**Post-hoc exploratory analysis of the Engager Randomised Controlled Trial data**

We wanted to understand if the baseline differences were related to any other factors and also to see of these baseline differences may have made a difference to the outcome measures at baseline. Therefore, we looked a differences between the groups on the primary and some of the secondary outcomes and also looked to see if the way in which people had ‘screened in’ to the trial may have had an impact.

**Method**

We compared baseline scores of the Engager group and usual care group using t-tests for continuous variables, chi2 tests for categorical variables and the Shapiro-Wilk test to test for normality of the distribution of CORE-OM scores. For the latter, we present p-values with p<0.05 indicating that the hypothesis of a normal distribution can be rejected. For outcomes where Shapiro-Wilk test indicated non-normality, we used Kruskal-Wallis test to test for differences. Pearson’s correlation coefficients was used to examine the relationship between baseline CORE-OM scores and other measures. All analyses were conducted using STATA version 15.

**Results**

Generally, participants in the Engager group and usual care group were similar in terms of their baseline scores on the outcome measures (See Table 1). There were no statistically significant difference in baseline scores of the CANFOR (Thomas et al 2003),1 ICECAP (Al-Janabi et al., 2012)2 and IOMI (Maguire et al 2017)3. There was, however, a significant difference in CORE-OM baseline scores with the Engager group having lower scores than the usual care group. CORE-OM severity at baseline was also more often categorised as “not distressed” or “mild” in the Engager group than in the usual care group, indicating overall less psychological distress in the Engager group. There were also significantly more participants with CORE-OM scores less than 10 (not distressed) in the Engager group than the usual care group (p=0.009).

**Table 1: Between-groups differences of outcome measures**

|  |  |  |
| --- | --- | --- |
|  | Engager Group | Usual Care Group |
|  |  |  |
| Outcomes (mean; SD) |  |  |
| CORE-OM\* | 15.2; 6.0 | 16.9; 6.2 |
| CANFOR | 11.0; 4.8 | 11.5; 5.5 |
| ICECAP | 0.6; 0.2 | 0.6; 0.2 |
| IOMI | 67.3; 11.8 | 65.9; 11.2 |
|  |  |  |
| Normal distribution (p-value) |  |  |
| CORE-OM | 0.033 | 0.485 |
| CANFOR | 0.116 | 0.423 |
| ICECAP | 0.016 | 0.087 |
| IOMI | 0.793 | 0.723 |
|  |  |  |
| CORE-OM Severity (n) |  |  |
| Not distressed\*\* | 34 | 17 |
| Mild | 34 | 33 |
| Moderate | 40 | 52 |
| Moderate to severe | 23 | 22 |
| Severe | 9 | 16 |

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

There were significant differences in screening in scores between the groups (Table 2). Participants in the Engager group had significantly lower GAD scores, indicating lower anxiety levels. Although not statistically significant, this group also had lower PHQ and PTSD scores, corresponding to lower depression and stress levels, and less participants had past/likely future common mental health problems.

Comparing the number of tools participants screened in on, more participants in the Engager group screened in on only one or two tools.

**Table 2: Between-groups differences of screening scores**

|  |  |  |
| --- | --- | --- |
|  | Engager Group | Control Group |
|  |  |  |
| Screening scores (mean; SD) |  |  |
| PHQ | 12.9; 5.6 | 13.9; 5.2 |
| GAD\* | 11.1; 5.3 | 12.4; 4.9 |
| PTSD | 2.1; 1.7 | 2.3; 1.7 |
| Past/Future common MH (n; %) | 121 | 121 |
|  |  |  |
| Number of tools screened in on (n) |  |  |
| 1 | 25 | 19 |
| 2 | 33 | 25 |
| 3 | 40 | 47 |
| 4 | 42 | 49 |
|  |  |  |

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Examining the relationship between screening in tools and baseline CORE-OM scores, we identified a strong correlation between CORE-OM and PHQ and GAD scores, as well as the number of tools a participant screened in on (Table 3). PTSD scores and past/future common mental health problems were both moderately correlated with CORE scores at baseline.

**Table 3: Correlation between screening tools and baseline CORE scores**

|  |  |
| --- | --- |
| Screening tools | Correlation coefficient |
|  |  |
| PHQ score | 0.6232 |
| GAD score | 0.6186 |
| PTSD score | 0.3664 |
| Past/future common MH | 0.3611 |
| Number of tools screened in on | 0.6016 |

Therefore, we found statistically significant imbalances in CORE-OM at baseline between the groups, with the Engager group scoring as less mentally unwell. We also found that there were imbalances between the groups on their ‘screening in’ scores with the Engager group having lower anxiety, depression and PTSD levels. This screening in difference was to be expected as CORE-OM scores correlated highly with GAD-7 and PHQ-9. This, in addition to the other baseline imbalances, suggested to us that the usual care group may have struggled with being in prison as they had higher levels of common mental health problems when in prison, but may have had less social disadvantage on release into the community. Conversely, the Engager group may have been less affected by being in prison, as they had greater social disadvantage in the community.

Due to these differences in CORE-OM at baseline and that scores appeared to be sensitive current experiences we investigated further the stability of the CORE-OM.

**Temporal changes to CORE-OM Methods**

We examined the changes in mean CORE-OM at baseline vs 6 months follow-up using Cohen’s effect size *d*. We calculated the effect size for all participants, as well as for each group. We also examined for follow-ups which took place in the community and those which took place in prison separately. The effect size was assumed to be small when Cohen’s *d* was at least 0.2, medium when at least 0.5 and large when at least 0.8.

**Results**

The drop in CORE-OM score was somewhat larger for the usual care group than for the Engager group. But the usual care group had a higher score to begin with, so had the potential to show a bigger change. The effect size was also larger in the subgroup of participants who were follow-up in the community compared to those with follow-up in prisons, with the usual care group showing a *d* value of 0.98. This pattern was not true for the Engager group, which had a slightly lower effect size in the sub-analysis of follow-ups in the community (see Tables 4-6).

**Table 4: Temporal changes in the CORE-OM between baseline and follow-up**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Baseline** | **6 months** | **Change** | **Cohen’s d1** |
|  | **n** | **Mean** | **S.D.** | **Mean** | **S.D.** | **Mean** | **S.D.** |  |
| All  | 182 | 15.86 | 5.97 | 12.28 | 7.28 | -3.58 | 7.58 | -0.6 |
| Engager group | 92 | 15.04 | 5.92 | 12.61 | 6.87 | -2.43 | 6.83 | -0.41 |
| Usual Care group | 90 | 16.70 | 5.94 | 11.94 | 7.7 | -4.77 | 8.16 | -0.8 |

1calculated as (M2-M1)/S.D.1; effect sizes: d=0.2=small; d=0.5=medium; d=0.8=large

**Table 5: Temporal changes in the CORE-OM between baseline and follow-up among participants with follow-up in the community**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Baseline** | **6 months** | **Change** | **Cohen’s d1** |
|  | **n** | **Mean** | **S.D.** | **Mean** | **S.D.** | **Mean** | **S.D.** |  |
| All  | 98 | 15.81 | 6.06 | 11.72 | 7.39 | -4.09 | 7.89 | -0.67 |
| Engager group | 44 | 14.73 | 6.01 | 12.87 | 7.28 | -1.86 | 6.95 | -0.31 |
| Usual Care group | 54 | 16.69 | 6.01 | 10.78 | 7.41 | -5.91 | 8.20 | -0.98 |

1calculated as (M2-M1)/S.D.1; effect sizes: d=0.2=small; d=0.5=medium; d=0.8=large

**Table 6: Temporal changes in the CORE-OM between baseline and follow-up among participants with follow-up in prisons**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Baseline** | **6 months** | **Change** | **Cohen’s d1** |
|  | **n** | **Mean** | **S.D.** | **Mean** | **S.D.** | **Mean** | **S.D.** |  |
| All  | 48 | 16.93 | 5.87 | 13.65 | 7.52 | -3.28 | 7.73 | -0.56 |
| Engager group | 25 | 15.99 | 6.03 | 13.64 | 6.93 | -2.35 | 7.44 | -0.39 |
| Usual Care group | 23 | 17.95 | 5.65 | 13.67 | 8.27 | -4.29 | 8.09 | -0.76 |

1calculated as (M2-M1)/S.D.1; effect sizes: d=0.2=small; d=0.5=medium; d=0.8=large

Therefore, mean scores for the CORE-OM dropped differentially over time depending on trial arm and location of follow-up. Compared to a previous study (Barkham et al., 2007)4 assessing the temporal stability of the CORE-OM, the effect size in our population was larger. The drops for those in the usual care arm followed up in the community were greatest, providing for the possibility that CORE-OM scores and location of follow up (prison or not) interact.

**References**

1. Thomas, S., Harty, M.A., Parrott, J., McCrone, P., Slade, M. & Thornicroft, G. CANFOR: Camberwell assessment of need-forensic version. London: Gaskell. 2003.
2. Al-Janabi, H., Flynn, T. & Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Quality of Life Research, 2012;21:167-176. doi: 10.1007/s11136-011-9927-2
3. Maguire, M., Disley, E., Liddle, M., Meek, R. & Burrowes, N. Developing a toolkit to measure intermediate outcomes to reduce reoffending. London: Ministry of Justice Analytical Series. 2017.
4. Barkham, M., Mullin, T., Leach, C., Stiles, W.B. & Lucock, M. Stability of the CORE-OM and the BDI-I prior to therapy: evidence from routine practice. Psychol Psychother. 2007; 80(Pt 2):269‐278.