**Hypnosis for the Management of Anxiety and Dyspnea in Pulmonary Rehabilitation - Rationale and design for a cluster-randomized active-control trial [HYPNOBPCO\_2]**

**SUPPLEMENTARY METHODS**

**Index**

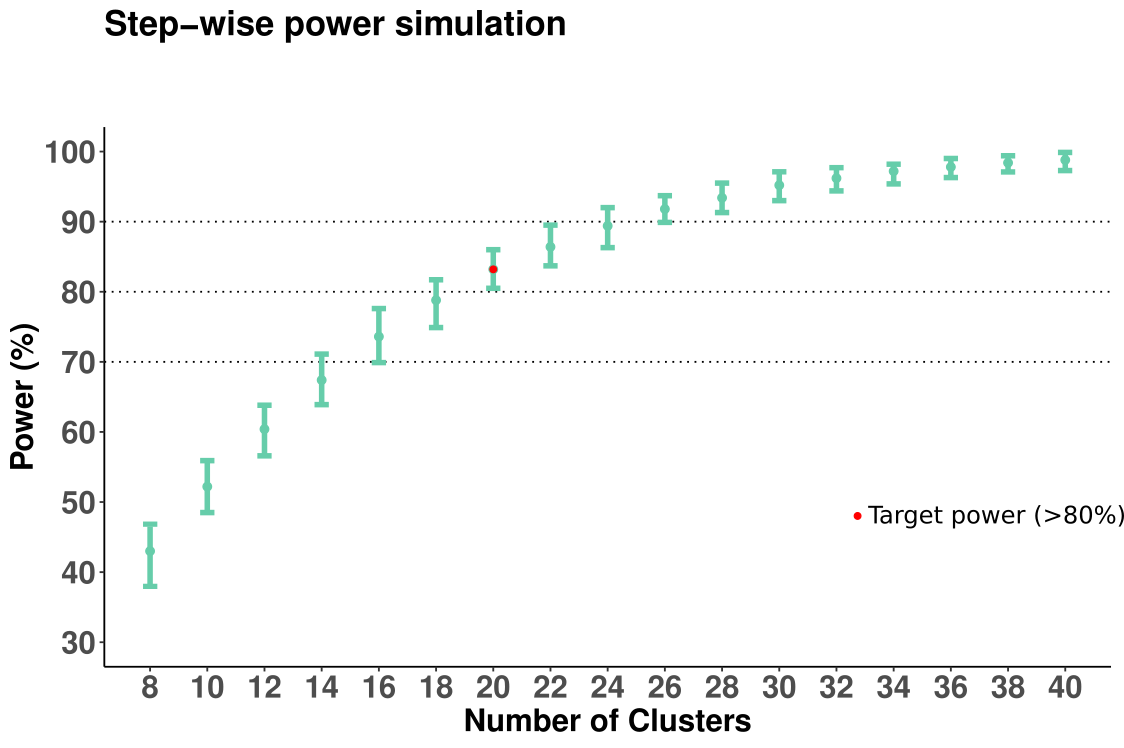
1. Power analyses………………………………………………………………...02
2. References……………………………………………………………………...04
3. ***Power analyses***

In order to estimate statistical power, a simulation-based approach was chosen. This technique has numerous advantages, particularly when in need of using mixed models to control for nested random effects (as is the case of this cluster-randomized trial) (Green & MacLeod, 2016). The procedure consists of iterating three consecutive steps: simulating response values using a model that fits the main hypothesis, refitting the model to the simulated response, and applying a statistical test to the simulated fit. Since the tested effect is assumed to exist, every positive test is considered a true positive, and every negative is treated as a Type II error. Hence, once a model is defined, the power of the test can be calculated from the number of successes and failures at step three.

The actual simulation of responses is straightforward. First, a mean expected effect (with its standard deviation) is established per condition, by piloting and/or looking at the literature. Then, a normal distribution with that mean and standard deviation is assumed for responses, and data per condition is sampled as many times as instances of measurement x participants x clusters the simulated sample is expected to have. For the present study, previous literature [18,19] and the investigators’ precedent trial [10], were used to estimate the size of the effect. It was determined that the PRP+Hypnosis intervention was expected to yield an additional 10% delta (Δ) in anxiety scores when compared to controls (Hypnosis ΔSTAI-6 = -24%, SD = 18% , Relaxation ΔSTAI-6 = -14%, SD = 18%).

Some additional elements had to be taken into account in order to simulate responses in accordance with study design. In the present trial, the primary end-point consists of change in anxiety, as measured by the STAI-6. However, anxiety is monitored in two different ways: additional to total change after PRP completion, anxiety change is assessed weekly throughout the PRP, as to observe if hypnosis effects wane over time. For the purpose of finding a suitable sample size, it was assumed that the effect of hypnosis would indeed wane over time. This was done because sufficient power for detecting a waning effect would also be trivially sufficient for detecting an effect that either grows or remains constant. Thus, simulated hypnosis responses exhibited an additional 10% decrease in anxiety scores from baseline when compared to controls (at Week 2), but this difference waned gradually over time until matching controls by PRP completion (at Week 4). Another crucial element to consider was the trial’s cluster structure: allocation to arms was done by group of patients (i.e. clusters) rather than per patient. Thus, the simulation had to replicate the cluster structure: data was generated for an even number of PRP+Hypnosis and PRP+Relaxation clusters, of 5 patients each (based on the typical number of patients that integrate a standard PRP at the CHB).

We used the lme4 and DoParallel packages in R to conduct these simulations, for sample sizes ranging from 8 to 40 clusters (in steps of 2 clusters).Clusters were divided evenly across arms. Alpha was set at 0.05. To evaluate anxiety changes over time, while simultaneously controlling for intracluster correlation, we fitted the hierarchical nested mixed model Anxiety Change ~ Intervention type X Instance of measurement + (1|Cluster ID/Participant ID). In this model, Intervention type was a 2-level categorical factor (“Hypnosis”, “Relaxation”), while Intervention type was a 3-level ordinal factor (“Week 2”, “Week 3”, “Week 4”, with week 1 being the baseline). It consisted of random intercepts per participant, and per participant/cluster. We established the detection of a significant Intervention Type main effect as criterion for success. For each set of clusters, 100 samples were generated, fitted and tested to obtain a power level. This operation was then repeated 250 times per step to obtain the 95% Confidence Interval for each power level. Figure SM 1 shows power for all samples. Under these conditions, a sample of n=100 (20 clusters, 10 clusters per condition) yielded a power of 83% (95% CI 80, 86). In anticipation of a 20% attrition level, the target sample was set at n = 120 (i.e., 12 clusters per arm).



**Figure SM1. Simulation-based power analysis.** Power for detecting a significant Intervention Type main effect as a function of sample size (alpha = 0.05). Error bars represent the 95% Confidence Interval for power at each sample. Black dotted lines: 70%, 80% and 90% power threshold; red dot: first projected sample size with power > 80%.

1. ***References***

Green, P. and MacLeod, C.J. (2016), SIMR: an R package for power analysis of generalized

linear mixed models by simulation. Methods Ecol Evol, 7: 493-498.

<https://doi.org/10.1111/2041-210X.12504>