Study title: We-PAP: A Couples-based Intervention for Sleep Apnea

NCT 04759157

Document date: March 12, 2020

4.4 Statistical Design and Power

General considerations

All statistical analyses will be estimated using Stata v15 or higher. Descriptive statistics for Aim 1 will be calculated using all available data, and tests of hypotheses for Aim 2 will be performed using intent-to-treat approaches that include all participants randomized to a treatment condition regardless of whether they complete the intervention or not. Following enrollment of the full sample, treatment group differences in baseline demographic and sleep related functioning variables will be tested. Any variables that emerge as being significantly different between the two treatment groups will be included as covariates in all tests of Aim 2. If more than 3 variables emerge as being significantly different between the treatment arms, the full set of demographic and sleep related functioning variables will be used to estimate a propensity score indicating likelihood of membership in We-PAP relative to information control (IC); this propensity score will be included as a covariate in all tests of Aim 2.

Given the pilot nature of the proposed study, significance tests of study hypotheses will be performed using one-sided tests with $\alpha = 0.10$. Research concerning statistical criteria for pilot testing of clinical trials[110] suggest that these relaxed statistical criteria, with an emphasis on a priori defined confidence intervals, are appropriate because the primary purposes of the proposed study are to evaluate whether there is reasonable evidence that We-PAP outperforms IC and, if so, to use this initial evidence to plan for a larger scale RCT of We-PAP. We will use these criteria with the primary adherence outcomes as a go/no go threshold for proceeding to a larger scale RCT. Provided that We-PAP outperforms IC as hypothesized, we will continue to use these criteria in estimating the 90% CI of the between group difference to determine the necessary sample size for a subsequent RCT. Because of the well-known phenomenon of pilot trials generating larger effect sizes than follow-on RCTs, we will use the lower limit of the 90% CI to anticipate this reduction in effect size and conservatively plan for a larger study.

Sample Size Determination

Study sample size was primarily determined by power analyses of primary adherence outcomes (adherence outcomes in patient and patient and partner sleep quality). Adherence outcomes (% nights with CPAP used >= 4 hours and numbers of hours of CPAP use per night) are available only for the patient and tests of study hypotheses depend on the precision of estimates of group effects at the highest level of nesting (i.e., the individual). As such, power estimates for tests of treatment effects on these variables were generated for standard linear models rather than MLMs.

Power estimates were generated using descriptive statistics for behavioral and cognitive behavioral CPAP interventions vs. IC reported in Wozniak et al.[111]. For % nights with CPAP usage >= 4 hours, We-PAP adherence was assumed to be 47% and IC adherence was assumed to be 28%. For total hours of CPAP usage per night, We-PAP usage was assumed to be 5.44 hours/night and IC usage was assumed to be 4 hours/night with a SD of 2.12 hours/night within each group. Power estimates based on these assumptions indicate that a sample size of 20 couples per group will provide power of .8 or higher to detect between group differences of $d = .68$ for average hours of CPAP usage and O.R. = 4.36. These required effect sizes are consistent with the medium to large differences in adherence outcomes between behavioral treatments for CPAP and IC reported in Wozniak et al., 2014[111]. Power estimates for patient and partners' sleep [actigraphy and diary-assessed] were generated using descriptive statistics for behavioral and cognitive behavioral CPAP interventions vs. IC as reported in Wozniak et al.,[111] as well as in Buysse et al.,[58] and Sweetman et al[112]. Assumptions for secondary outcomes were: sleep efficiency, We-PAP: $M = 82.2$, $SD = 1.16$, IC: $M = 79.99$, $SD = 1.15$; daytime sleepiness, We-PAP: $M = 4.9$, $SD = 0.3$, IC: $M = 5.2$, $SD = 0.45$). Power estimates for secondary outcomes were not available for differential treatment effects of CBT vs. IC for health-related quality of life, cognitive functioning, or relationship quality. Therefore, consistent with other psychosocial outcomes (e.g., depression, anxiety, etc.), differential treatment effects were assumed to be moderate to large, $d = .7$. The availability of these outcomes for both patients and partners permitted estimation of power for MLMs with adjustment made for assumed associations ($r = .3$) between relationship partners within a couple. Power estimates based on these assumptions indicate that a sample size of 20 couples per group will provide power of .8 or higher to detect between group differences of $d = .54$ for all outcomes, which is smaller than or equal to effect sizes for all outcomes described above.

A final consideration in the determination of the sample size for the proposed study was the necessary sample size for stable and precise estimates of treatment effects using multilevel modeling. Maas and Hox's (2005)[113] canonical work on necessary sample sizes for group level effects in MLM determines that group sizes of 20 members and larger is sufficient for generating unbiased estimates. Given that primary questions center on treatment effects in the proposed study, 20 couples per treatment ensures an adequate sample size for generating unbiased estimates of treatment effects thus permitting use of MLM to test study hypotheses.

Analysis Plan

Aim 2

Both primary outcomes will be analyzed using a two-level MLMs as described by the following series of equations:

Level-1: $Y(\text{Adherence})_{ij} = \beta_{0j} + e_{ij}$
Level-2: $\beta_{0j} = \pi_{0j} + \pi_{0[1\text{ to }n]}$*[covariates determined by tests of between group differences at baseline in demographic and sleep variables] + $\pi_{0[n+1]}$*[Treatment] + μ_{0j}

where *i* represents daily measures total hours of CPAP use/whether total hours of CPAP >= 4 hours (0 = no, 1 = yes) over the 3 month follow-up and *j* represents individuals. Treatment is a dummy coded variable with 0 = IC and 1 = We-PAP. The level two intercept in this model, π_{0j}, represents the maximum likelihood estimate of average adherence and $\pi_{0[n+1]}$ provides a test of the primary study hypotheses. Separate models will be run for total hours of sleep and >=4 hours of CPAP use as outcomes estimated using models with Gaussian and binomial distributions respectively.

Secondary outcomes will be analyzed using a three-level MLMs as described by the following series of equations:

Level-1: $Y(\text{Outcome})_{ijk} = \beta_{0jk} + \beta_{1jk}$*(Pre vs. 3 month follow-up) + e_{ijk}
Level-2: $\beta_{0jk} = \pi_{00k} + \pi_{0[1\text{ to }n]k}$*[covariates determined by tests of between group differences at baseline in demographic and sleep variables] + $\pi_{0[n+1]k}$ *(Patient vs. partner) + μ_{0jk}
$\beta_{1jk} = \pi_{1jk} + \pi_{1[n+1]k}$ *(Patient vs. partner)
Level-3: $\pi_{0jk} = \gamma_{000} + \gamma_{001}$*[Treatment] +$r_{00k}$
$\pi_{1jk} = \gamma_{100} + \gamma_{101}$*[Treatment]
$\pi_{1[n+1]k} = \gamma_{1[n+1]0} + \gamma_{1[n+1]1}$*[Treatment]

where *i* represents repeated measurement of the outcome variable, *j* represents individuals, and *k* represents couples. Treatment is a dummy coded variable with 0 = IC and 1 = We-PAP; Pre vs. 3 month follow-up is a dummy coded variable with 0 = pre-treatment and 1 = 3 month follow-up; and, patient vs. partner is a dummy coded variable with 0 = patient and 1 = partner. The level three intercept in this model, π_{0jk}, represents the maximum likelihood estimate of average outcome, π_{1jk} provides a test of differential treatment effects, and $\pi_{1[n+1]k}$ provides a test of whether these effects differ for patients and partners. Separate models will be run for each outcome.

Missing Data. Mechanisms of missingness will be evaluated in outcomes and potential mediators with missing data. If missingness patterns are consistent with MAR assumptions, recently developed fully conditional specification with chained equations methods for multiple imputation of missing data in MLMs will be used to handle missing data[114,115]. All variables in hypothesis tests plus additional auxiliary variables will be included; potential auxiliary variables will be carefully screened using methods recommended for small sample sizes[103]. If missingness patterns are consistent with MNAR assumptions, we will consider the use of pattern mixture models[104] and conditionally dependent dropout[116] models to minimize any potential bias in estimation of treatment effects.