

Running title: Long-term treatment outcome for two types of MBT

Day hospital versus intensive outpatient mentalization-based treatment: 3-year follow-up of patients treated for borderline personality disorder in a multicentre randomized clinical trial

Maaïke L. Smits*, Dine J. Feenstra, Dawn L. Bales, Matthijs Blankers, Jack J. M. Dekker, Zwaan Lucas, Jan H. Kamphuis, Jan J. V. Busschbach, Roel Verheul, Patrick Luyten

* Corresponding Author

Maaïke Smits

Viersprong Institute for Studies on Personality Disorders

PO Box 7, 4660 AA Halsteren, the Netherlands.

E-mail: maaïke.smits@deviersprong.nl

Trial registration: Nederlands Trial Register. NTR2292. Registered April 16, 2010.

Keywords: mentalization-based treatment, borderline personality disorder, randomized clinical trial, treatment intensity, long-term follow-up

This research was supported in part by a grant from ZonMw (grant no. 171202012).

Word count: 4676

Number of figures and tables: 5

Abstract

Background. Two types of mentalization-based treatment (MBT), day hospital MBT (MBT-DH) and intensive outpatient MBT (MBT-IOP), have been shown to be effective in treating patients with borderline personality disorder (BPD). This study evaluated trajectories of change in a multi-site trial of MBT-DH and MBT-IOP at 36 months after the start of treatment.

Methods. All 114 patients (MBT-DH $n = 70$, MBT-IOP $n = 44$) from the original multicentre trial were assessed at 24, 30 and 36 months after the start of treatment. The primary outcome was symptom severity measured with the Brief Symptom Inventory. Secondary outcome measures included borderline symptomatology, personality and interpersonal functioning, quality of life, and self-harm. Data were analysed using multilevel modelling and the intention-to-treat principle.

Results. Patients in both MBT-DH and MBT-IOP maintained the substantial improvements made during the intensive treatment phase and showed further gains during follow-up. Across both conditions, 83% of patients improved in terms of symptom severity, and 97% improved on borderline symptomatology. No significant differences were found between MBT-DH and MBT-IOP at 36 months after the start of treatment. However, trajectories of change were different. Whereas patients in MBT-DH showed greater improvement during the intensive treatment phase, patients in MBT-IOP showed greater continuing improvement during follow-up.

Conclusions. Patients in both conditions showed similar large improvements over the course of 36 months, despite large differences in treatment intensity. MBT-DH and MBT-IOP were associated with different trajectories of change. Cost-effectiveness considerations and predictors of differential treatment outcome may further inform optimal treatment selection.

Previous studies have provided evidence for the efficacy of two types of mentalization-based treatment (MBT) for borderline personality disorder (BPD): day hospital MBT (MBT-DH), a treatment involving day hospitalization of patients 5 days per week, and intensive outpatient MBT (MBT-IOP), an outpatient treatment program conducted 2 days per week (Bales et al., 2014; Bales et al., 2012; Barnicot & Crawford, 2019; Bateman & Fonagy, 1999, 2001, 2008, 2009; Jørgensen et al., 2014; Jørgensen et al., 2013; Kvarstein et al., 2015; Laurensen et al., 2018). However, only one trial has directly compared the two treatment programs. Smits et al. (2019) found that both MBT-DH and MBT-IOP were associated with substantial improvements on both primary and secondary outcome measures, representing moderate to large effect sizes 18 months after the start of treatment. Although MBT-DH was not superior to MBT-IOP in terms of changes on the primary outcome measure (symptom severity), MBT-DH showed a trend towards superiority on secondary outcome measures, particularly on measures of relational functioning. Longer-term follow-up data are thus needed to further determine the relative efficacy of MBT-DH and MBT-IOP.

As MBT aims to improve mentalizing, and improvements in mentalizing are thought to underlie 'broaden-and-build' cycles leading to improved emotion regulation, feelings of autonomy and agency, and improved capacity for relatedness (Fredrickson, 2001; Luyten & Fonagy, 2014), patients in MBT are expected to show ongoing improvement after treatment termination. Consistent with this hypothesis, previous follow-up studies have shown that the gains made during the intensive treatment phase of MBT-DH were maintained and that patients continued to improve over 18-month and 5-year follow-up periods for the MBT-DH program (Bales et al., 2014; Bateman & Fonagy, 2001, 2008), although for MBT-IOP, ongoing improvement was not found in one 18-month follow-up study (Jørgensen et al., 2014).

Several studies have focused on the impact of treatment modality or intensity on treatment outcome in BPD. For instance, a non-randomized study by Bartak et al. (2011) investigated the differential effectiveness of three treatment modalities (inpatient, day hospital and outpatient treatment) and consequent varying degrees of treatment intensity for cluster B personality disorders, yielding somewhat inconclusive results. Although patients improved in all treatment modalities, at 18-month follow-up inpatient treatment was associated with marginally significant better outcomes in terms of psychiatric symptoms, compared with outpatient treatment. There were no differences in terms of quality

Running title: Long-term treatment outcome for two types of MBT

of life and interpersonal functioning (Bartak et al., 2011). At 60-month follow-up, inpatient treatment was associated with slightly better outcomes compared with outpatient treatment in terms of quality of life, but not on other outcome measures (Horn et al., 2016). Another non-randomized study, by Chiesa, Fonagy, Holmes, and Drahorad (2004), showed significantly better outcomes in a heterogeneous sample of patients with personality disorder in a step-down program compared with longer-term inpatient treatment and general psychiatric services. At 1- and 2-year follow-up, patients in the step-down group showed a greater improvement in symptom severity, social functioning and self harming behaviors and reported less use of outpatient treatment and readmissions to psychiatric services after discharge. These findings were maintained at 6-year follow-up (Chiesa, Fonagy, & Holmes, 2006). Nevertheless, a more recent meta-analysis of psychotherapy for BPD by Cristea et al. (2017) found that neither treatment duration nor treatment intensity was related to treatment outcome. A number of recent randomized controlled trials are addressing the issue of treatment intensity (Juul et al., 2019; McMain et al., 2018), but these studies are still ongoing. Clearly, more research in this area is needed.

The current paper reports on the findings at 3-year follow-up from the abovementioned randomized clinical trial comparing MBT-DH and MBT-IOP for patients with BPD (Smits et al., 2019). Because MBT-DH and MBT-IOP differ markedly in terms of treatment intensity, we expected that the tendency for MBT-DH to be superior to MBT-IOP would be more pronounced at 36-month follow-up, given the substantially higher dose of treatment.

Method

This study was approved by the Medical Ethical Committee of Erasmus Medical Center, Rotterdam, the Netherlands (NL38571.078.12). A total of 114 patients from three sites in the Netherlands were randomized to MBT-DH ($n = 70$) or MBT-IOP ($n = 44$) and received treatment as allocated for a maximum duration of 18 months. Inclusion and exclusion criteria, patient characteristics and randomization procedures have been described in detail by Laurensen et al. (2014). All 114 patients included in the 18-month treatment outcome study (Figure 1) were approached again for long-term follow-up assessments at 24, 30, and 36 months after start of treatment. Written informed consent was obtained from all patients.

[Insert Figure 1]

Outcome measures

The primary outcome measure was symptom severity as assessed by the Global Severity Index of the Brief Symptom Inventory (BSI; De Beurs, 2011; Derogatis, 1975). Secondary outcomes included: severity of borderline symptoms as measured with the Personality Assessment Inventory (PAI-BOR; Distel, De Moor, & Boomsma, 2009); personality functioning as assessed by the Severity Indices of Personality Problems-Short Form (SIPP; Verheul, 2006; Verheul et al., 2008); interpersonal problems as assessed by the Inventory of Interpersonal Problems (IIP; Horowitz, Alden, Wiggins, & Pincus, 2000; Zevalkink et al., 2012); quality of life as assessed by the EQ-5D-3L (Brooks, Rabin, & de Charro, 2003); and frequency of suicide attempts and self-harm as assessed by the Suicide and Self-Harm Inventory (SSH; Bateman & Fonagy, 2004). The intensity of care consumption during the follow-up period was based on registered number of minutes of treatment at the research sites. To exclude terminative or administrative visits related to the previous intensive treatment phase from counting towards additional care consumption, a minimum of 180 registered treatment minutes after termination of the intensive treatment was assumed to be relevant care consumption.

Treatment interventions

A detailed description of MBT-DH and MBT-IOP is provided elsewhere (Smits et al., 2019). Briefly, treatment components and features in MBT-DH and MBT-IOP are very similar, with weekly individual sessions in both programs, but the intensity of group therapy differs markedly. MBT-IOP involves two group therapy sessions per week, while MBT-DH entails a day hospital program 5 days per week, with nine group therapy sessions per week. Treatment adherence to the MBT model in the intensive treatment phase was rated as adequate by three independent raters and did not differ between MBT-DH and MBT-IOP. Overall dropout rate during the intensive treatment phase was 12%, $n = 14$, comprising one-sided termination of treatment by the patient ($n = 12$) or push-out by staff ($n = 2$), with no differences between the groups ($n = 5$, 11% for MBT-IOP and $n = 9$, 13% for MBT-DH; $\chi^2(1) = 0.056$, $p =$

0.813). The duration of the main treatment phase was somewhat shorter in MBT-DH (mean = 14.3 months, $SD = 4.2$) compared with MBT-IOP (mean = 15.9 months, $SD = 3.1$; $t(109) = 2.223$, $p = 0.028$). There were no differences between the groups in the proportion of patients using medication at baseline ($\chi^2(1) = 0.001$, $p = 0.972$), at 18-month follow-up ($\chi^2(1) = 2.276$, $p = 0.131$), or 36-month follow-up ($\chi^2(1) = 0.185$, $p = 0.667$).

After termination of the intensive treatment phase, patients were offered individually tailored follow-up care generally consisting of individual booster sessions of mentalizing therapy, crisis management or psychiatric consultation, as detailed in the MBT manual (Bateman, Bales, & Hutsebaut, 2014). Overall, 76.3% ($n = 87$) of patients received such care at their initial treatment site after ending the intensive treatment phase, with a trend for patients in MBT-DH to more often receive follow-up care ($n = 58$, 82.9%) than patients in MBT-IOP ($n = 29$, 65.9%; $\chi^2(1) = 3.41$, $p = 0.065$). Additionally, for patients who received follow-up care, the intensity of this care was significantly higher in the MBT-DH group (median = 82 hours) than in the MBT-IOP group (median = 51 hours; Mann-Whitney $U = 1049.00$, $z = -2.04$, $p = 0.041$), with large individual variability in both groups (range 3–350 hours).

Statistical analyses

Differences in demographic and clinical features at baseline were investigated using two-tailed chi-square tests and independent sample t -tests, as appropriate. Mann-Whitney U tests were used to examine differences in follow-up care consumption.

Multilevel modelling was used to examine treatment outcomes over time to best accommodate the missing data that are an inevitable feature of longitudinal follow-up and to deal with the dependency of repeated measures within subjects over time; this was conducted using the XTMIXED procedure of Stata Statistical Software Release 12. All outcome analyses were based on the intention-to-treat principle. Time points were coded -6 , -5 , -4 , -3 , -2 , -1 , and 0 , implying that regression coefficients involving time measured the rate of change from baseline to 36-month follow-up and regression intercepts referenced group differences at the last time point. SSHI scores were log-transformed because they were positively skewed. Maximum likelihood was used to assess whether random or fixed slopes should be assumed in

models for each outcome variable. Subsequently, quadratic and cubic time variables were added to the model if likelihood ratio tests showed a significant improvement in fit.

In interpreting the multilevel models; a significant main effect of time (polynomials) represents significant differences over the course of time independent of group; intercepts reference potential significant group differences at 36 months after start of treatment. Finally, significant interaction effects between group and (any polynomial of) time represent significant differences in the slopes of MBT-DH compared with MBT-IOP, and hence significant differences in the trajectories of change over time.

Reported estimates and Cohen's *d* effect sizes (Cohen, 1988) were based on predicted values. The *a priori* set superiority margin was $d \geq 0.50$, as this represents a clinically meaningful difference in the treatment of BPD (Laurensen et al., 2018; Smits et al., 2019). Furthermore, clinically significant change was calculated for symptom severity and borderline symptomatology. Patients were classified into the following categories: (1) recovered (i.e., statistically reliable change and movement from a dysfunctional range to a functional range), (2) improved (i.e., statistically reliable change in the direction indicative of improvement without crossing the cut-off); (3) unchanged (i.e., no statistically reliable change), (4) deteriorated (i.e., statistically reliable change in the opposite direction to that indicative of improvement), and (5) relapsed (i.e., statistically reliable change in the opposite direction to that indicative of improvement and movement from a functional to a dysfunctional range). To deal with missing data, recovery scores were based on predicted estimates of multilevel modeling. Following Jacobson and Truax (1991), reliable change (RC) from baseline to 18 months, baseline to 36 months, and 18 to 36 months after the start of treatment was computed based on the formula $RC = 1.96 \times \sqrt{2(SE)^2}$. To calculate the standard error of measurement (*SE*), Cronbach's alpha of 0.97 was used for the BSI (De Beurs, 2011) and 0.81 for the PAI-BOR (Distel et al., 2009) based on the following formula: $SE = \sqrt{(1-\alpha)}$. The cut-off score for movement from a dysfunctional to a normative range was based on the following formula: $(SD_{normal} \times M_{clinical}) + (SD_{clinical} \times M_{normal}) / (SD_{normal} \times SD_{clinical})$. Means and standard deviations for the clinical and non-clinical populations were based on values reported in the manual of the Dutch version of the BSI (De Beurs, 2011). For the PAI-BOR, means and standard deviations for the non-clinical population were based on norms of Distel et al. (2009), and respective values for the clinical population were based on our own sample. Owing to missing data, clinically significant change could not be

computed for all participants: it was computed for $n = 112$ participants on the BSI and $n = 111$ on the PAI-BOR (Table 1). Chi-square tests were used to determine whether MBT-DH and MBT-IOP differed in terms of recovery as calculated from baseline to 36 months after start of treatment and from 18 to 36 months.

Randomization, missing data and sensitivity analyses

There were no significant differences between the two treatment groups at baseline, except for a higher percentage of patients who reported self-harm at baseline in MBT-IOP (63%) compared with MBT-DH (42%), $\chi^2(1) = 3.96$, $p < 0.001$. The proportion of missing data increased somewhat with each follow-up assessment owing to difficulties in contacting patients who were no longer in treatment, and ranged from 53% to 58% depending on the outcome measure and follow-up time point. Overall, 61% of patients completed at least one of the three follow-up assessments on the primary outcome measure. There was no difference between MBT-IOP and MBT-DH in terms of the number of patients who completed at least one follow-up assessment: $n = 27$ (61%) for MBT-IOP and $n = 42$ (60%) for MBT-DH ($\chi^2(1) = 0.021$, $p = 0.885$). There were also no significant baseline differences between patients who completed a follow-up assessment and those who did not. This suggests that there was no selective study drop-out as a function of baseline severity.

Although multilevel modelling is quite robust in dealing with missing data, we re-ran all analyses using state-of-the-art data imputation procedures as described in Smits et al. (2019). These analyses yielded similar results, hence only results on the non-imputed data set are reported. Results on the imputed data are available upon request from the first author.

We also performed completer analyses, excluding treatment drop-outs using the pre-defined criterion of one-sided termination of treatment by the patient or push-out by staff ($n = 14$). These analyses yielded similar results as the intention-to-treat analyses for all outcome measures (see Supplemental Table S1).

Results

Primary outcome

Improvement over time in terms of symptom severity between baseline and 36-month follow-up was statistically significant, representing large effect sizes in both MBT-IOP ($d = 1.00$) and MBT-DH ($d = 1.12$). There was no significant difference between the two groups at 36 months after the start of treatment ($\beta = -0.20$, 95% CI $[-0.62, 0.22]$, $z = -0.93$, $p = 0.350$), nor did the rate of change differ between the two groups ($\beta = -0.02$, 95% CI $[-0.09, 0.05]$, $z = -0.51$, $p = 0.610$). The between-group effect size of $d = 0.26$ also indicated that MBT-DH was not superior to MBT-IOP in terms of improvement in symptom severity based on the *a priori* specified clinically meaningful Cohen's $d \geq 0.5$ margin at 36 months after the start of treatment (Table 1).

[Insert Table 1]

[Insert Figure 2]

Secondary outcomes

Likewise, no significant differences were observed for any of the secondary outcome measures between the two groups at 36 months after the start of treatment (Table 1). Between-group effect sizes were small, with the exception of quality of life as measured with the EQ-5D-3L. Patients in MBT-DH showed higher scores than patients in MBT-IOP on this measure ($d = 0.55$), although this difference did not reach significance in the multilevel model. Yet, significant interaction effects between time(polynomials) and group indicated significant differences in terms of a differential rate of change between patients in MBT-DH and MBT-IOP on several other secondary outcome variables, suggesting differences in trajectories of improvement between the treatment groups. This was the case for several domains of personality functioning as assessed by the SIPP (identity integration, self-control, and relational capacities), borderline symptomatology as assessed by the PAI-BOR, and interpersonal problems as assessed by the IIP (Table 1). For all these secondary outcome measures, patients in MBT-DH maintained treatment gains made in the first 18 months of treatment or showed small additional gains, as represented by small within-group effect sizes from 18- to 36-month follow-up (range $d = -0.08$ to

0.24). Patients in MBT-IOP, by contrast, showed clear continued improvement on these outcome measures during the follow-up period, as is also shown by medium within-group effect sizes between 18 and 36 months (range $d = 0.29$ to 0.56). For relational functioning (SIPP relational capacities and IIP interpersonal problems), patients in MBT-IOP even showed a larger improvement during the follow-up period compared with the active treatment phase, whereas patients in MBT-DH did not show continued improvement from 18 to 36 months after the start of treatment, in terms of effect sizes.

Clinically significant change

In terms of symptom severity as assessed by the BSI, for the two treatment groups combined, 83% of patients ($n = 93$) were categorized as improved at 36-month follow-up, over a quarter of whom ($n = 25$, 26.9%) were classified as recovered. A small proportion of patients ($n = 16$, 14.3%) did not show reliable change and three patients (2.7%) deteriorated over the course of the 36-month follow-up period. There were no differences between MBT-IOP and MBT-DH at 18 months ($\chi^2(2, n = 112) = 1.75, p = 0.418$) or 36 months ($\chi^2(3, n = 112) = 5.43, p = 0.143$) after the start of treatment in terms of clinically significant change. Looking specifically at the follow-up period from 18 to 36 months, 95 patients maintained their initial improvement (84.8% unchanged), 12 patients continued to improve with additional reliable change (10.7%), and four patients deteriorated (3.6%). There were no significant differences in clinically significant change from 18 to 36 months between MBT-DH and MBT-IOP ($\chi^2(3, n = 112) = 1.6816, p = 0.641$).

In terms of borderline symptomatology (PAI-BOR), nearly all patients ($n = 108$, 97.3%) improved at 36-month follow-up, more than 50% of whom ($n = 60$, 55.6%) could be classified as recovered. Three patients who had shown improvement at 18 months after the start of treatment showed deterioration at 36-month follow-up. There were no differences between MBT-IOP and MBT-DH at 18 months ($\chi^2(1, n = 111) = 1.33, p = .248$) or 36 months ($\chi^2(3, n = 111) = 3.18, p = 0.204$) after the start of treatment in terms of clinically significant change on the PAI-BOR. Looking specifically at the follow-up period from 18 to 36 months, 89 patients showed additional reliable change (80.1% improved or recovered), 11 patients (9.9% unchanged) maintained their initial improvement, and 11 patients (9.9%) showed deterioration, among whom one patient relapsed. Over the follow-up period from 18 to 36 months, MBT-DH and MBT-IOP

differed significantly in terms of clinically significant change on borderline symptomatology ($\chi^2(4, n = 111) = 15.70, p = 0.003$), with more patients in MBT-IOP showing continued improvement or recovery during follow-up ($n = 39, 92.9\%$ improved or recovered in MBT-IOP versus $n = 50, 72.4\%$ in MBT-DH; $\chi^2(1) = 7.7021, p = 0.006$). This observation is consistent with findings from the multilevel models indicating that patients in MBT-IOP showed greater continued improvement during the follow-up period than patients in MBT-DH.

[Insert Table 2]

Discussion

Patients in both MBT-DH and MBT-IOP maintained the substantial improvements made during the intensive treatment phase, and continued to improve during long-term follow up from 18 to 36 months after the start of treatment. In terms of clinically significant change on the primary outcome measure of symptom severity, 83% of patients improved over this period, over a quarter of whom (26.9%) met criteria for recovery—that is, they moved from a dysfunctional to a functional range on symptom severity. For borderline symptomatology specifically, 97% of patients improved, and over half of these patients could be classified as recovered. Contrary to our hypothesis, but consistent with the outcome results at 18 months, MBT-DH did not show superiority in terms of reduction of symptom severity at 36 months after the start of treatment. Similarly, there were no apparent differences between the two groups in terms of clinically significant change over the course of 36 months for both symptom severity and borderline symptomatology. The trend towards superiority of MBT-DH on several secondary outcome measures, specifically those in the domain of relational functioning, observed at 18 months after the start of treatment, was no longer evident at 36 months.

Furthermore, trajectories of change during follow-up were notably different for patients in MBT-DH and MBT-IOP. Patients in MBT-DH tended to show the largest improvement during the intensive treatment phase, and these gains were largely maintained or slightly increased during the follow-up period. By contrast, whereas the rate of improvement in MBT-IOP was smaller than that in MBT-DH during the first 18 months, patients in MBT-IOP showed larger additional gains during follow-up. In other words, patients in MBT-IOP caught up with patients in MBT-DH during follow-up. This differential

trajectory of change was most apparent in the domain of relational functioning, with MBT-DH patients showing faster improvement initially that seemed to level off during follow-up, whereas MBT-IOP patients showed a larger improvement during follow-up. This finding is consistent with our earlier hypothesis (Smits et al., 2019), based upon work by Fonagy, Luyten, and Allison (2015), that the ‘safety net’ and scaffolding provided by the day hospital setting may result in earlier improvements in mentalizing and social learning because it may provide patients with greater opportunity to generalize therapeutic gains within a relatively safe social context. In contrast, patients in MBT-IOP are less likely to have access to a supportive environment, and they may have to face everyday problems related to relationships, work, and social activities sooner than patients in a day hospital program. Consequently, it may take more time for ‘virtuous cycles’ associated with increasing mentalizing and social learning to emerge for these patients. However, once these cycles are established, these patients seem to be able to catch up with patients in MBT-DH.

Importantly, the greater improvements in MBT-IOP during the follow-up period were not related to a higher intensity of follow-up care. In fact, patients in MBT-DH received follow-up care more often and at a higher intensity than those in MBT-IOP. Taken together, from a developmental psychopathology perspective, these results suggest that recovery in patients in MBT-IOP may follow a more ‘natural’ course, with patients being challenged from the start of treatment to generalize the capacities they acquire in terms of mentalizing, attachment, and social learning to their everyday life. Patients in MBT-DH, by contrast, may show faster improvement in these capacities because they mainly function within a relatively protected environment, but then may face the same problem when they are no longer in the day hospital setting—that is, how to generalize what they have learned in treatment to dealing with the everyday challenges of life. Further research is needed to investigate these hypotheses.

There are important limitations that should be kept in mind when interpreting the current results. First, the proportions of missing data at the 24-, 30-, and 36-month follow-up points were substantial, although this limitation is somewhat mitigated by the fact that imputation analysis showed comparable results. Moreover, there were no differences in the percentages of missing data between the two treatment groups, and completer analyses showed similar results. Yet, it cannot be ruled out that missingness might be related to factors related to treatment outcome. Second, the superiority margin set

in this trial corresponded to a medium effect size. It could be argued that smaller between-group differences may be clinically relevant; further research in larger samples is needed to address this issue. Third, there was some evidence for differential care consumption between MBT-IOP and MBT-DH after the end of the main treatment phase. However, we do not have systematic data on the exact nature of the care that was provided at the treatment sites during follow-up, or on in-session adherence to the MBT model outside the main treatment phase, although the same therapists were involved in providing care during follow-up as in the main treatment phase. In addition, we did not use data on treatment-seeking outside the treatment site for this study. This issue will be addressed in more detail in future cost-effectiveness analyses. Finally, use of medication may have influenced treatment outcomes, although there were no differences between the two treatment arms in the proportion of patients using medication at baseline or 18 or 36 months after the start of treatment.

Despite these limitations, the results of this study suggest that MBT-IOP and MBT-DH are both valuable treatment options for patients with BPD. However, cost-effectiveness analyses are needed, given the large differences in intensity and cost of the two treatments, to further investigate whether MBT-IOP is more cost-effective than MBT-DH. As observations from an organizational perspective have shown that MBT-DH is more difficult to implement (Bales, Verheul, & Hutsebaut, 2017), without support for its cost-effectiveness, it will be hard to maintain MBT-DH as a viable treatment option. In terms of the optimization of use of health care resources, the implementation of less intensive treatment programs might enable more patients to be treated with similar resources, reducing the iatrogenic harm related to the choice of less ideal treatment options or long waiting lists. Still, however, MBT-DH may be more (cost-)effective in specific subgroups of BPD patients (e.g., those with more severe and chronic psychosocial problems), including those for whom MBT-IOP might even be contraindicated. Therefore, more research is needed to identify potential factors that might predict differential treatment outcomes. For example, for patients who are highly chaotic and fragmented, the stronger holding environment and structure provided by an intensive day hospital setting may be more effective than an outpatient setting. On the basis of the current results, we also hypothesize that an assessment of the social environment and opportunities for creating a sufficiently benign context in which treatment gains can be generalized might be important to consider when determining which treatment is likely to work best for individual patients.

Acknowledgements

We would like to thank all the research assistants for collecting the data. We are also very grateful to the patients who participated in this study.

Financial support

This research was supported in part by a grant from ZonMw (grant no. 171202012).

Conflicts of interest

PL and DLB have been involved in the development, training and dissemination of mentalization-based treatment. The other authors declare that they have no competing interests.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. This study was approved by the Medical Ethical Committee of Erasmus Medical Center, Rotterdam, The Netherlands (NL38571.078.12).

Author names and affiliations

Maaïke L. Smits, MSc (maaïke.l.smits@gmail.com), health care psychologist, De Viersprong and Viersprong Institute for Studies on Personality Disorders, Halsteren, The Netherlands; Dine J. Feenstra, PhD (dineke.feenstra@deviersprong.nl), clinical psychologist-psychotherapist, De Viersprong and principal investigator, Viersprong Institute for Studies on Personality Disorders, Halsteren, The Netherlands; Dawn L. Bales, PhD (dawn.bales@mbtnederland.nl), clinical psychologist-psychotherapist, De Viersprong and director, Expertcentre MBT-Nederland, Bergen op Zoom, The Netherlands; Matthijs Blankers, PhD (matthijs.blankers@arkin.nl), senior researcher, Department of Research, Arkin Mental Health Care, Amsterdam, The Netherlands; Trimbos Institute, The Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands and Amsterdam UMC, Location AMC, Department of Psychiatry, University of Amsterdam, Amsterdam, The Netherlands; Jack J. M. Dekker, PhD (Jack.Dekker@arkin.nl),

Running title: Long-term treatment outcome for two types of MBT

Professor Dr. at Department of Research, Arkin Mental Health Care, Amsterdam, The Netherlands and Department of Clinical Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; Zwaan Lucas, MSc (za.lucas@ Groningen.psyq.nl), clinical psychologist-psychotherapist, Lentis, Groningen, The Netherlands; Jan H. Kamphuis, PhD (j.h.kamphuis@uva.nl); Professor Dr. of Clinical Psychology and chair of Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands and senior researcher, Viersprong Institute for Studies on Personality Disorders, Halsteren, The Netherlands; Jan J. V. Busschbach, PhD (j.vanbusschbach@erasmusmc.nl), Professor Dr., Department of Psychiatry, Section Medical Psychology and Psychotherapy, Erasmus MC, Rotterdam, The Netherlands and senior researcher, Viersprong Institute for Studies on Personality Disorders, Halsteren, The Netherlands; Roel Verheul, PhD (rv2611@hotmail.com); CEO De Viersprong, Halsteren, The Netherlands; Patrick Luyten, PhD (patrick.luyten@kuleuven.be), Professor Dr., Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium; Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom and senior researcher, Viersprong Institute for Studies on Personality Disorders, Halsteren, The Netherlands.

Author contributions

PL, RV, JJMD, JJVB, JHK, DLB and DJF designed the study and directed the trial. DLB and PL were responsible for MBT quality aspects within the trial. DJF and MLS coordinated the trial and data collection. MLS, ZL and JJMD were responsible for trial implementation and data collection at the treatment sites. MLS, MB and PL performed the data analysis. MLS, DJF, JJVB, RV and PL interpreted the data and drafted the article. JHK, DLB, MB, JJMD, ZL and JHK revised the manuscript. All authors approved the final version to be submitted for publication.

References

Bales, D., Timman, R., Andrea, H., Busschbach, J. J., Verheul, R., & Kamphuis, J. H. (2014). Effectiveness of day hospital mentalization-based treatment for patients with severe borderline personality disorder: a matched control study. *Clinical Psychology & Psychotherapy*, 22(5), 409-417. doi:10.1002/cpp.1914

Running title: Long-term treatment outcome for two types of MBT

- Bales, D., van Beek, N., Smits, M. L., Willemsen, S., Busschbach, J. J. V., Verheul, R., & Andrea, H. (2012). Treatment outcome of 18-month day hospital mentalization-based treatment in patients with severe borderline personality disorder in the Netherlands. *Journal of Personality Disorders*, 26, 568-582. doi:10.1521/pedi.2012.26.4.568
- Bales, D. L., Verheul, R., & Hutsebaut, J. (2017). Barriers and facilitators to the implementation of mentalization-based treatment (MBT) for borderline personality disorder. *Personality and Mental Health*, 11(2), 118-131. doi:10.1002/pmh.1368
- Barnicot, K., & Crawford, M. (2019). Dialectical behaviour therapy v. mentalisation-based therapy for borderline personality disorder. *Psychological Medicine*, 49, 2060-2068. doi:S0033291718002878
- Bartak, A., Andrea, H., Spreeuwenberg, M. D., Ziegler, U. M., Dekker, J., Rossum, B. V., . . . Emmelkamp, P. M. G. (2011). Effectiveness of Outpatient, Day Hospital, and Inpatient Psychotherapeutic Treatment for Patients with Cluster B Personality Disorders. *Psychotherapy and Psychosomatics*, 80(1), 28-38. doi:10.1159/000321999
- Bateman, A., Bales, D., & Hutsebaut, J. (2014). A quality manual for MBT. Retrieved from <http://www.annafreud.org/media/1217/a-quality-manual-for-mbt-edited-april-23rd-2014-2.pdf>
- Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. *American Journal of Psychiatry*, 156(10), 1563-1569. doi:doi:10.1176/ajp.156.10.1563
- Bateman, A., & Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: an 18-month follow-up. *American Journal of Psychiatry*, 158(1), 36-42. doi:10.1176/appi.ajp.158.1.36
- Bateman, A., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: mentalization-based treatment*. Oxford: Oxford University Press.
- Bateman, A., & Fonagy, P. (2008). 8-year follow-up of patients treated for borderline personality disorder: mentalization-based treatment versus treatment as usual. *American Journal of Psychiatry*, 165(5), 631-638. doi:10.1176/appi.ajp.2007.07040636

Running title: Long-term treatment outcome for two types of MBT

- Bateman, A., & Fonagy, P. (2009). Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *American Journal of Psychiatry*, 166(12), 1355-1364. doi:10.1176/appi.ajp.2009.09040539
- Brooks, R., Rabin, R., & de Charro, F. (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. London: Kluwer Academic Publishers.
- Chiesa, M., Fonagy, P., & Holmes, J. (2006). Six-Year Follow-Up of Three Treatment Programs to Personality Disorder. *Journal of Personality Disorders*, 20(5), 493-509. doi:10.1521/pedi.2006.20.5.493
- Chiesa, M., Fonagy, P., Holmes, J., & Drahorad, C. (2004). Residential Versus Community Treatment of Personality Disorders: A Comparative Study of Three Treatment Programs. *American Journal of Psychiatry*, 161(8), 1463-1470. doi:10.1176/appi.ajp.161.8.1463
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of Psychotherapies for Borderline Personality Disorder: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, 74(4), 319-328. doi:10.1001/jamapsychiatry.2016.4287
- De Beurs, E. (2011). *Brief Symptom Inventory 18, -BSI 18-, Handleiding herziene editie*. Leiden: PITS B.V.
- Derogatis, L. R. (1975). *Brief Symptom Inventory*. Baltimore, MD: Clinical Psychometric Research.
- Distel, M. A., De Moor, M. H. M., & Boomsma, D. I. (2009). Dutch translation of the Personality Assessment Inventory – borderline features scale (PAI-BOR): norms, factor structure and reliability. *Psychologie en Gezondheid*, 37, 38-46. doi:10.1007/BF03080362
- Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: a new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders*, 29(5), 575-609. doi:10.1521/pedi.2015.29.5.575
- Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. *American Psychologist*, 56(3), 218-226.
- Horn, E. K., Bartak, A., Meerman, A. M. M. A., Rossum, B. V., Ziegler, U. M., Soons, M., . . . Busschbach, J. J. (2016). Long-term effectiveness of psychotherapy in personality disorders. *Under review*.

Running title: Long-term treatment outcome for two types of MBT

Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *Inventory of Interpersonal Problems: Manual*. San Antonio, TX: The Psychological Corporation.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19. doi:10.1037//0022-006x.59.1.12

Jørgensen, C. R., Bøye, R., Andersen, D., Døssing Blaabjerg, A. H., Freund, D., Jordet, H., & Kjølbye, M. (2014). Eighteen months post-treatment naturalistic follow-up study of mentalization-based therapy and supportive group treatment of borderline personality disorder: clinical outcomes and functioning. *Nordic Psychology, 66*(4), 254-273. doi:10.1080/19012276.2014.963649

Jørgensen, C. R., Bøye, R., Boye, R., Jordet, H., Andersen, D., & Kjolbye, M. (2013). Outcome of mentalization-based and supportive psychotherapy in patients with borderline personality disorder: a randomized trial. *Acta Psychiatrica Scandinavica, 127*(4), 305-317. doi:10.1111/j.1600-0447.2012.01923.x

Juul, S., Lunn, S., Poulsen, S., Sørensen, P., Salimi, M., Jakobsen, J. C., . . . Simonsen, S. (2019). Short-term versus long-term mentalization-based therapy for outpatients with subthreshold or diagnosed borderline personality disorder: a protocol for a randomized clinical trial. *Trials, 20*(1), 196. doi:10.1186/s13063-019-3306-7

Kvarstein, E. H., Pedersen, G., Urnes, O., Hummelen, B., Wilberg, T., & Karterud, S. (2015). Changing from a traditional psychodynamic treatment programme to mentalization-based treatment for patients with borderline personality disorder-does it make a difference? *Psychology and Psychotherapy: Theory, Research and Practice, 88*(1), 71-86. doi:10.1111/papt.12036

Laurensen, E. M. P., Luyten, P., Kikkert, M. J., Westra, D., Peen, J., Soons, M., . . . Dekker, J. J. M. (2018). Day hospital mentalization-based treatment v. specialist treatment as usual in patients with borderline personality disorder: randomized controlled trial. *Psychological Medicine, 48*(15), 1-8. doi:10.1017/S0033291718000132

Laurensen, E. M. P., Smits, M. L., Bales, D. L., Feenstra, D. J., Eeren, H. V., Noom, M. J., . . . Verheul, R. (2014). Day hospital Mentalization-based treatment versus intensive outpatient Mentalization-

- based treatment for patients with severe borderline personality disorder: protocol of a multicentre randomized clinical trial. *BMC Psychiatry*, 14, 301. doi:10.1186/s12888-014-0301-0
- Luyten, P., & Fonagy, P. (2014). Mentalising in attachment contexts. In P. Holmes & S. Farnfield (Eds.), *The Routledge Handbook of Attachment: Theory*. London: Routledge.
- McMain, S. F., Chapman, A. L., Kuo, J. R., Guimond, T., Streiner, D. L., Dixon-Gordon, K. L., . . . Hoch, J. S. (2018). The effectiveness of 6 versus 12-months of dialectical behaviour therapy for borderline personality disorder: the feasibility of a shorter treatment and evaluating responses (FASTER) trial protocol. *BMC Psychiatry*, 18(1), 230. doi:10.1186/s12888-018-1802-z
- Smits, M. L., Feenstra, D. J., Eeren, H. V., Bales, D., Laurensen, E. M. P., Blankers, M., . . . Luyten, P. (2019). Day hospital versus intensive out-patient mentalisation-based treatment for borderline personality disorder: a multicentre randomised clinical trial. *British Journal of Psychiatry*, 216(2), 79-84. doi:10.1192/bjp.2019.9
- Verheul, R. (2006). SIPP-SF: Severity Indices for Personality Problems Short Form. Retrieved from (<http://www.deviersprong.nl/over-de-viersprongonderzoek/onderzoekslijn-diagnostiek/onderzoekslijn-assessment-enindicatiestelling/sipp-main-menu/>).
- Verheul, R., Andrea, H., Berghout, C. C., Dolan, C., Busschbach, J. J. V., van der Kroft, P. J. A., . . . Fonagy, P. (2008). Severity indices of personality problems (SIPP-118): Development, factor structure, reliability, and validity. *Psychological Assessment*, 20(1), 23-34. doi:10.1037/1040-3590.20.1.23
- Zevalkink, J., de Geus, J., Hoek, W., Berghout, C. C., Brouwer, D., Riksen-Walraven, J. M. A., . . . Katzko, M. (2012). *Manual IIP-64-NL: Inventory of Interpersonal Problems - Dutch version (version 4.2)* Internal report Nederlands Psychoanalytisch Instituut, Amsterdam.

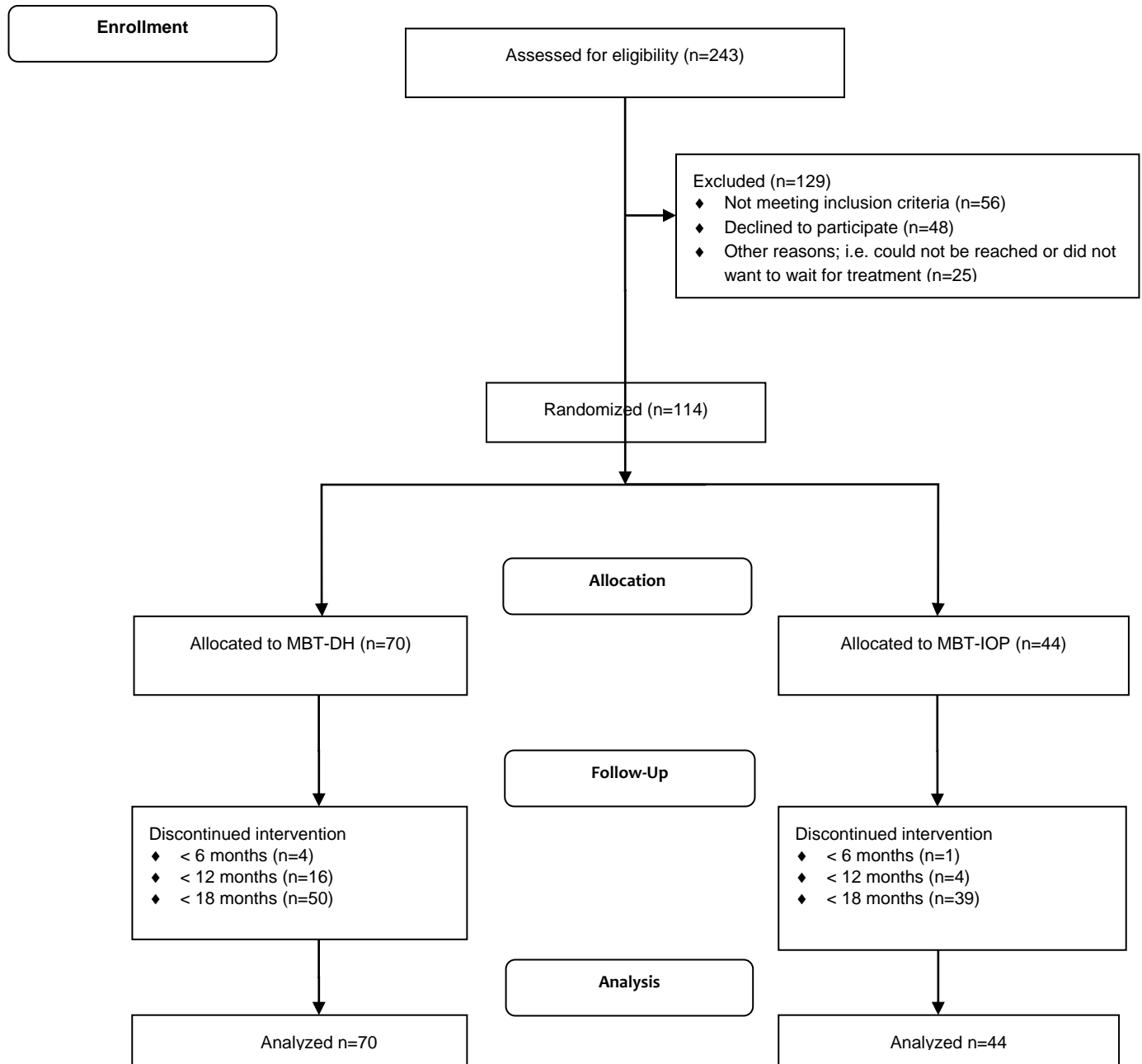


Figure 1. CONSORT Flow Diagram. MBT-DH = day hospital mentalization-based treatment
 MBT-IOP = intensive outpatient mentalization-based treatment.

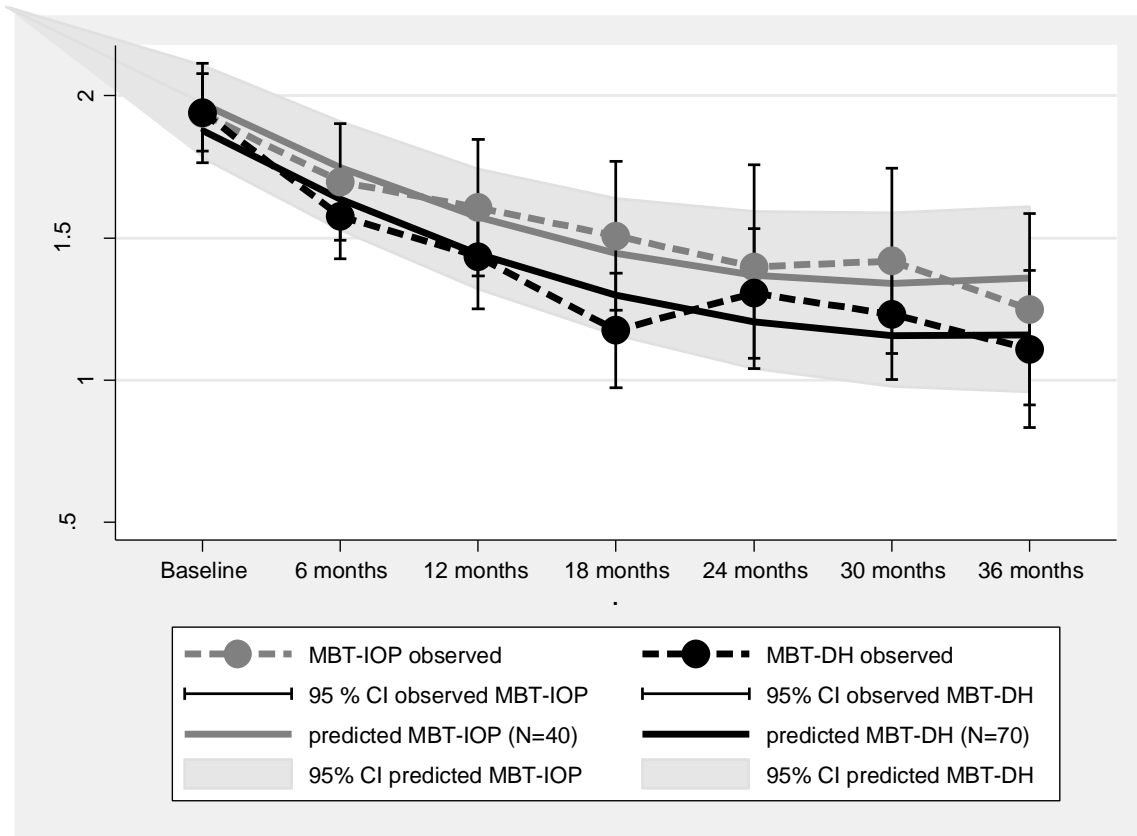


Figure 2. Observed and predicted Global Severity Index scores from baseline to 36 months after start of treatment for intensive outpatient mentalization-based treatment (MBT-IOP) ($n = 44$) and day hospital mentalization-based treatment (MBT-DH) ($n = 70$) with 95% confidence intervals.

Table 1. Predicted means, results from multilevel models and effect sizes 36 months after start of treatment on primary and secondary outcome measures for patients randomly assigned to intensive outpatient mentalization-based treatment (MBT-IOP) (*n* = 44) or day hospital mentalization-based treatment (MBT-DH) (*n* = 70)

	Symptom Severity (GSI)				Identity Integration (SIPP)				Self-Control (SIPP)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI
Baseline	1.97	1.84 to 2.11	1.88	1.78 to 1.97	1.88	1.76 to 1.99	1.86	1.79 to 1.94	2.05	1.95 to 2.15	2.16	2.08 to 2.23
18 months	1.45	1.25 to 1.64	1.30	1.16 to 1.44	2.34	2.13 to 2.54	2.64	2.78 to 2.50	2.57	2.40 to 2.74	2.86	2.75 to 2.98
36 months	1.36	1.11 to 1.61	1.16	0.96 to 1.36	2.71	2.43 to 2.99	2.59	2.36 to 2.81	2.88	2.64 to 3.13	3.00	2.82 to 3.19
Model: Wald X ²	69.26 (<i>df</i> = 4)				96.82 (<i>df</i> = 6)				104.31 (<i>df</i> = 5)			
Linear change	-0.04	-0.05 to -0.14		<i>p</i> = 0.352	0.25*	0.05 to 0.469		<i>p</i> = 0.016	0.09*	0.01 to 0.18		<i>p</i> = 0.032
Quadratic change	-0.02***	-0.09 to 0.05		<i>p</i> < 0.001	-0.07	-0.14 to 0.01		<i>p</i> = 0.069	-	-		-
Cubic change	-	-		-	0.01*	0.00 to 0.02		<i>p</i> = 0.037	0.00	-0.00 to 0.00		<i>p</i> = 0.138
Δ Linear change	-0.02	-0.09 to 0.05		<i>p</i> = 0.610	-0.27**	-0.44 to -0.10		<i>p</i> = 0.002	-	-		-
Δ Quadratic change	-	-		-	0.04**	0.02 to 0.07		<i>p</i> = 0.001	-	-		-
Δ Cubic change	-	-		-	-	-		-	0.00*	0.00 to 0.00		<i>p</i> = 0.048
Δ Group 36 months	-0.20	-0.62 to 0.22		<i>P</i> = 0.350	-0.13	-0.60 to 0.35		<i>p</i> = 0.606	0.12	-0.30 to 0.54		<i>p</i> = 0.576
Within-group ES	0.86/0.12		1.04/0.21		0.80/0.49		1.58/-0.08		1.07/0.49		1.61/0.24	
(Baseline-18 /18-36 months)	1.00		1.12		1.45		1.15		1.71		1.54	
(Baseline-36 months)	0.24 / 0.26		0.50 / 0.15		0.57/0.17							
Between-group ES	0.24 / 0.26		0.50 / 0.15		0.57/0.17							
(18/36 months)												
	Social Concordance (SIPP)				Responsibility (SIPP)				Relational Capacities (SIPP)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI
Baseline	2.63	2.53 to 2.73	2.75	2.68 to 2.83	2.49	2.37 to 2.61	2.67	2.60 to 2.75	2.29	2.15 to 2.42	2.16	2.08 to 2.24
18 months	2.81	2.65 to 2.97	3.00	2.89 to 3.11	2.74	2.55 to 2.92	2.86	2.74 to 2.97	2.41	2.21 to 2.61	2.58	2.46 to 2.70
36 months	2.99	2.78 to 3.20	3.25	3.11 to 3.39	2.98	2.76 to 3.21	3.04	2.88 to 3.20	2.54	2.29 to 2.79	2.58	2.41 to 2.75
Model: Wald X ²	41.06 (<i>df</i> = 3)				43.27 (<i>df</i> = 3)				37.30 (<i>df</i> = 5)			
Linear change	0.06**	0.02 to 0.10		<i>p</i> = 0.001	0.08***	0.05 to 0.12		<i>p</i> < 0.001	0.08	-0.02 to 0.19		<i>p</i> = 0.135
Quadratic change	-	-		-	-	-		-	-0.00	-0.02 to 0.01		<i>p</i> = 0.570
Δ Linear change	0.02	-0.03 to 0.07		<i>p</i> = 0.362	-0.02	-0.07 to 0.02		<i>p</i> = 0.337	-0.15*	0.29 to -0.01		<i>p</i> = 0.033

Running title: Long-term treatment outcome for two types of MBT

Δ Quadratic change							0.03**	0.01 to 0.05	p = 0.005			
Δ Group 36 months	0.26	-0.05 to 0.56	p = 0.101	0.06	-0.26 to 0.37	p = 0.728	-0.04	-0.18 to 0.86	p = 0.859			
Within-group ES (Baseline–18/18-36 months)	0.38/0.32		0.59/0.53		0.45/0.39		0.15/0.62		0.21/0.29		0.91/0.00	
(Baseline–36 months)	0.77		1.09		0.90		0.75		0.55		0.79	
Between-group ES (18/36 months)	0.40/0.45				0.00/0.09				0.29/0.05			

	Borderline Symptomatology (PAI-BOR)				Quality of Life (EQ-5D-3L)				Interpersonal Problems (IIP)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	M	95% CI	M	95% CI	M	95% CI	M	95% CI	M	95% CI	M	95% CI
Baseline	49.61	48.14 to 51.09	47.52	46.32 to 48.73	0.44	0.41 to 0.47	0.48	0.45 to 0.51	108.44	103.74 to 113.15	108.95	105.48 to 112.42
18 months	40.07	38.15 to 43.24	33.91	33.35 to 37.23	0.60	0.55 to 0.64	0.66	0.63 to 0.69	105.293	96.27 to 114.31	89.75	83.07 to 96.42
36 months	35.42	31.77 to 39.07	32.95	29.92 to 35.97	0.61	0.56 to 0.66	0.69	0.65 to 0.73	90.94	77.58 to 104.30	87.74	77.05 to 98.43
Model: Wald X ²	116.44 (df = 5)				44.68 (df = 4)				19.44 (df = 5)			
Linear change	-1.56*	-3.01 to -0.10	p = 0.036		-0.00	-0.03 to 0.03	p = 0.816		-6.65	-14.47 to 1.17	p = 0.096	
Quadratic change	-	-	-		-	-	-		0.62	-0.51 to 1.76	P = 0.283	
Cubic change	-0.02	-0.05 to -0.01	p = 0.160		0.00**	0.00 to 0.00	p = 0.007		-	-	-	
Δ Linear change	-1.32	-0.55 to 3.20	p = 0.165		0.01	-0.02 to 0.03	p = 0.606		-8.85	-1.31 to 19.00	p = 0.088	
Δ Quadratic change	-	-	-		-	-	-		-1.58*	-3.05 to 0.10	p = 0.036	
Δ Cubic change	-0.04	-0.08 to 0.00	p = 0.061		-	-	-		-	-	-	
Δ Group 36 months	-2.47	-9.26 to 4.31	p = 0.475		0.08	-0.04 to 0.20	p = 0.199		-3.20	-28.28 to 21.87	p = 0.802	
Within-group ES (Baseline–18/18-36 months)	1.23/0.56		1.68/0.24		1.10/0.09		1.34/0.24		0.12/0.40		0.81/0.00	
(Baseline--36 months)	1.95		1.63		1.20		1.57		0.67		0.70	
Between-group ES	0.66/0.23				0.43/0.56				0.53/0.08			

	Frequency of self-harm (SSHI)				Frequency of suicide attempts (SSHI)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	Log M	95% CI	Log M	95% CI	Log M	95% CI	Log M	95% CI
Baseline	1.63	1.38 to 1.87	0.89	0.70 to 1.08	0.17	0.09 to 0.24	0.22	0.11 to 0.20
18 months	0.33	0.09 to 0.56	0.42	0.24 to 0.59	0.19	0.02 to 0.15	0.13	0.05 to 0.12

Running title: Long-term treatment outcome for two types of MBT

36 months	0.27	0.11 to 0.42	0.41	0.290 to 0.53	0.11	0.00 to 0.07	0.04	0.00 to 0.02
Model: Wald X ²	83.04 (df = 7)			14.06 (df = 3)				
Linear change	-0.23	-0.69 to 0.22	<i>p</i> = 0.315	-0.03**	-0.05 to 0.01	<i>p</i> = 0.012		
Quadratic change	0.14	-0.04 to 0.32	<i>p</i> = 0.127	–	–	–		
Cubic change	-0.02*	-0.04 to -0.00	<i>p</i> = 0.018	–	–	–		
Δ Linear change	0.47	-0.12 to 1.05	<i>p</i> = 0.119	0.00	-0.02 to 0.03	<i>p</i> = 0.830		
Δ Quadratic change	-0.25*	-0.48 to -0.01	<i>p</i> = 0.039	–	–	–		
Δ Cubic change	0.03*	0.01 to 0.06	<i>p</i> = 0.012	–	–	–		
Δ Group 36 months	0.14	-0.33 to 0.62	<i>p</i> = 0.558	0.00	-0.09 to 0.09	<i>p</i> = 0.919		
Within-group ES	1.45/0.10		0.56/0.01		0.31/0.57		0.39/0.72	
(Baseline–18/18-36 months)								
(Baseline–36 months)	1.51		0.64		0.62		0.90	
Between-group ES			0.13/0.31		0.02/0.09			
(18/36 months)								

Note. MBT-IOP = intensive outpatient mentalization-based treatment. MBT-DH = day hospital mentalization-based treatment. GSI = Global Severity Index of the Brief Symptom Inventory. SIPP = Severity Indices of Personality Problems. PAI-BOR = Personality Assessment Inventory – Borderline Personality Disorder section. EQ-5D-3L = EuroQoL-5D-3L. IIP = Inventory of Interpersonal Problems. SSHI = Suicide and Self-Harm Inventory. linear/quadratic/cubic change represent main effects of (polynomials of) time and indicate changes over time independent of group; Δ linear/quadratic/cubic change = interaction term of time (polynomial) with group indicating differential change of MBT-DH compared to MBT-IOP over time. Δ group 36 months = difference of MBT-DH compared to MBT-IOP at 36 months. ES = Effect size by means of Cohen's *d*. **P* < 0.05 ** *P* < 0.01 *** *P* < 0.001.

Table 2. Distribution over categories of recovery based on symptom severity (GSI) and borderline symptomatology (PAI-BOR) for patients randomly assigned to intensive outpatient mentalization-based treatment (MBT-IOP) or day hospital mentalization-based treatment (MBT-DH)

GSI	Baseline–18 months			Baseline–36 months			18–36 months		
	MBT-IOP (n = 43)	MBT-DH (n = 69)	Total (n = 112)	MBT-IOP (n = 43)	MBT-DH (n = 69)	Total (n = 112)	MBT-IOP (n = 43)	MBT-DH (n = 69)	Total (n = 112)
Recovered, n (%)	5 (11.6)	11 (15.9)	16 (14.3)	8 (18.6)	17 (24.6)	25 (22.3)	2 (4.7)	3 (4.3)	5 (4.5)
Improved, n (%)	30 (69.8)	51 (73.9)	81 (72.3)	26 (60.5)	42 (60.9)	68 (60.7)	1 (2.3)	6 (8.7)	7 (6.3)
Unchanged, n (%)	8 (18.6)	7 (10.1)	15 (13.4)	9 (20.9)	7 (10.1)	16 (14.3)	39 (90.7)	56 (82.2)	95 (84.8)
Deteriorated, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	3 (4.3)	3 (2.7)	1 (2.3)	3 (4.3)	4 (3.6)
Relapsed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
PAI-BOR	(n = 42)	(n = 69)	(n = 111)	(n = 42)	(n = 69)	(n = 111)	(n = 42)	(n = 69)	(n = 111)
Recovered, n (%)	19 (45.2)	39 (56.5)	59 (53.2)	22 (52.4)	38 (55.1)	60 (54.1)	2 (4.8)	0 (0)	2 (1.8)
Improved, n (%)	23 (54.8)	30 (43.5)	53 (47.7)	20 (47.6)	28 (40.6)	48 (43.2)	37 (88.1)	50 (72.5)	87 (78.4)
Unchanged, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (7.1)	8 (11.6)	11 (9.9)
Deteriorated, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	3 (4.3)	3 (2.7)	0 (0)	10 (14.5)	10 (9.0)
Relapsed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1.4)	1 (0.9)

Note. MBT-IOP = intensive outpatient mentalization-based treatment. MBT-DH = day hospital mentalization-based treatment. GSI = Global Severity Index of the Brief Symptom Inventory. PAI-BOR = Personality Assessment Inventory – Borderline Personality Disorder section. Recovery scores are calculated between baseline and 18 months and baseline and 36 months after start of treatment. Recovered, statistically reliable change and movement from a dysfunctional range to a functional range. Improved, statistically reliable change in the direction indicative of improvement without crossing cut-off. Unchanged, no statistically reliable change. Deteriorated, statistically reliable change in the opposite direction of that indicative of improvement. Relapsed, statistically reliable change in the opposite direction of that indicative of improvement and movement from a functional to a dysfunctional range.

Supplemental Table S1. Predicted means, results from multilevel models and effect sizes 36 months after start of treatment on primary and secondary outcome measures for treatment completers: Outpatient Mentalization-Based Treatment (MBT-IOP) (*n* =39) or Day Hospital Mentalization-Based Treatment (MBT-DH) (*n* = 61)

	Symptom Severity (GSI)				Identity Integration (SIPP)				Self-Control (SIPP)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI
Baseline	2.01	1.86 to 2.16	1.93	1.83 to 2.03	1.88	1.75 to 2.00	1.83	1.75 to 1.90	2.03	1.92 to 2.14	2.14	2.07 to 2.21
18 months	1.47	1.27 to 1.68	1.33	1.18 to 1.48	2.32	2.11 to 2.53	2.64	2.50 to 2.78	2.54	2.37 to 2.72	2.87	2.75 to 2.99
36 months	1.39	1.12 to 1.66	1.18	0.97 to 1.39	2.70	2.39 to 3.00	2.58	2.35 to 2.81	2.86	2.60 to 3.12	3.02	2.82 to 3.22
Model: Wald X ²	71.75 (<i>df</i> = 4)				100.02 (<i>df</i> = 6)				106.55 (<i>df</i> = 5)			
Linear change	0.05	-0.04 to 0.14	<i>p</i> = 0.312		0.25*	0.05 to 0.452	<i>p</i> = 0.016		0.09*	0.01 to 0.18	<i>p</i> = 0.028	
Quadratic change	-0.03***	-0.04 to 0.01	<i>p</i> < 0.001		-0.06	-0.13 to 0.01	<i>p</i> = 0.078		-	-	-	
Cubic change	-	-	-		0.01*	0.00 to 0.01	<i>p</i> = 0.047		0.00	-0.00 to 0.00	<i>p</i> = 0.159	
Δ Linear change	-0.02	-0.09 to 0.05	<i>p</i> = 0.528		-0.28**	-0.45 to -0.11	<i>p</i> = 0.001		-0.08	-0.19 to 0.03	<i>p</i> = 0.169	
Δ Quadratic change	-	-	-		0.04***	0.02 to 0.07	<i>p</i> < 0.001		-	-	-	
Δ Cubic change	-	-	-		-	-	-		0.00*	0.00 to 0.00	<i>p</i> = 0.028	
Δ Group 36 months	-0.21	-0.64 to 0.23	<i>p</i> = 0.347		-0.12	-0.60 to 0.36	<i>p</i> = 0.630		0.16	-0.27 to 0.58	<i>P</i> = 0.470	
Within-group ES (Baseline–18/18–36 months)	0.83/0.11		1.07/0.21		0.74/0.47		1.67/0.09		1.02/0.47		1.69/0.25	
(Baseline–36 months)	0.96		1.14		1.36		1.17		1.65		1.56	
Between-group ES (18/36 months)	0.22/0.26				0.50/0.14				0.61/0.21			
	Social Concordance (SIPP)				Responsibility (SIPP)				Relational Capacities (SIPP)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI
Baseline	2.60	2.48 to 2.71	2.75	2.67 to 2.83	2.49	2.35 to 2.62	2.68	2.60 to 2.76	2.29	2.10 to 2.40	2.16	2.08 to 2.24
18 months	2.79	2.62 to 2.96	3.00	2.90 to 3.11	2.73	2.55 to 2.91	2.74	2.75 to 2.98	2.41	2.19 to 2.60	2.58	2.46 to 2.70
36 months	2.98	2.76 to 3.20	3.25	3.11 to 3.39	2.97	2.75 to 3.20	3.05	2.90 to 3.21	2.54	2.32 to 2.86	2.58	2.41 to 2.75
Model: Wald X ²	42.42 (<i>df</i> = 3)				42.04 (<i>df</i> = 3)				42.59 (<i>df</i> = 5)			
Linear change	0.06**	0.03 to 0.10	<i>p</i> = 0.001		0.08***	0.05 to 0.12	<i>p</i> < 0.001		0.08	-0.03 to 0.18	<i>p</i> = 0.151	
Quadratic change	-	-	-		-	-	-		-0.00	-0.02 to 0.01	<i>p</i> = 0.674	
Δ Linear change	0.02	-0.03 to 0.07	<i>p</i> = 0.480		-0.02	-0.06 to 0.03	<i>p</i> = 0.409		-0.15*	-0.28 to -0.01	<i>p</i> = 0.031	

Running title: Long-term treatment outcome for two types of MBT

Δ Quadratic change	–	–	–	–	–	–	0.03**	0.01 to 0.05	$p = 0.005$			
Δ Group 36 months	0.26	–0.05 to 0.58	$p = 0.098$	0.08	–0.23 to 0.40	$p = 0.608$	–0.01	–0.41 to 0.38	$p = 0.946$			
Within-group ES (Baseline–18/18–36 months)	0.40/0.32		0.62/0.53		0.44/0.38		0.15/0.63		0.23/0.27		1.00/0.01	
(Baseline–36 months)	0.79		1.11		0.87		0.77		0.55		0.86	
Between-group ES (18/36 months)	0.44/0.45				0.03/0.13				0.32/0.02			

	Borderline Symptomatology (PAI-BOR)				Quality of Life (EQ-5D-3L)				Interpersonal Problems (IIP)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI	<i>M</i>	95% CI
Baseline	49.53	47.94 to 50.42	48.47	47.30 to 49.64	0.43	0.39 to 0.47	0.47	0.44 to 0.50	108.51	103.36 to 113.66	112.11	108.49 to 115.74
18 months	41.14	38.50 to 43.78	35.63	33.67 to 37.59	0.59	0.55 to 0.64	0.66	0.62 to 0.69	103.45	94.10 to 112.81	91.07	84.28 to 97.86
36 months	35.68	31.83 to 39.53	33.37	30.29 to 36.45	0.61	0.56 to 0.66	0.69	0.65 to 0.73	89.77	77.85 to 103.70	89.80	79.89 to 100.72
Model: Wald X^2	116.37 ($df = 5$)				45.01 ($df = 4$)				20.93 ($df = 5$)			
Linear change	–1.66*	–3.11 to –0.21	$p = 0.025$	–0.00	–0.03 to 0.03	$p = 0.804$	–6.00	–13.85 to 1.85	$p = 0.134$			
Quadratic change	–	–	–	–	–	–	0.48	–0.67 to 1.63	$p = 0.414$			
Cubic change	–0.02	–0.05 to 0.01	$p = 0.268$	0.00**	0.00 to 0.00	$p = 0.007$	–	–	–			
Δ Linear change	1.49	–0.37 to 3.36	$p = 0.117$	0.01	–0.02 to 0.03	$p = 0.573$	–8.87	–1.33 to 19.07	$p = 0.088$			
Δ Quadratic change	–	–	–	–	–	–	–1.58*	–3.07 to –0.08	$p = 0.039*$			
Δ Cubic change	–0.05*	–0.09 to –0.01	$p = 0.024$	–	–	–	–	–	–			
Δ Group 36 months	–2.31	–9.16 to 4.54	$p = 0.508$	0.08	–0.04 to 0.20	$p = 0.175$	0.03	–25.30 to 25.37	$p = 0.998$			
Within-group ES (Baseline–18/18–36 months)	1.13/0.55		1.86/0.23		1.11/0.09		1.36/0.25		0.19/0.36		0.90/0.00	
(Baseline–36 months)	1.86		1.70		1.20		1.61		0.70		0.73	
Between-group ES	0.67/0.21				0.44/0.58				0.42/0.00			

	Frequency of self-harm (SSHI)				Frequency of suicide attempts (SSHI)			
	MBT-IOP		MBT-DH		MBT-IOP		MBT-DH	
	Log <i>M</i>	95% CI	Log <i>M</i>	95% CI	Log <i>M</i>	95% CI	Log <i>M</i>	95% CI
Baseline	1.66	1.41 to 1.92	0.79	0.60 to 0.98	0.19	0.10 to 0.27	0.14	0.09 to 0.18
18 months	0.33	0.10 to 0.57	0.35	0.18 to 0.52	0.10	0.03 to 0.17	0.07	0.04 to 0.11

Running title: Long-term treatment outcome for two types of MBT

36 months	0.27	0.11 to 0.43	0.37	0.25 to 0.49	0.01	-0.01 to 0.02	0.01	-0.00 to 0.02
Model: Wald X^2	78.11 ($df = 7$)			12.39 ($df = 3$)				
Linear change	-0.25	-0.71 to 0.20	$p = 0.278$	-0.03**	-0.05 to 0.01	$p = 0.008$		
Quadratic change	0.15	-0.03 to 0.34	$p = 0.107$	-	-	-		
Cubic change	-0.02*	-0.04 to -0.00	$p = 0.014$	-	-	-		
Δ Linear change	0.47	-0.12 to 1.06	$p = 0.118$	0.01	-0.02 to 0.04	$p = 0.520$		
Δ Quadratic change	-0.24*	-0.48 to -0.01	$p = 0.043$	-	-	-		
Δ Cubic change	0.03*	0.01 to 0.06	$p = 0.013$	-	-	-		
Δ Group 36 months	0.10	-0.38 to 0.58	$p = 0.678$	0.00	-0.09 to 0.10	$p = 0.933$		
Within-group ES (Baseline-18/18-36 months)	1.47/0.10		0.56/0.04		0.33/0.61		0.34/0.63	
(Baseline-36 months)	1.54		0.59		0.65		0.77	
Between-group ES (18/36 months)	0.02/0.21				0.14/0.07			

Note. MBT-IOP = intensive outpatient mentalization-based treatment. MBT-DH = day hospital mentalization-based treatment. GSI = Global Severity Index of the Brief Symptom Inventory. SIPP = Severity Indices of Personality Problems. PAI-BOR = Personality Assessment Inventory – Borderline Personality Disorder section. EQ-5D-3L = EuroQol-5D-3L. IIP = Inventory of Interpersonal Problems. SSHI = Suicide and Self-Harm Inventory. linear/quadratic/cubic change represent main effects of (polynomials of) time and indicate changes over time independent of group; Δ linear/quadratic/cubic change = interaction term of time (polynomial) with group indicating differential change of MBT-DH compared to MBT-IOP over time. Δ group 36 months = difference of MBT-DH compared to MBT-IOP at 36 months. ES = Effect size by means of Cohen's d . * $P < 0.05$ ** $P < 0.01$ *** $P < 0.001$.