5.1. Sample size calculation

A sample size calculation was performed using the method described by Cohen (1992). There is no precedent for this type of trial in this setting, so it is not known whether exposure to the Home Goals intervention may be associated with small, moderate or large effects on measures of depression, anxiety and physical activity. We have therefore followed conventional sample size calculation methods described by Cohen (1992), expecting a moderate effect size as a conservative assumption. In order to detect a moderate effect size using between-groups ANOVA, with 80% power, and an alpha level of 0.05, we estimate that at least 67 participants are needed per group. This would yield a sample size of 134. Expecting a 30% dropout rate, which is common in studies of psychological interventions, we would need to inflate the recruitment target to 192. If we recruit more than the 192, we will accept the young people onto the trial, but we will not collect or analyse their data.

5.2. Primary analysis

Trial data will be summarised using a CONSORT diagram and all analyses will be based on *intention-to-treat* principles. Missing data will be imputed using an expectation-maximization algorithm (Schafer & Olsden, 1998), prior to conducting formal analyses.

The primary hypothesis test (A) will be based on comparing mean outcome measures between groups at week 8, as shown in Figure 1. Mean PHQ-A scores will be compared between groups using analysis of covariance (ANCOVA), controlling for baseline severity.

Sensitivity analyses will be performed to assess the robustness of the main findings. This will involve repeating the above analysis including a random intercept for each trial site and additionally introducing "role" as a covariate (admin; mental health; other clinical role) and any other baseline demographics that may be unbalanced between groups.

1.3. Secondary analyses

ANCOVA (and sensitivity analyses) described above will be repeated at each of the time-points illustrated in Figure 1 (weeks 0, 6, 12, 36), using the PHQ-A, and using the GAD-A as an outcome, controlling for baseline scores. Outcomes at 6-months follow-up (pooled for both groups) will be compared to outcomes at the baseline assessment, using paired-samples t-tests (or an appropriate non-parametric test depending on the distribution of the data).

These analyses will be repeated at each of the post-intervention time-points illustrated in Figure 1 (weeks 6, 36), using the WEMWBS sub-domain scores. These between and within-group comparisons will also be summarised using effect sizes (Cohen's *d*).

Post-intervention measures (weeks 6, 36) 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).

We will carry out an exploratory analysis of knowledge and levels of activity before and after the intervention.

5.4. Exploratory analyses of mechanisms of change

Change scores will be computed denoting changes in theoretical mechanisms measured using the questionnaires listed in pg. 7. These change scores will be computed between time-point 0 and time-points 6, and 12, denoting changes within the intervention phase.

Residualized change scores across multiple theoretical mechanisms will be entered into a Bayesian network analysis carried out separately for each intervention group, using the post-treatment (weeks 6, 12) PHQ-A as the dependent variable. Each Bayesian network will be trained using a Tree-Augmented Naïve Bayes (TAN) algorithm with adjustment for small cell counts (Friedman, Geiger, & Goldszmidt, 1997). This method offers a data-driven way to model a network of relationships (called *attribute dependencies*) between predictors and their joint influence over a

target outcome (post-treatment OLBI). TAN produces a simple and parsimonious network model where each predictor is allowed to depend on one additional predictor, thus modelling multiple two-way interactions. Variable selection will be performed using a 10-fold cross-validation procedure (Rodriguez, Perez, & Lozano, 2010). The resulting treatment-specific network models will be visualised using directed acyclic graphs (Shrier & Platt, 2008), which enable an intuitive interpretation of key mechanisms of change, their relative importance, and their interrelationships.