Statistical Analysis Plan

5.2. Primary analysis

Trial data will be summarised using a CONSORT diagram and all analyses will be based on *intention-to-treat* principles (e.g., including all available data for participants who never attended intervention sessions, those who attended all sessions, and those who dropped out after a few sessions). Intention-to-treat principles will be followed to minimise well-known biases present in "completer analysis" (e.g., only including data from those who complete the intervention, who may not be representative of the wider population of eligible participants in need for treatment) and to follow best practice principles in healthcare research (White et al., 2011). Missing data will be imputed using an expectation-maximization algorithm, prior to conducting formal analyses.

The primary hypothesis test will be based on comparing mean OLBI (total severity) scores between groups at week 10 (post-intervention), as shown in Figure 1. Mean OLBI scores (dependent variable) will be compared between groups using ANCOVA, adjusting for baseline severity and entering "intervention group" as an independent variable. Adjusted 95% confidence intervals will be calculated around the adjusted mean difference between interventions. The analysis will be conducted using the Statistical Package for the Social Sciences (SPSS) by a researcher who will be blind to the label of the interventions. The primary analysis will be conducted using imputed data, following intention-to-treat principles. Secondary analyses will repeat the above ANCOVA model using per protocol analysis (only including participants that actually started the intervention; and excluding data from those in the 'non-intervention' group) and an unimputed dataset (e.g., completers analysis), to test the robustness and stability of the main analysis.

5.3. Secondary analyses

The analysis described above will be repeated at the 44-week follow-up. In addition, these analyses will be repeated at each of the post-intervention time-points illustrated in Figure 1 (weeks 10, 44), using the OLBI sub-domain scores (*exhaustion*; *disengagement*), the WEMWBS well-being measure, and the JDSS job satisfaction measure, controlling for baseline scores. These between-group comparisons will also be summarised using effect sizes (Cohen's *d*).

Post-intervention measures (weeks 10, 20, 44) will be compared to baseline measures (week 0) within each group, using paired-samples t-tests or an appropriate non-parametric test depending on the distribution of the data. Within-sample pre-post treatment effect sizes will also be computed using the method proposed by Minami et al. (2008).

Dropout will be defined as attending less than 4 (half) of the intervention sessions. We will examine predictors of dropout and treatment response (defined as reliable improvement in the OLBI measure), using all available baseline measures as candidate predictors. Reliable predictors of each of these outcomes of interest will be identified using a variable selection method called Elastic Net Regularization (Zou & Hastie, 2005) applied in separate logistic regression models for each dependent variable.

Additional secondary analyses will examine variability in burnout, wellbeing and job satisfaction across groups defined by their job role and demographic characteristics. Between-group comparisons will be made using t-tests (or Mann-Whitney U Tests if the data are not normally distributed).

Qualitative data collected via online surveys after all participants had the opportunity to access the intervention (week 20) will be analysed using thematic analysis (TA). TA will follow the 6-step phase of analysis proposed by Braun and Clarke (2006). We will triangulate qualitative themes generated via the TA method with baseline survey data in order to explore if the themes vary according to job role (e.g., junior doctors, consultants, etc.) and setting (e.g., community vs. hospital services).