior to wait-list if the time-by-group interaction was statistically significant at the 0.05 level, i.e., Internet-CBT participants showed larger pre-to post-treatment change on the outcome than the wait-list participants. All models included random intercepts to adjust for baseline differences between individuals. Random slopes, which accounted for between-individual differences in development over time, were included if they significantly improved the goodness of fit of the model. Because most adolescents had two parents who provided data, analyses on the parent-rated outcomes were performed as 3-level models where the assessments were clustered within the parents, who were clustered within adolescents.

The longitudinal analyses of measures that were assessed weekly during treatment included 11 assessment points: the pre-treatment, 9 weekly, and the post-treatment assessment point. One participant in the treatment group provided a post-treatment score on the primary outcome measure that was 2.61 standard deviations above the mean score of the treatment group at post-treatment. Including this score on the analysis reduced the between group effect size by 14% on the primary outcome and between 6% and 8% on the other weekly measured outcomes. Thus, the participant was considered to be an outlier and the provided post-treatment assessment was excluded from the weekly analysis.

Six-month follow-up data provided by the treatment group were analyzed using the MLMM framework, including all the participants in the treatment group regardless of missing data at post-treatment or 6-month follow-up. Separate slopes were estimated for the pre- to post-treatment assessment (Slope 1) and post-treatment to 6-month follow-up assessment (Slope 2). A non-significant Slope 2 was interpreted as maintenance of improvement during the follow-up period while a significant Slope 2 was interpreted as improvement or deterioration during the follow-up period. Slope 1 and 2 were then summed to form the estimated overall pre to 6-month follow-up change in the treatment group. Standard errors for the summed slopes were calculated as

and used to determine the statistical significance of the overall within-group effect over the pre-treatment to 6-month follow-up period.

Estimated means and standard errors at pre-treatment, post-treatment, and 6-month follow-up assessments were calculated by summing the appropriate fixed effects obtained in the MLMM analyses, using the formula above. Between-group effect sizes were calculated as Cohen´s *d* by dividing the beta-coefficient of the interaction effect divided by the estimated MLMM pre-treatment standard deviation 2. Within-group effect sizes *d* were calculated as the estimated pre- to post-treatment change in each group (and Slope 1 and Slope 1+Slope 2 in the 6-month follow-up analyses) divided by the sample standard deviation at pre-treatments. Confidence intervals with a 95% margin for the effect sizes were calculated using 5000 bootstrap replications 3. The bootstrap replications were clustered on adolescents, i.e., if one adolescent was included in a bootstrap replication, all his or her assessment points were included in the bootstrap replication, regardless if they were provided by the adolescent or parent 4. Within-group effect sizes with a confidence interval that did not cross zero were interpreted as a significant pre- to-post treatment improvement within the groups.

**References**

1. Team RC. R: A language and environment for statistical computing. http://www.R-project.org/.

2. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods. 2002;7(1):105-125. doi:10.1037/1082-989X.7.1.105.

3. Kelley K. The Effects of Nonnormal Distributions on Confidence Intervals Around the Standardized Mean Difference: Bootstrap and Parametric Confidence Intervals. Educational and Psychological Measurement. 2005;65(1):51-69. doi:10.1177/0013164404264850.

4. Ren S, Lai H, Tong W *et al*. Nonparametric bootstrapping for hierarchical data. Journal of Applied Statistics. 2010;37(9):1487-1498. doi:10.1080/02664760903046102.