Supplement Table of Contents

Further specification of statistical analyses

The fixed part of the (G)LMM analyses contained treatment (as centered dummy, 0.5 for ST and -0.5 for TAU), time (0,1,2,3,4,5,6 for the half-year intervals) and their interaction. Note that for GLMM survival analysis the time variable has to be categorical, in contrast to the other (G)LMM analysis, where time was dimensional. Before analyses the time development in the treatments was inspected, and where appropriate, either a piecewise regression model was employed (two linear time slopes for different periods in the 3-year period) or a quadratic time effect and a treatment-by-time^2 interaction were added. Covariance structures for the repeated part were determined by comparing fit indices of different models, notably Chi-square tests on the -2LL values. We calculated two effect sizes: for the size of the effects in the (G)LMM analyses the effect size *r*, with $r=\sqrt{(t^2/(t^2+d.f.))}$, which corresponds to the t-tests of the fixed part of the (G)LMM; and as a conventional effect size Cohen's *d*, with *d*=estimated change/SD, with the estimated change derived from the fixed part of the (G)LMM analysis and the SD being the square root of the variance based on the random and residual parts of the (G)LMM analysis. For the F-tests in the fixed part we used as effect size $r = \sqrt{(d.f._n * F/(d.f._n * F + d.f._d)}$, with $_n$ = numerator, $_d$ = denominator. Note that for F-tests *f* is often used as effect size, with $f = \sqrt{(r^2/(1-r^2)}$, and for small *r*, $f \approx r$.

Primary outcomes

Supervised and unsupervised leave **S**upervised leave and unsupervised leave were analyzed by means of GLMM survival analysis. The first half year that stable supervised respectively unsupervised leave was attained was target variable, with treatment, time (6-months periods as factor), and time by treatment interaction as predictors. A binomial model with a complementary log-log link was used. Site was added as random intercept.

SNAP-FV PD Scales The SNAP ratings on the four PD scales were analyzed with LMM with in the random part the four sources of SNAP ratings (patient, observers 1-3) nested within patient, SNAP subscale (antisocial, borderline, narcissistic, paranoid) nested within rater (within patient), and time (linear time effect) nested within SNAP subscale nested within rater. The fixed part contained time (linear), treatment, and their interaction, rater, SNAP-subscale. For the initial model all interactions of rater and SNAP-subscale with time and treatment were added. Stepwise deletion was used to delete higher order interactions not significant at p=.05, starting with the 4-way interaction. Main effects were retained in the model, as was the treatment x time interaction. The effect size (Cohen's *d*) over 3 years per treatment and of the difference between treatments was calculated on the basis of the fixed effect coefficients, and for the *SD* the residual variance and the patient intercept variance of the random part.

Secondary outcomes

SNAP-FV Temperament Scales The SNAP ratings on the three temperament scales were analyzed with LMM with in the random part the four sources of SNAP ratings (participant, observers 1-3) nested within participant, SNAP subscale (Positive Emotions (reversed), Negative Emotions, Disinhibition) nested within rater (within participant), and time (linear time effect) nested within SNAP subscale (within rater within participant). Addition of a random intercept for site led to estimation problems, and was therefore not done for the final model. The initial fixed part contained time (linear), treatment, rater, and SNAP-subscale, and their interactions. Stepwise deletion was used to delete higher order interactions not significant at p=.05, starting with 4-way interaction. Main effects were retained in the model, as was the treatment x time interaction. The effect size (Cohen's *d*) over 3 years per treatment and of the difference between treatments was calculated on the basis of the fixed effect coefficients and the residual variance and the participants' intercept of the random part.

START The START scales (with the Strengths-scale inversed) showed a highly similar pattern over time and were therefore analyzed with a multivariate analogue of mixed regression, with START-scale as level

nested in participant, and time nested within START-scale (nested within participant). The random part contained these nested variables (variance structure) as well as a random intercept for site. As distributions were skewed, mixed gamma regression with a log-link was used, with the scales transformed so that 0.1 (and not 0) was the minimum (as gamma regression cannot handle zero's). Visual inspection further indicated a curvilinear time development; hence time-squared was added to the fixed part (adding it to the random part created estimation problems).

HCR-20^{V2} HCR-out scores were analyzed with mixed regression with CS covariance structure for the repeated part, as best fitting. Visual inspection showed a different time development between treatments with a change in linear slope in ST at 1.5 year, therefore a piecewise regression model was used with the knot value at 1.5 years. The fixed part contained the piecewise time model, treatment, and their interactions; the random part an intercept and time-slope for site. Cohen's d was estimated from the coefficients of the fixed part with the *SD* based on the residual variance (*SD*=4.89).

Early maladaptive schemas YSQ scores had a skewed distribution and were therefore analyzed with mixed gamma regression with a log link, and CS covariance structure for the repeated part, as the best fitting structure. The fixed part contained time (linear), treatment, and their interaction. Site was random intercept. Cohen's d was estimated from the coefficients of the fixed part with the *SD* based on the CS variance (*SD*=0.303).

Maladaptive schema modes SMI maladaptive scores showed a skewed distribution and a curvilinear change over time. Scores were therefore analyzed with GLMM gamma regression with a log-link, a CS covariance structure for the repeated part, as the best fitting structure, and in the fixed part time (linear), time squared, treatment, and treatment by time linear and time squared interactions. The random part contained a random slope for participant (nested under site). Addition of a random intercept for site created estimation problems, and was therefore not done. Cohen's *d* was estimated from the coefficients of the fixed part with the *SD* based on the residual variance (*SD*=.313).

Healthy schema modes SMI healthy scores showed a skewed distribution and a curvilinear change over time. Inversed scores were therefore analyzed with GLMM gamma regression with a log-link, a CS covariance structure for the repeated part, as the best fitting structure, and in the fixed part time (linear), time squared, treatment, and treatment by time linear and time squared interactions. The random part contained an intercept for site and a random slope for participant (nested under site). Cohen's *d* was estimated from the coefficients of the fixed part with the *SD* based on the residual variance (*SD*=.286).

Incidents The weighted incident sum was analyzed with GLMM negative binomial regression with a log-link with AR1 covariance structure for the repeated part, as the best fitting structure, and a random intercept for site and random intercept and slope (time) for participant. The fixed part contained time (linear), treatment, and their interaction. Cohen's *d* was estimated from the coefficients of the fixed part with the *SD* based on the sum of the AR1 and the random intercept variances (*SD*=1.40; thus based on the transformed scale).

SCL-90 The LMM analysis with CS covariance structure for the repeated part, as the best fitting structure, did not converge when site was added as random intercept. Therefore the random intercept was left out of the model. The fixed part contained time (linear), treatment, and their interaction. Cohen's d was estimated from the coefficients of the fixed part with the SD based on the CS variance (*SD*=.40).

Treatment retention

Treatment retention was analyzed with GLMM survival analysis, with treatment, time (categorical) and their interaction as predictors in the fixed part. The random part had site as intercept. Because of quasi-separation in the data when divided per 6 months the data were analyzed with year as time variable. Robust covariance estimators were used.