BJPsych | The British Journal of Psychiatry
1. doi: 10.1192/bjp.bp.111.098681

Online supplement

Statistical analysis

Primary analysis

All analyses were done in Stata version 11 on an intention-to-treat basis, i.e. all patients and all available data were included in the analysis. We fitted a mixed model with random intercept for the 36-item Short Form Health Survey (SF-36) aggregate score (primary outcome) at baseline, 4 months, 10 months and 16 months, with separate treatment effects calculated for each time point (i.e. the model treated time as a fixed categorical effect). The model adjusted for gender, age, work status, lifetime psychiatric comorbidity and clinician-rated impairment, and was corrected with a cluster effect for treatment group. Adjustment variables were defined before commencing analyses, and chosen since we regarded them to be potential moderators of change. The model was checked by diagnostic plots of the residuals.

Using this mixed model we first tested whether the two groups differed with regard to changes over time on the primary outcome, i.e. tested the hypothesis that the interaction estimates of time group for all three follow-up time points were equal to zero. This was done by means of the Wald χ^2 test (see Fig. 3). In a next step, adjusted change scores from baseline to 4 months, 10 months and 16 months were calculated for each group for the primary outcome. We then calculated comparison effect sizes (adjusted Cohen's *d*) for each time point by dividing the interaction estimate of group time (which represents the adjusted between-group difference in mean change from baseline to this time point) by the pooled standard deviation at baseline (see Fig. 3(a)). The same statistical model was used for the analysis of the standard Physical Component Summary (PCS) of the SF-36, which is provided for comparison only, and for the secondary outcome measures (see Fig. 3(b–f)).

Calculation of probability of treatment response and number needed to treat

Probabilities of treatment response (i.e. improvement of at least 4 points on the SF-36 aggregate score from baseline to 16 months) in both groups were calculated from log odds derived from a simple logistic regression model (i.e. a model with intervention as the only covariate). The calculation of probabilities on group level does not allow adjustment for baseline variables. We calculated relative risks (RRs) from a simple generalised linear model with a binomial family and a log link function. Number needed to treat (NNT) was estimated through 1/risk difference between the two groups. Both RR and NNT were estimated using unadjusted data and stated with 95% confidence intervals.

Sensitivity analysis

A sensitivity analysis was carried out for the primary outcome, using the same mixed model with random intercept as described above, but based on multiple imputation of missing outcomes. We did this analysis since we found considerable drop-out at 16 months, to ensure that the results were stable and that the reported group differences could not be explained by attrition. The multiple imputations were made by means of a multivariate normal data augmentation method (50 unique data-sets), applied to the intervention and comparison group separately, and where each used an additional measurement of the respective scale (obtained at referral, i.e. before the clinical assessment) together with the adjustment variables as covariates.

Results based on multiple imputations (sensitivity analysis)

Results based on multiple imputations differed only marginally from the main analysis, and did not change the interpretation of the trial results. Based on multiple imputation, the adjusted difference in mean SF-36 aggregate score change from baseline to 16 months was 4.1 points (95% CI 1.5–6.7, $P=0.002$). Patients in the Specialised Treatment for Severe Bodily Distress Syndromes (STreSS) intervention group improved 4.0 points (95% CI 2.0–6.0, $P<0.001$), whereas patients in the enhanced usual care group showed no improvement (-0.1 points, 95% CI -1.7 to 1.6, $P=0.94$).

Based on multiple imputation, probability of treatment response (i.e. improvement of at least 4 points on the SF-36 aggregate score) was 0.47 (95% CI 0.32–0.61) in the STreSS group and 0.26 (95% CI 0.14–0.38) in the enhanced usual care group. The relative risk for treatment response was 1.84 (95% CI 1.02–3.21, $P=0.04$) in favour of STreSS. The number needed to treat (NNT) to achieve on additional treatment response with STreSS compared with enhanced usual care was 5 (95% CI 3–53, $P=0.03$).

Conclusions

The sensitivity analysis where all missing values were replaced by means of multiple imputations yielded very similar results to the intention-to-treat-analysis based on observed data. Attrition did not introduce a bias in the trial results, and the results reported in our paper seem to be statistically robust.