STATISTICAL ANALYSIS PLAN

Evaluation of Tavistock Relationships MBT-PP to improve child outcomes by reducing inter-parental conflict: a pragmatic efficacy randomised controlled trial with internal pilot

Sheffield Hallam University

Principal investigator: Prof. Abigail Millings

YOUTH ENDOWMENT FUND

Evaluation of Tavistock Relationships MBT-PP to improve child outcomes by reducing inter-parental conflict: a pragmatic efficacy cluster randomised controlled trial with internal pilot
Statistical analysis plan

Evaluating institution: Sheffield Hallam University
Principal investigator(s): Prof. Abigail Millings

Project title	*Evaluation of Tavistock Relationships MBT-PP to improve child outcomes by reducing inter-parental conflict: a pragmatic efficacy cluster randomised controlled trial with internal pilot*
Developer (Institution)	*Tavistock Relationships*
Evaluator (Institution)	*Sheffield Hallam University*
Principal investigator(s)	*Prof. Abigail Millings*
SAP author(s)	*Sean Demack, Prof. John Reidy, Prof. Abigail Millings, Dr Elaine Clarke*
Trial design	*Two armed balanced 2 level clustered RCT with children clustered into (referred) families and randomisation at the family level.*
Trial type	Efficacy with internal pilot
Evaluation setting	3 x local authorities (Bristol, Bournemouth, Christchurch & Poole (BCP) and Dorset). Intervention will be embedded into Early Help, creating a 'reducing parental conflict' referral pathway
Target group	*Families with children/young people aged 7-14 experiencing inter-parental conflict*
Number of participants	*Protocol*

Primary outcome and data source	*Child externalising and internalising problems (as measured by the mothers reported SDQ 'total difficulties' scale)*
Secondary outcome and data source	*Inter-parental conflict (parent report and child report)* *Child psychological well-being (child report)* *Parenting style (parent report)* *Parent mentalising capacity (parent report)* *Parent anger expression (parent report)* *Parent emotional adaptation (parent report)*

SAP version history

Version	Date	Changes made and reason for revision
1.1 *[latest]*	17/01/25	Responding to reviewer comments.
1.0 [*original*]		*[leave blank for the original version]*

Any changes to the design or methods need to be discussed with the YEF Evaluation Manager and the developer team prior to any change(s) being finalised. Describe in the table above any agreed changes made to the evaluation design. Please ensure that these changes are also reflected in the SAP (CONSORT 3b, 6b).

Table of Contents

Introduction

Interparental conflict (IPC), regardless of whether parents are still together or separated/divorced, has long lasting negative impacts on children and young people (CYP), including contributing to emotional and behavioural problems (van Eldik et al., 2020). The intervention under evaluation in this efficacy trial is Mentalization Based Therapy for Parenting Under Pressure (MBT-PP), delivered by Tavistock Relationships. MBT-PP is a 10-session intervention delivered over 16 weeks, to parents experiencing high conflict, by a trained therapist. Sessions are delivered online and are, after individual assessment sessions, attended by both parents simultaneously if possible. There is no direct therapeutic work with CYP. MBT-PP is offered to parents experiencing high levels of intense and poorly resolved conflict. The goal of MBT-PP is to help parents to reduce their conflict, which is expected to lead to improves outcomes for CYP. The theory of change for MBT-PP is that by increasing parents' capacity to mentalise (think about the thoughts and feelings of the other person, in real time) about their co-parent's and their child(ren)'s experience, parents become more aware of the emotional impact of their conflict on their child(ren), and express less anger towards their co-parent (Hertzmann et al., 2016). They may also become more emotionally adapted to the reality of their situation (e.g., if newly separated). This should lead to a reduced level of conflict between parents (Millings et al., 2020), which means that CYP perceive less conflict between parents. CYP then have fewer experiences of blaming themselves for, or fearing becoming the target of the conflict, which in turn leads to CYP having fewer emotional and behavioural problems, and better overall psychological well-being.

Design overview

This pragmatic efficacy trial with internal pilot is a two-armed balanced 2 level clustered RCT, in which MBT-PP is compared against Treatment As Usual (TAU). TAU is support for inter-parental conflict based around the digital resources 'Getting It Right for Children' for separated parents, and 'Argue Better' for intact parents. Both resources are created by OnePlusOne and are designed to be used as guided self-help, in which clients work through the materials and also check in with a practitioner about their progress (Cavanagh & Millings, 2013). TAU is delivered by local authority staff.

This trial is running in 3 local authorities, Bristol, Dorset, and Bournemouth, Christchurch, and Poole (BCP). In each local authority, this evaluation funds a 'Gateway Lead' post. Postholders are staff with experience of working with CYP and families and have backgrounds in primary teaching and family support.

Gateway Leads have created referral pathways into the project and promoted awareness of the project in schools and other relevant local authority staff. When referred, families are screened for eligibility against exclusion and inclusion criteria and taken through the consent

process. Once consent is obtained, families complete baseline assessments, and are subsequently randomly allocated to receive either MBT-PP or TAU. Allocation is conducted by the evaluator according to a minimisation protocol to ensure a balanced sample. Families then undergo their allocated treatment. Once treatment has completed, families complete post-intervention assessments. Follow up assessments are completed 3 months later.

The primary outcome variable for this trial is the 'Total Difficulties' (TD) subscale of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) obtained from the mothers' post-intervention assessment, reporting on the CYP. As such, analyses will focus on comparing mothers' post-intervention TD scores between MBT-PP and TAU. Secondary outcome variables are fathers' post-intervention TD score; and child psychological well-being (as measured by the Stirling Children's Well-being Scale, Liddle & Carter, 2015).

The trial is also assessing proposed mechanisms of action, by assessing: parent report IPC (as measured by the O'Leary-Porter Scale, OPS, Porter & O'Leary, 1980, adapted as described in the protocol); child perception of IPC (as measured by the Perceptions of Interparental Conflict-Intensity/Frequency Scale, PIC-I/F, Kline, Wood & Moore, 2003); parent anger (as measured by the Dimensions of Anger Reactions-Revised, DAR-R, Nederlof et al., 2009); parent mentalising capacity (as measured by the Parental Reflective Function Questionnaire, PRFQ, Luyten et al., 2009); parenting style (as measured by the Parenting Scale Short Form, PS-8, Kliem et al., 2019); and, for separated parents only, parent emotional adaptation to separation (as measured by the Emotional Adaptation to Relationship Dissolution Assessment, EARDA, Millings et al., 2020).

Table 1. Trial design overview

Trial design, including number of arms		Two armed balanced 2 level clustered RCT with children clustered into (referred) families and randomisation at the family level.
Unit of randomisation		Family
Stratification variables (if applicable)		Age group (all CYP aged 8-11 / all CYP aged 12-14 / CYP aged 8-11 and 12-14), Minority ethnic group status of one or both parents (yes / no), relationship status (separated/intact)
Primary outcome	**Variable**	a) Child internalising and externalising behaviours
	Measure (instrument, source, scale)	a) Mother reported SDQ, total difficulties scale (Goodman, 1997). Scale 0 to 40.
Secondary outcome(s)	**Variable(s)**	b) Child psychological well-being c) Parent report IPC

		d) Child perception of IPC e) Parent anger f) Parent mentalising capacity g) Parenting style h) Parent emotional adaptation to separation (separated parents only)
	Measure(s) **(instrument, source, scale)**	b) Stirling Children's Well-being Scale (SCWBS; Liddle & Carter, 2015). Scale 12 to 60. c) O'Leary-Porter Scale (OPS; Porter & O'Leary, 1980). Scale 0 to 40. d) Perceptions of Interparental Conflict-Intensity/Frequency Scale (PIC-I/F; Kline, Wood & Moore, 2003). Scale 13 to 78. e) Dimensions of Anger Reactions- Revised (DAR-R; Nederlof, Hovens, Muris & Novaco, 2009). Scale 0 to 28. f) Parental Reflective Function Questionnaire (PRFQ; Luyten et al., 2009). Scale 1 to 7. g) Parenting Scale Short Form (PS-8; Kliem et al., 2019), a short form of the Parenting Scale (Arnold et al., 1993). Scale 1 to 7. h) Emotional Adaptation to Relationship Dissolution Assessment (EARDA; Millings et al., 2020). Scale 0 to 5.
Baseline for primary outcome	**Variable**	Scores at baseline (pre-randomisation) on variable listed above
	Measure (instrument, source, scale)	As above
Baseline for secondary outcome	**Variable(s)**	Scores at baseline (pre-randomisation) on variables listed above
	Measure(s) (instrument, source, scale)	As above

Sample size calculations

Table 2. Sample size calculations overview

		Protocol	SAP (drawing on empirical estimates for explanatory power and ICC)
Minimum Detectable Effect Size (MDES)		0.15-0.20 sds	0.17 sds
Pre-test/ post-test correlations	Level 1 (participant)	0.50 to 0.70 (R^2 between 0.25 & 0.49)	0.87 (R^2 = 0.75)
	Level 2 (cluster)	0.50 to 0.70 (R^2 between 0.25 & 0.49)	0.78 (R^2 = 0.61)
Intracluster correlations (ICCs)	Level 1 (participant)	n/a	n/a
	Level 3 (cluster)	0.01 to 0.15	0.27
Alpha		0.05	0.05
Power		0.80	0.80
One-sided or two-sided?		Two	Two
Average cluster size		2 CYP per family	1.4 CYP per family
Number of clusters	Intervention	175	125
	Control	175	125
	Total	350	250
Number of participants (CYP)	Intervention	350	175
	Control	350	175
	Total	700	350

The power analyses published in the protocol included key parameters for which no empirical estimates could be obtained. Specifically, these were covariate explanatory power (from the baseline mothers SDQ measure) at CYP and family levels and the family level ICC for the mothers SDQ measure at outcome. The protocol used relatively wide ranges of values for CYP and family level explanatory power (between 0.25 and 0.49) and ICC (0.01 to 0.15).

As specified in the protocol, analyses of data from the internal pilot provides empirical point estimates for these parameters for use in the updated power analyses presented here. At the time of writing (August 2024), a total of 43 families (21 allocated to MBT and 22 to TAU) with 61 CYP (30 MBT, 31 TAU) had complete baseline/T1 and outcome (T2, post intervention) mothers SDQ. Analyses of these 61 CYP in 43 families has been used to provide empirical estimates for ICC and explanatory power at the family and CYP levels.

The power analyses were undertaken using equation 1.0 below and checked using the PowerUp! software (Dong & Maynard, 2013, sheet CRA2_2r).

[1.0] $$MDES_{2LCRT} \sim M_{J-m-2}\sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{ICC(1-R^2_{Fam})}{J} + \frac{(1-ICC)(1-R^2_{CYP})}{nJ}}$$

For a 2-level CRT design, the MDES is influenced by (estimates updated from analyses of pilot data):

- n=number of CYP per family = **1.4** (updated from 2.0 in protocol).
- J = number of families (**90** at pilot, allowed to vary up to **400** see below).
- P = proportion of families allocated to intervention group (= **0.50**).
- m= number of (level 2) covariates used (which will include: group membership, family-level pre-test, 4 (dummy) variables used for minimisation ~ **6** variables).
- M_{J-m-2} is the group effect multiplier value of the t-distribution for a 2-tailed test with alpha=0.05 & beta=0.80.
- ICC is the family level ICC = **0.27** (updated from 0.01 to 0.15 in protocol). This is the proportion of variance of the outcome at level 2 (between-family variance).
- R^2_{CYP} = **0.75** (updated from 0.25 to 0.49 in protocol). The proportion of within-family child level variance that is reduced by covariate(s). This is the within-family, between-CYP explanatory power.
- R^2_{Fam} = **0.61** (updated from 0.25 to 0.49 in protocol). The proportion of between-family variance that is reduced by covariate(s) – between-family explanatory power.

Four changes between the protocol and SAP have been made. As shown in equation 1, two of these (reduced CYP per family and increased ICC) will result in reducing the statistical sensitivity of the clustered RCT design. At the same time, two changes (increased explanatory power at CYP and family levels) will result in increasing the statistical sensitivity. To illustrate

the differences, Table 3 shows MDES estimates for a range of sample sizes from power analyses presented in the protocol and updated in this SAP.

Table 3. Sample sizes and MDES estimates for the protocol and SAP stages of the Tavistock MBT-PP impact evaluation

MDES estimates	Protocol • 2.0 CYP per family • ICC ~ 0.01-0.15 • CYP explanatory power ~ 0.25-0.49 • Fam explanatory power ~ 0.25-0.49	SAP • 1.4 CYP per family • ICC = 0.27 • CYP explanatory power = 0.75 • Fam explanatory power = 0.61
90 Families	0.30-0.40	0.29
150 Families	0.23-0.31	0.22
200 Families	0.20-0.27	0.19
250 Families	0.18-0.24	0.17
300 Families	0.16-0.22	0.16
350 Families	0.15-0.20	0.15
400 Families	0.14-0.19	0.14

The SAP power analyses resulted in MDES estimates closer to the lower range estimates presented in the protocol. At protocol stage, it was decided that 350 families (with 700 CYP) would be required for the clustered RCT design to be able to detect an effect size of 0.20 or higher as statistically significant ($p<0.05$, two-tailed) with a statistical power of 0.80 or higher. With 350 families, the MDES estimates ranged between 0.15 and 0.20 sds (depending in the ICC and explanatory power). With an attrition rate of 10%, assuming attrition was random, the indicative MDES estimates ranged between 0.16 and 0.21 sds. With an attrition rate of 20%, assuming attrition was random, the indicative MDES estimates ranged between 0.17 and 0.23 sds.

The updated power analyses presented here illustrate that, even with fewer CYP per family than initially estimated, a smaller sample of families is likely to be sufficient sensitivity to detect an MDES of 0.20 sds or higher than was estimated at protocol. With 250 families (and 350 CYP), the MDES estimate is 0.17 sds. With an attrition rate of 10%, assuming attrition was random, the indicative MDES estimate is 0.18 sds. With an attrition rate of 20%, assuming attrition was random, the indicative MDES estimate is 0.19 sds.

Therefore, following the updated power analysis, the recommended sample size for the evaluation of the Tavistock MBT-PP programme is reduced from 350 families (700 CYP) to be 250 families (350 CYP).

Analysis

Research questions were operationalised as follows:

RQ1 Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower externalising and internalising behaviours in children aged 8-14 (as measured by the SDQ total 'difficulties' scale)?

This splits into two separate RQs (SDQ total difficulties for mothers and for fathers).

> **RQ1a** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower externalising and internalising behaviours in children aged 8-14 as measured by the mother's SDQ total 'difficulties' scale? – PRIMARY OUTCOME RQ1
>
> **RQ1b** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower externalising and internalising behaviours in children aged 8-14 as measured by the father's SDQ total 'difficulties' scale? – SECONDARY OUTCOME RQ1

RQs 2-8 inclusive also all examine secondary outcomes, and, apart from those where the outcome variable comes from CYP rather than parents (RQ2 regarding child wellbeing and RQ4 regarding child perception of IPC), they split into a) maternal report and b) paternal report, as per RQ1.

RQ2 Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to higher wellbeing in children aged 8-14 (as measured by the Stirling Children's Well-being Scale)?

RQ3 Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower IPC reported by parents (as measured by the O-Leary-Porter Scale)?

> **RQ3a** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower IPC reported by mothers (as measured by the O-Leary-Porter Scale)?
>
> **RQ3b** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower IPC reported by fathers (as measured by the O-Leary-Porter Scale)?

RQ4 Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower IPC reported by children (as measured by the Children's Perception of Interparental Conflict Scale)?

RQ5 Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower parent anger expression (as reported by the Dimensions of Anger Reactions-Revised)?

> **RQ5a** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower mother anger expression (as reported by the Dimensions of Anger Reactions-Revised)?
>
> **RQ5b** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to lower father anger expression (as reported by the Dimensions of Anger Reactions-Revised)?

RQ6 Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to higher mentalising ability in parents (as reported by the Parental Reflective Function Questionnaire)?

> **RQ6a** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to higher mentalising ability in mothers (as reported by the Parental Reflective Function Questionnaire)?
>
> **RQ6b** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to higher mentalising ability in fathers (as reported by the Parental Reflective Function Questionnaire)?

RQ7 Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to more positive parenting (as measured by the Parenting Scale Short Form PS-8)?

> **RQ7a** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to more positive parenting by mothers (as measured by the Parenting Scale Short Form PS-8)?
>
> **RQ7b** Does MBT-PP (compared to TAU) delivered to parents experiencing IPC lead to more positive parenting by fathers (as measured by the Parenting Scale Short Form PS-8)?

RQ8 Does MBT-PP (compared to TAU) delivered to separated parents experiencing IPC lead to better parent emotional adaptation to relationship dissolution (as measured by the EARDA)?

> **RQ8a** Does MBT-PP (compared to TAU) delivered to separated parents experiencing IPC lead to better mother emotional adaptation to relationship dissolution (as measured by the EARDA)?

RQ8b Does MBT-PP (compared to TAU) delivered to separated parents experiencing IPC lead to better father emotional adaptation to relationship dissolution (as measured by the EARDA)?

Error! Not a valid bookmark self-reference. presents the primary and secondary outcomes for the Tavistock MBT-PP impact evaluation.

Table 4. Tavistock MBT-PP Impact Evaluation: Primary & Secondary Outcomes

Outcome Measure	Baseline (Pre minimisation)	T2 (Post Intervention)	T3 (Longitudinal 3-month follow-up)
Mothers reported SDQ total difficulties score	✓	✓(Primary)	✓(Secondary)
Fathers reported SDQ total difficulties score	✓	✓(Secondary)	✓(Secondary)
Stirling Children's Wellbeing Scale	✓	✓(Secondary)	✓(Secondary)
Perceptions of Interparental Conflict-Intensity/Frequency Scale	✓	✓(Secondary)	✓(Secondary)
Mothers reported O'Leary-Porter Scale	✓	✓(Secondary)	✓(Secondary)
Fathers reported O'Leary-Porter Scale	✓	✓(Secondary)	✓(Secondary)
Mothers reported Dimensions of Anger Reactions-Revised	✓	✓(Secondary)	✓(Secondary)
Fathers reported Dimensions of Anger Reactions-Revised	✓	✓(Secondary)	✓(Secondary)
Mothers reported Parental Reflective Function Questionnaire pre-mentalising modes subscale	✓	✓(Secondary)	✓(Secondary)
Fathers reported Parental Reflective Function Questionnaire pre-mentalising modes subscale	✓	✓(Secondary)	✓(Secondary)
Mothers reported Parental Reflective Function Questionnaire certainty about mental states subscale	✓	✓(Secondary)	✓(Secondary)
Fathers reported Parental Reflective Function Questionnaire certainty about mental states subscale	✓	✓(Secondary)	✓(Secondary)
Mothers reported Parental Reflective Function Questionnaire interest/ curiosity about mental states subscale	✓	✓(Secondary)	✓(Secondary)
Fathers reported Parental Reflective Function Questionnaire interest/ curiosity about mental states subscale	✓	✓(Secondary)	✓(Secondary)
Mothers reported Parenting Scale Short Form	✓	✓(Secondary)	✓(Secondary)
Fathers reported Parenting Scale Short Form	✓	✓(Secondary)	✓(Secondary)

Outcome Measure	Baseline (Pre minimisation)	T2 (Post Intervention)	T3 (Longitudinal 3-month follow-up)
Mothers reported Emotional Adaptation to Relationship Dissolution Assessment [a]	✓	✓(Secondary)	✓(Secondary)
Fathers reported Emotional Adaptation to Relationship Dissolution Assessment [a]	✓	✓(Secondary)	✓(Secondary)

[a] Separated parents only.

The analysis described in this plan follows YEF analysis guidance[1]. An Intention to Treat (ITT) approach will be taken whereby CYP are identified as members of the intervention or control group solely based on their random allocation and regardless of whether their parent(s) participate in the intervention or not. The ITT approach best preserves random allocation. This means that the only difference between the two group (of families/parents) is that one group received the Tavistock MBT-PP programme and the other groups (TAU) did not. All other differences between these groups are (fixed to be) random at baseline and, assuming zero attrition, this assumption is maintained at outcome. It is the ITT analysis of the primary outcome (mothers reported SDQ total difficulties score) that this clustered RCT is statistically powered to detect. Specifically and assuming zero attrition, if a difference between the intervention and control group is observed to be 0.17 sds or higher, this would represent evidence that participation in the Tavistock programme caused this positive (or negative) impact within pre-specified levels of statistical significance ($p<0.05$, two-tailed) and statistical power (0.80 or higher). As detailed below, following the ITT analysis of the primary outcome, missing data and Compliers Average Causal Effect (CACE) analyses will be undertaken.

An ITT approach will also be used for analyses of secondary outcomes. Follow-on analyses of the primary outcome will estimate the Compliers Average Causal Effect (CACE); the impact of the programme for CYP located in families/parents known to have engaged in the MBT-PP programme as was intended (see below).

All analyses will be undertaken using STATA v18. Specifically, parameter estimates will be obtained using the 'mixed' suite of commands in the STATA software. For primary and secondary outcomes, pre-test covariates will be included at both CYP and cluster (family) levels. These will be centred as recommended by Hedges & Hedberg (2007 & 2013) and Konstantopoulos (2008). Specifically, at the CYP level, scores will be centred around the cluster (family/parent) mean and at the family/parent level the cluster(family) means will be centred around the overall cluster-level grand mean. Other than the constant term, centring

[1] https://youthendowmentfund.org.uk/resources-for-evaluators/

will not result in different estimates of coefficients (and hence impact) compared with using raw scores at CYP level and aggregated mean scores at the family/parent level but brings other statistical advantages. Hedges & Hedberg (2007) note that using cluster (mean) centring results in more stable estimates of variance components when covariate values vary substantially across clusters. Centring also ensures no issues of multicollinearity because the correlation between the cluster- and CYP-level centred variables will be zero[2]. Finally, centring results in a more realistic coefficient estimate for the constant term because this relates to a mean score (rather than a score of zero) on the T1 mothers SDQ-TD score.

Primary outcome analysis

A multilevel linear regression model will be used to estimate the impact of the Tavistock MBT-PP intervention, on mothers-reported SDQ total difficulties (SDQ-TD) score for their children. Whilst mothers-reported SDQ -TD score represents the single primary outcome for the impact evaluation, fathers SDQ-TD score will also be analysed as a secondary outcome (see below).

For the analysis of the primary outcome, an estimate of impact will be obtained from a hierarchical (multilevel) linear model described in equation 2.0:

$$[2.0] \quad Y_{ij} = \beta_0 + \beta_1 T_j + \beta_2 (X_{ij} - \bar{X}_j) + \beta_3 \left(\bar{X}_j - \bar{X}\right) + \beta_{4..7} [MIN\ DUM]_j + \vartheta'_j + \varepsilon'_{ij}$$

Where:

- Y_{ij} is the mothers reported SDQ-TD for CYP i in family j at T2 (16 weeks following minimisation).
- T_j is the family-level group identifier which will take the value 1 (=MBT) or zero (=TAU)
- β_1 is the coefficient that estimates the difference between the MBT and TAU groups adjusted for baseline scores and minimisation variables; i.e. $\beta_1 = (MBT - TAU)_{adjusted}$. The $\beta_1$coefficient will be used to estimate the AITT (see below).
- X_{ij} is the mothers reported SDQ-TD for CYP i in family j at T1/baseline centred around $\bar{X}_j$, the family-level mean (i.e. $(X_{ij} - \bar{X}_j)$)
- $\bar{X}$ is the family-level grand mean mothers reported SDQ-TD score used to centre $\bar{X}_j$, the family-level mean (i.e. $\bar{X}_j - \bar{X}$))
- $[MIN\ DUM]_j$ is the collection of four family-level dummy variables used in the minimisation; ages of CYP in family (two variables); whether one parent was not White British (one variable) and relationship status (one variable).
- $\beta_{4..7}$ are the coefficient estimates for the four family-level dummy variables.

[2] This because centring the CYP level scores around their cluster (family) mean is an orthogonal transformation, CYP scores are placed relative to their family mean and so will not share any variance (r = 0).

- ϑ_j' and ε_{ij}' are random effects that remain at the school and pupil levels and are assumed to be distributed normally in the population with zero means, variances σ_c^2 and τ_c^2 respectively, and for these variances to be conditionally uncorrelated.

The estimate of impact will be standardised as an effect size by dividing by the standard deviation obtained from the variance partition estimates for the empty/null model described in equation 2.1:

$$[2.1] \quad Y_{ij} = \beta_0 + \vartheta_j + \varepsilon_{ij}$$

Where:

- β_0 is the constant term for the empty model and represent the mean mothers reported SDQ-TD for CYP at T2 (16 weeks following minimisation).
- ϑ_j and ε_{ij} are random effects at the school and pupil levels and are assumed to be distributed normally in the population with zero means, variances σ^2 and τ^2 respectively.

Table 5 summarises the ITT analysis for the mothers SDQ-TD primary outcome described in equation 2.0.

Table 5. Example analysis model for mothers SDQ-TD primary outcome (RQ1a)

Analysis and Sample	*Level 1 (CYP) Covariates*	*Level 2 (Family) Covariates*	*Outcome Variable*
Null/Empty model			CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)
Final Model	Mothers SDQ-TD score at T1/baseline (Centred around family mean)	Mean mothers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Four Minimisation Dummy Variables: CYP ages (2); parental ethnicity (1), relationship status (1);	CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)

The empty model is described by equation 2.1 and the final model by equation 2.0. The impact of the MBT-PP programme (compared with TAU) amongst parents experiencing IPC on

mothers SDQ-TD scores will be estimated using the coefficient for the group membership family-level covariate included in the final model (β_1 in equation 2.0). Impact will be standardised into an effect size by dividing the β_1 coefficient shown by the unconditional standard deviation from the null/empty model described in equation 6.0 below.

Secondary outcome analysis

Table 4 above lists a total of 35 secondary outcomes. For all 35 secondary outcomes, an ITT approach will be taken.

For the analysis of all secondary outcomes, an estimate of impact will be obtained from a hierarchical linear model described in equation 3.0:

$$[3.0] \quad Y_{ij} = \beta_0 + \beta_1 T_j + \beta_2 (X_{ij} - \bar{X}_j) + \beta_3 \left(\bar{X}_j - \bar{X}\right) + \beta_{4..7} [MIN\ DUM]_j + \vartheta_j + \varepsilon_{ij}$$

Where:

- Y_{ij} is secondary outcome score for CYP i in family j at T2 (16 weeks following minimisation).
- T_j is the family-level group identifier which will take the value 1 (=MBT) or zero (=TAU)
- β_1 is the coefficient that estimates the difference between the MBT and TAU groups adjuested for basline scores and mimimisation variables; i.e. $(MBT - TAU)_{adjusted}$
- X_{ij} is the baseline score for the secondary outcome for CYP i in family j at T1/baseline centred around $\bar{X}_j$, the family-level mean (i.e. $(X_{ij} - \bar{X}_j)$)
- $\bar{X}$ is the family-level grand mean for the secondary outcome and used to centre $\bar{X}_j$, the family-level mean (i.e. $\bar{X}_j - \bar{X}$)
- $[MIN\ DUM]_j$ is the collection of four family-level dummy variables used in the minimisation; ages of CYP in family (two variables); whether one parent was not White British (one variable) and relationship status (one variable).
- $\beta_{4..7}$ are the coefficient estimates for the four family-level dummy variables.
- ϑ_j' and ε_{ij}' are random effects that remain at the school and pupil levels and are assumed to be distributed normally in the population with zero means, variances σ_c^2 and τ_c^2 respectively.

Table 6 summarises the ITT analysis for the fathers SDQ-TD secondary outcome as an example, a similar approach will be used for all secondary outcomes at T2.

Table 6. Example analysis model for fathers SDQ-TD primary outcome (RQ1b)

Analysis and Sample	*Level 1 (CYP) Covariates*	*Level 2 (Family) Covariates*	*Outcome Variable*

Empty model			CYP level fathers SDQ-TD (raw) score at T2 (16 weeks following minimisation)
Final Model	Fathers SDQ-TD score at T1/baseline (Centred around family mean)	Mean fathers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Four Minimisation Variables: CYP ages (2), parental ethnicity (1), relationship status (1).	CYP level fathers SDQ-TD (raw) score at T2 (16 weeks following minimisation)

Subgroup analyses

To examine evidence of differential impact for Tavistock MBT, three subgroup analyses are planned for the (mothers SDQ-TD) primary outcome. Differential impact will be examined relating to CYP age and ethnicity and the relationship status of the parents. All analyses of subgroups are exploratory.

Two analysis stages will be used to examine evidence of differential impact. Stage one will add two terms to the ITT analysis model. A main effects term will be added along with a term that interacts with group membership. If the interaction term is observed to be statistically significant (p<0.05, two tailed), this provides evidence of differential impact and so separate impact analyses will be undertaken for different CYP subsamples.

For example, to examine evidence of differential impact for different aged CYP, CYP age will be included as a main effect (CYP_Age_{ij}) and as an interaction with the family-level group identifier ($T_j \times CYP\_Age_{ij}$), as described in equation 4.0

[4.0] $Y_{ij} = \beta_0 + \beta_1 T_j + \beta_2(X_{ij} - \bar{X}_j) + \beta_3(\bar{X}_j - \bar{X}) + \beta_4 CYP\_Age_{ij} + \beta_5 T_j \times CYP\_Age_{ij} + \beta_{6..9}[MIN\ DUM]_j + \vartheta_j + \varepsilon_{ij}$

If the β_5 interaction coefficient is observed to be statistically significant, a second stage analysis will estimate impact for two subsamples; CYP aged 7-11 and CYP aged 12-14[3]. This

[3] These age groupings were selected to reflect the Primary (Key Stage 2) and Secondary (Key Stage 3) educational phases and used as controls in the minimisation.

will be done by running separate analyses (described in equation 2.0 above) for these two CYP age groups.

To examine evidence of differential impact relating to the relationship status of parents, a relationship status binary family-level main effect(s) term ($e.g. FAM\_RELSTATUS_{ij}$) would be added along with interaction with the family-level group identifier ($T_j * FAM_RELSTATUS_{ij}$etc). $FAM\_RELSTATUS_{ij}$ takes the value of 0 (intact) or 1 (separated). If the interaction term is observed to be statistically significant, a second stage analysis will estimate impact separately for CYP from intact and from separated families.

Analyses that will be used to examine evidence of differential impact relating to CYP ethnicity are less simple to specify here because of unknowns around the size of subsamples for some specified ethnic groups. The value of examining evidence of differential impact relating to CYP ethnicity is greater when ethnicity can be defined using specific and multiple groups (e.g., the ONS Minor classification[4]; Indian, Bangladesh, Black Caribbean, Black African, White British etc) compared with more aggregated classifications (e.g. the ONS Major Classification; White, Asian, Black) and (even more so) the binary Black and Asian Minoritised Ethnicities (BAME) classification. The aim will be to undertake the analyses using as defined measure of ethnicity that is possible with the data without introducing a risk of statistical disclosure. Subsamples of (n =) 30 or higher will be examined. The trial has a sample size aim of 350 CYP, and so subsamples of around 10% of this would be included.

Table 1 in the protocol provided some summary ethnicity statistics for the three local authority areas included in the evaluation which ranged from being 78% (Bristol) and 94% (Dorset) White British (86% across the three areas). If the recruited sample of CYP reflect these wider statistics, it seems likely that analyses would be limited to using the simple BAME binary ethnicity classification. In this event, the analyses would progress in a similar way to that specified for family 'relationship status' above; the CYP BAME binary classification would be included as a main effect term (CYP_BAME_{ij}) and as an interaction with the family-level group identifier ($T_j * CYP_BAME_{ij}$etc). CYP_BAME_{ij} takes the value of 0 (White British) or 1 (BAME). If the interaction term is observed to be statistically significant, a second stage analysis will estimate impact for White British and BAME CYP groups. These analyses would be caveated by the limitations of classifying such a broad group of ethnicities into a single (BAME) group.

[4] See https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021

Table 7 summarises subsample analysis used to examine evidence of differential impact of the MBT-PP programme relating to family relationship status for the mothers SDQ-TD primary outcome as an example. A similar approach will be used to examine differential impact relating to CYP age and ethnicity.

Table 7. Example (relationship status) subgroup analysis for mothers SDQ-TD primary outcome

Analysis and Sample	*Level 1 (CYP) Covariates*	*Level 2 (Family) Covariates*	*Outcome Variable*
Main Effects Final Impact Model (Complete ITT Sample)	Mothers SDQ-TD score at T1/baseline (Centred around family mean)	Mean mothers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Four Minimisation Variables: CYP ages (2); parental ethnicity (1), **relationship status (1);**	CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)
Interaction Model (Complete ITT Sample)	Mothers SDQ-TD score at T1/baseline (Centred around family mean)	Mean mothers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Four Minimisation Variables: CYP ages (2), parental ethnicity (1), **relationship status (1)** **Interaction: Group Membership*relationship status**	CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)
Assuming a statistically significant interaction term…			
Null/Empty Models ***Two Subsamples:*** *CYP in intact families AND*			CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)

CYP in separated families.			
Subgroup Impact Models ***Two Subsamples:*** *CYP in intact families AND CYP in separated families.*	Mothers SDQ-TD score at T1/baseline (Centred around family mean)	Mean mothers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Three Remaining Minimisation Variables: CYP ages (2), parental ethnicity (1).	CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)

Further analyses

Sensitivity analyses will be undertaken to examine two things; first to examine whether time (between baseline and outcome data collection) influenced the outcome and estimated impact; second, to examine impact at the pilot and stage 2 evaluations (which are combined to form this efficacy study with an internal pilot).

To examine the influence of time we will first statistically summarise the time (in weeks[5]) between baseline (T1) and outcome (T2) and the bivariate relationship between time and the mothers SDQ TD outcome at T2 for the MBT and TAU groups. Following this we will take a similar approach to the CYP age subsample analysis and introduce time (in weeks) as a CYP-level main effects term (centred around the CYP level mean; CYP_CWeeks_{ij}) and as an interaction with the family-level group identifier ($T_j \times CYP\_CWeeks_{ij}$) in a hierarchical linear model, as described in equation 5.0:

[5.0] $Y_{ij} = \beta_0 + \beta_1 T_j + \beta_2 (X_{ij} - \bar{X}_j) + \beta_3 \left(\bar{X}_j - \bar{X}\right) + \beta_4 CYP\_CWeeks_{ij} + \beta_5 T_j \times CYP\_CWeeks_{ij} + \beta_{6..9}[MIN\ DUM]_j + \vartheta_j + \varepsilon_{ij}$

If the β_5 main effects term for time is statistically significant, this would indicate that mothers SDQ TD at T2 was associated with the number of weeks between baseline (T1) and outcome (T2) measures. This would not necessarily indicate a problem with the estimate of impact from the ITT analysis but would provide useful context. However, if the β_5 interaction coefficient is observed to be statistically significant, this would raise a question on the validity

[5] Measured in days but included as decimalised 'weeks' with the assumption of a mean of 16.0 weeks.

of the ITT estimate (because the interaction term would suggest that this estimate depended on the amount of time between T1 and T2). If the β_5 interaction coefficient is observed to be statistically significant, the sign would indicate whether increased time (weeks between T1 and T2) resulted in higher (positive) or lower (negative) estimates of impact. Follow-on analyses would be undertaken on subsamples of participants determined by time. This might be three subsamples (below mean, mean, above mean) but is dependent on subsample size and will be determined following an examination of the data.

To examine evidence of a different impact at pilot and stage 2 of the evaluation we will first statistically summarise the mothers SDQ TD scores at T2 for the two stages alongside the scores for the combined 'efficacy with internal pilot' sample. These analyses will be substantively informed by the IPE and evidence of any changes in the theory, delivery, fidelity and compliance of MBT between pilot and stage 2. Following the descriptive statistical summary, we will adopt an similar approach to the relationship status subsample analysis specified above by including two additional variables to the ITT hierarchical linear model. A binary family-level 'Stage' main effect(s) term ($e.g. FAM\_TrialStage_{ij}$) would be added along with interaction with the family-level group identifier ($T_j * FAM_TrialStage_{ij}$etc). $FAM\_TrialStage_{ij}$ will take the value of 0 (pilot) or 1 (stage 2). If the interaction term is observed to be statistically significant, a follow-on analysis will estimate impact separately for CYP in the pilot and stage 2 evaluations. This will be compared with the ITT estimate for the combined 'efficacy trial with internal pilot' evaluation and be discussed with reference to IPE findings on changes to the theory, implementation, fidelity and compliance for the MBT-PP programme between the two stages. This is illustrated in Table 8 below:

Table 8. Sensitivity Analysis: mothers SDQ-TD primary outcome at pilot and stage 2

Analysis and Sample	*Level 1 (CYP) Covariates*	*Level 2 (Family) Covariates*	*Outcome Variable*
Main Effects Final Impact Model (Complete ITT Sample)	Mothers SDQ-TD score at T1/baseline (Centred around family mean)	Mean mothers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Four Minimisation Variables: CYP ages (2), parental ethnicity (1), relationship status (1);	CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)

Interaction Model (Complete ITT Sample)	Mothers SDQ-TD score at T1/baseline (Centred around family mean)	Mean mothers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Four Minimisation Variables: CYP ages (2), parental ethnicity (1), relationship status (1) **TrialStage (0=pilot, 1=stage 2)** **Interaction: Group Membership*TrialStage**	CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)
Assuming a statistically significant interaction term…			
Null/Empty Models ***Two Subsamples:*** *Pilot & Stage 2 evaluations.*			CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)
Subgroup Impact Models ***Two Subsamples:*** *Pilot & Stage 2 evaluations.*	Mothers SDQ-TD score at T1/baseline (Centred around family mean)	Mean mothers SDQ-TD score at T1/baseline (Centred around grand mean) Group membership (MBT or TAU) Four Minimisation Variables: CYP ages (2), parental ethnicity (1), relationship status (1)	CYP level mothers SDQ-TD (raw) score at T2 (16 weeks following minimisation)

Longitudinal follow-up analyses

The T3 3 month follow up data provides 18 of the secondary outcome variables (see Table 4). As specified above, these will be subjected to the same analyses as the primary outcome variable.

Testing the theory of change mechanisms

To examine the causal mechanisms proposed in the theory of change we will undertake multigroup structural equation modelling (SEM) comparing the model in Figure 1 across the MBT and TAU groups. The initial model will be undertaken using the post-intervention measures of the variables included in Figure 1.

Figure 1. Proposed causal model for SEM analysis with child internalising and externalising behaviours (mothers' SDQ-TD) as the final latent variable in the causal chain on post-intervention measures.

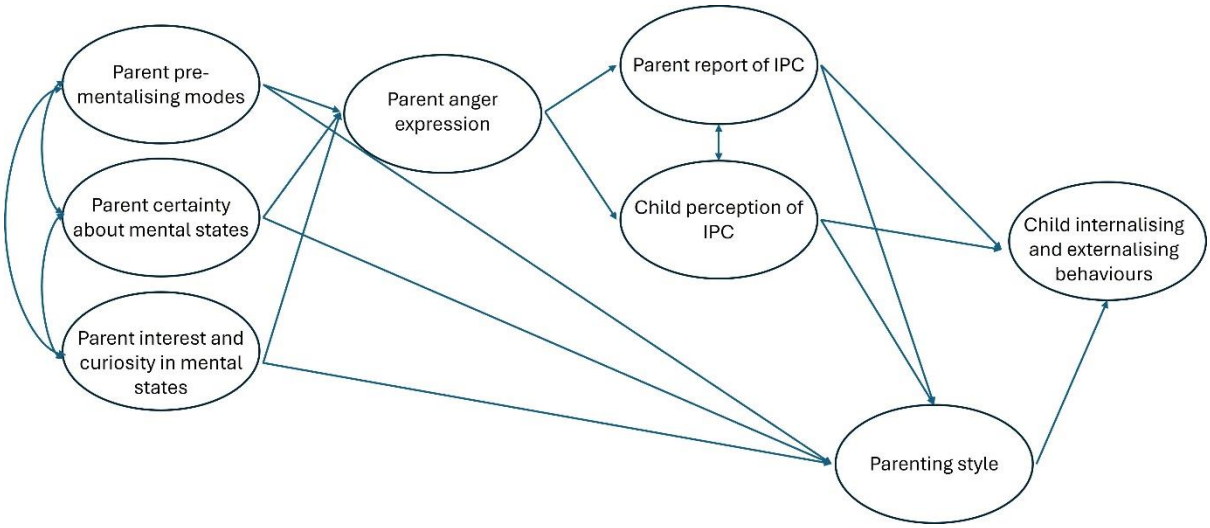
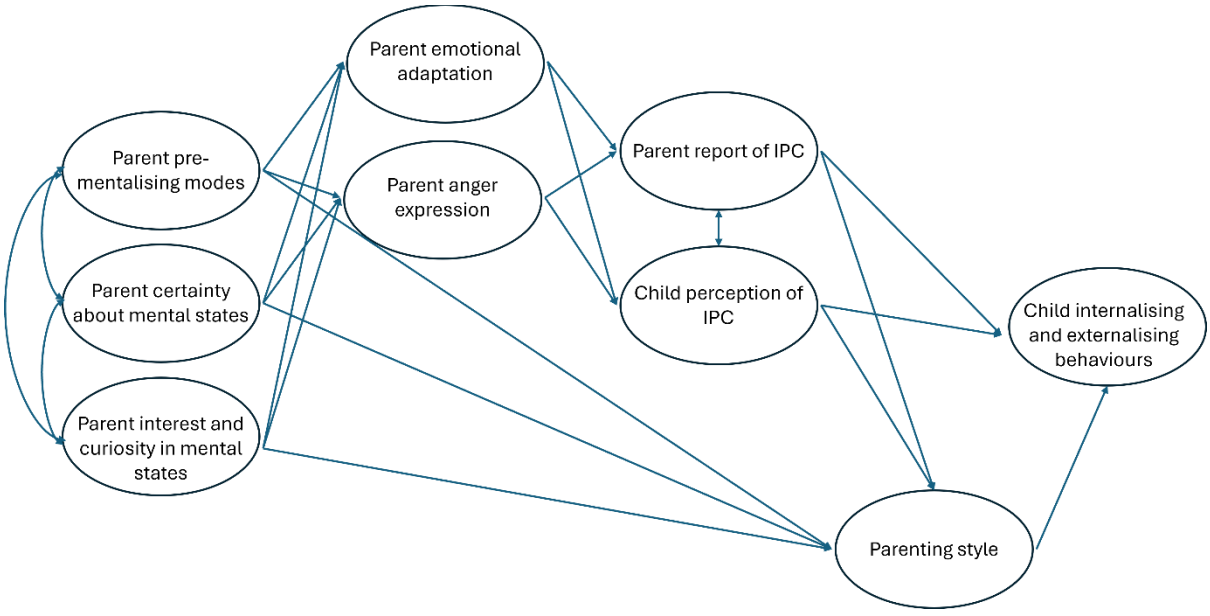


Figure 2. Proposed causal model for SEM analysis for separated parents, including the parental emotional adaptation measure.



Using the data from separated families only, we will also test the model in Figure 2, which treats emotional adaptation to relationship dissolution as an additional mediator, alongside anger expression. Again, this will be undertaken using the post-intervention measures.

Structural equation modelling (SEM) is the most appropriate technique with which to test these models because unlike path analysis it enables us to estimate measurement error within the model. The SEM analyses will be conducted using the Lavaan package in R. MBT and TAU groups will be included in the analyses together with the initial analysis fitting the same model to both groups with all parameters set to be free and model parameters being generated for each group separately. Fit of all models tested will be appraised using commonly used fit indices. Fit for the whole model and the models for each group will be determined with chi-square, CFI, RMSEA, SRMR, and AIC. Following Hu and Bentler (1999) the following criteria will be used to determine good fit (CFI > .95; RMSEA < .06; and SRMR < .08).

Before comparing latent parameters across the groups, we will examine measure invariance across the groups. Due to the random allocation of CYPs and parents to the MBT and TAU groups we would not expect differences between the groups in terms of interpretations of the constructs being measured. However, as the groups will have differing experiences during the intervention period, particularly in relation to aspects of mentalising examining measure invariance is advised prior to comparing latent variable means and variances across the two groups. We will undertake this invariance testing by examining configural variance, metric invariance and scalar invariance (see Putnick & Bornstein, 2016). Configural invariance checks that the model structure is equivalent across the groups and is tested by allowing the various parameters in the model to be free across the groups. Metric invariance ensures that, for both groups, each of the observed variables contributes to their relevant latent variables to an equal degree. This is assessed by constraining the various factor loadings to be equivalent across the groups. Scalar invariance tests the degree to which the latent variables are measured on the same scales and that any differences between the group means of the latent variables is captured by the differences of the means (intercepts) of the observed variables. This is assessed by specifying that item intercepts are equivalent across the groups. The test of measure invariance involves comparing the three modes outlined above (the configural, metric & scalar models) and undertaking likelihood ratio tests which compare chi-square statistic for the metric test model with the configural model and then a further test comparing the scalar model with the metric model. If either of these two tests is significant then this suggests potential problems with measure noninvariance. In the case of noninvariance an exploration of the models would be conducted to try to identify possible reasons for noninvariance. In the case of noninvariance we might not be able to then compare the latent

means across the two groups. In such a scenario we would test the models in each group separately and not directly compare any parameters across the two groups.

If the assumption of measure invariance is upheld, we will look to compare latent means across group with a particular focus on the mentalising latent variables as these are the key targets of the MBT intervention. This would be undertaken by restricting the models to have equal means in the mentalising latent variables and then conducting likelihood ratio tests to compare this constrained model with the model where the means are free to vary. If the likelihood ratio test is not significant then this would suggest that groups do not differ in terms of these latent means. We would also compare the regression coefficients from the mentalising latent variables to the anger expression latent variable in a similar way.

We will first utilise the above method to test a model featuring post-intervention SDQ-TD score (mothers) as the final endogenous variable in the causal chain. We will then test the same models with child well-being score at post-intervention as the final endogenous variable (figures 3 and 4). We will repeat this whole process using fathers' SDQ-TD scores. We will run these SEM models even in the absence of a main effect on each final endogenous variable, because it is possible to detect indirect effects in the absence of a direct effect, and indeed important to do so to provide information as to why the main effect is present or not (O'Rourke & Mackinnon, 2018).

Additionally, if there are sufficient responses from the three month follow-up testing for the primary (child internalising and externalising behaviours) and secondary (child psychological well-being) measures we will re-test the models in Figures 1 & 3 (and if possible Figures 2 & 4) but with replacing the post-intervention measures of these primary and secondary outcome with the three-month follow-up measures of these variables. This potentially enables us to test whether changes to mentalising, anger expression, interparental conflict and parenting style at time 2 predicts child behaviour and well-being at follow-up and thus potentially further supporting any causal relationships established from the post-intervention SEM modelling.

Figure 3. Proposed causal model for SEM analysis with the secondary outcome measure of child psychological well-being.

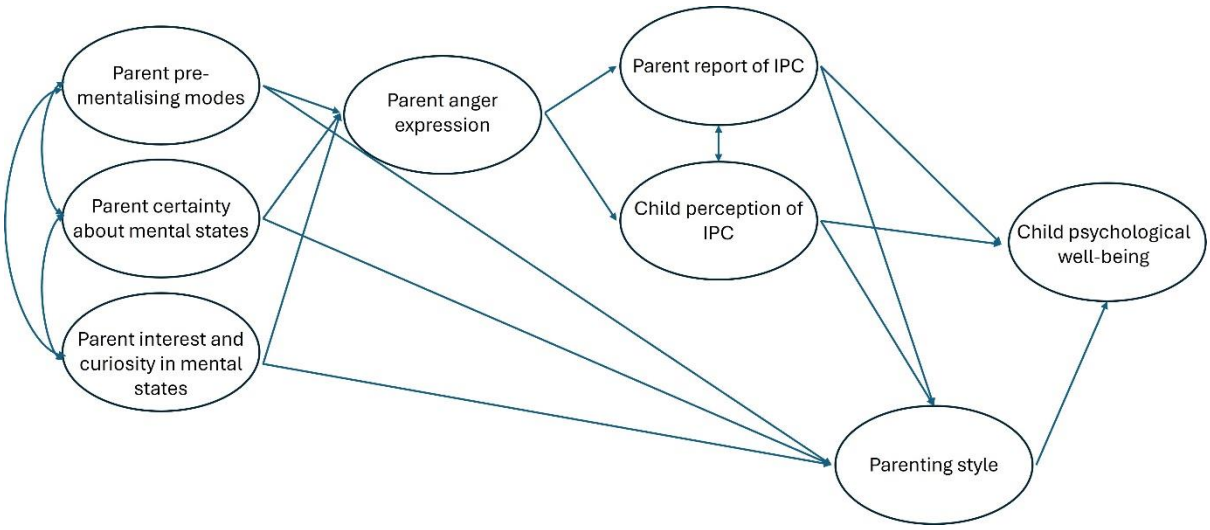
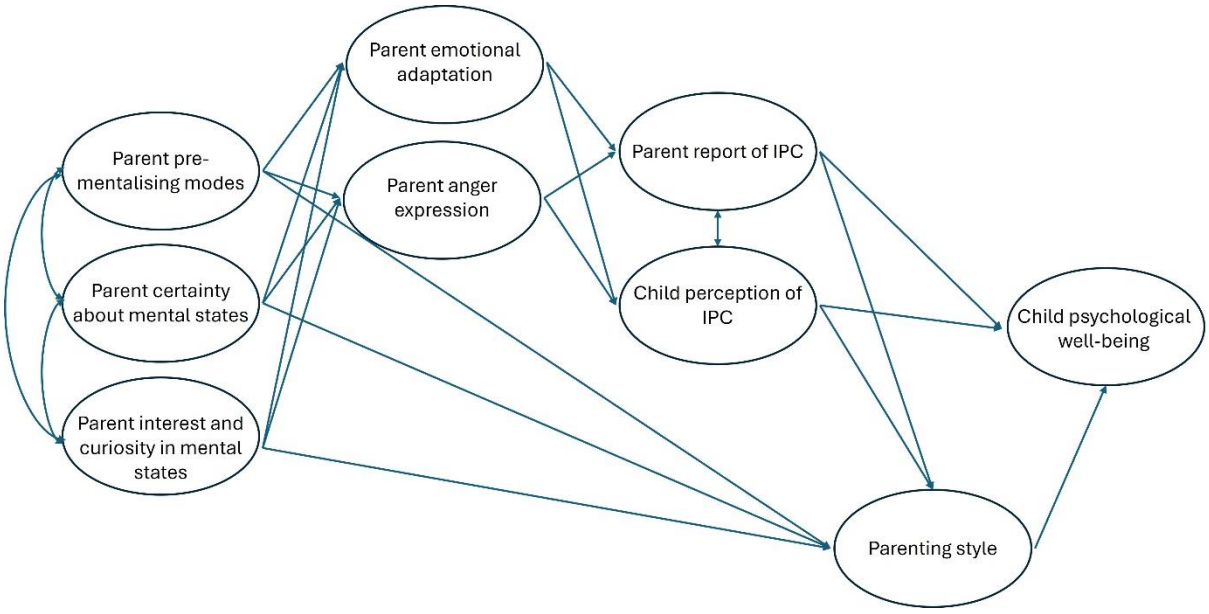


Figure 4. Proposed causal model for SEM analysis for separated parents with the secondary outcome measure of child psychological wellbeing, including the parental emotional adaptation measure.



These latter models will be tested in the same way as outlined above for the models with child internalising and externalising behaviours as the final latent variable in the causal chain.

Imbalance at baseline

The characteristics of the MBT and TAU group families and CYP as measured prior to randomisation will be compared. At the CYP level, this will include the T1 mothers SDQ-TD primary outcome and X secondary outcomes, ethnicity, age, gender, EAL and SEND status. At the family level this will include the three minimisation variables (relationship status, CYP ages and parental ethnicity] along with family size and the aggregated primary and secondary outcomes. For scale/continuous variables, means and standard deviations for the MBT and TAU groups will be reported and the difference converted into a Hedges g effect size by dividing the mean difference (MBT – TAU) by the pooled standard deviation. For categorical variables, counts and percentages will be reported.

The analyses of balance at baseline will be replicated for the subsample of CYP included in the final impact analysis for the primary outcome; a complete case subsample. In the case of very low attrition, baseline and complete case analyses should reflect each other. However, as attrition increases, differences are likely to be seen. These analyses will inform the missing data analyses described in the next section.

Missing data

The amount of missing outcome data for the mothers SDQ-TD primary outcome will be summarised, and reasons for missing data discussed in the report, where available. If greater than 5% of CYP have missing mothers SDQ-TD, multi-level logistic regression will be used to model presence or absence of the primary outcome including all available CYP and family-level baseline data as fixed effects, and family as a random effect. Significant variables and possible reasons for the missing data will be discussed in the report.

A binary variable that distinguishes CYPs with a missing T2 mothers SDQ-TD (= 1) from CYPs with a T2 score (= 0) will be created. Bivariate patterns of missingness will be described (e.g. % missing across all variables included in the ITT analyses plus auxiliary variables at the CYP and family levels) and then modelled using a multilevel logistic regression model with the CYP-level binary missing variable as outcome and the same variables used to examine baseline balance included as explanatory variables at CYP- and family-levels. This represents the 'drop-out' model which will provide estimates of the association between the included variables and the probability of being missing from the final impact analysis for the primary outcome.

If none of the covariates included in the 'drop-out' model are observed to account for a statistically significant amount of variation in the missing data outcome, imputation will not be feasible. This would lead us to cautiously conclude that the data are missing at random and would discuss the implications of this (which will depend upon the quantity of missing data) in the evaluation report.

If we observe that covariates account for a statistically significant amount of variation in the missing data outcome, we will conduct multiple imputation (MI) analysis to impute the

missing data. The MI analysis will be undertaken using the STATA 'MI Impute' command used to impute missing values on the primary outcome (StataCorp, 2023). A 'burn-in' of 20 will be used (meaning that the first 20 iterations will be discarded to allow the iterations to converge to the stationary distribution before the imputation) and 100 imputed datasets will be created. The impact analysis for the primary outcome will then be rerun within the imputed datasets using the STATA 'MI Estimate' command with Rubin's rules (Rubin, 1987) used to combine the multiply imputed estimates of the β_1 coefficient in equation 2.0. The coefficient estimates for the original complete case and MI analyses will be compared and discussed in the evaluation report.

Compliance

Compliance will be measured at the family level and is defined as families where at least one parent received 6+ sessions of MBT-PP (i.e., a family where two parents received up to 5 sessions each would not be classed as compliant). Six sessions is selected as the cut off due to previous research finding significant effects when 6 sessions was used as the minimum clinical offer (Hertzmann et al., 2016)

Compliance is assumed to be one-sided. Specifically, whilst it is possible for a family allocated to the MBT group to have not met the pre-specified compliance criteria (i.e., attending 6+ MBT-PP sessions), it is not possible for families allocated to the TAU group to have experienced any of the MBT-PP programme.

Compliance will be operationalised as a family-level binary variable which identifies CYPs in families classed as 'compliant' (= 1) and 'not compliant' (= 0). This will be used for the Compliers Average Causal Effect (CACE) analysis. An instrumental variable (IV) approach will be used with two-staged least squares to obtain CACE estimates. The first stage models the compliance binary outcome with the treatment indicator (T_j) included as a covariate. The predicted value for the compliance outcome (C_j') is then entered into a second stage model in place of the treatment indicator (T_j) as described in equations 6.1 (step 1) and 6.2 (step 2) below.

$$[6.1] \quad C_j = \beta_0 + \beta_1 T_j + \beta_2 (X_{ij} - \bar{X}_j) + \beta_3 \left(\bar{X}_j - \bar{X}\right) + \beta_{4..7} [MIN\ DUM]_j + \vartheta_j' + \varepsilon_{ij}'$$

$$[6.2] \quad Y_{ij} = \beta_0 + \beta_1 C_j' + \beta_2 (X_{ij} - \bar{X}_j) + \beta_3 \left(\bar{X}_j - \bar{X}\right) + \beta_{4..7} [MIN\ DUM]_j + \vartheta_j' + \varepsilon_{ij}'$$

These two stages are included in the STATA 'ivregress' command. Standard errors of estimates will be adjusted for clustering of CYP in families using the 'vce (cluster robust)' subcommand.

Intra-cluster correlations (ICCs)

For the mothers SDQ-TD measures at both baseline/T1 and outcome/T2, unconditional ICCs at the family level will be estimated using a null (empty) 2-level multilevel variance components model. Additionally, for the T2 primary outcome, a conditional ICC will be estimated using the partitioned variance that remains at family and CYP levels after the final impact model is fitted. Within the analyses, a table will present the variance decomposition for the two levels (family and CYP) along with the ICC estimates calculated using equation 7.0 (obtained directly using the STATA 'estat icc' command).

[7.0] $$\text{ICC} = \frac{\text{Variance}_{\text{family}}}{\text{Variance}_{\text{family}} + \text{Variance}_{\text{CYP}}}$$

Presentation of outcomes

The effect size measure to be used will be Hedges' g. This will be calculated using equation 8.0.

[8.0] $$ES = \frac{(MBT - TAU)_{adjusted}}{\sqrt{\delta^2_{Fam} + \delta^2_{CYP}}}$$

Where:

$(MBT - TAU)_{adjusted}$ is the mean difference in mothers SDQ-TD for the MBT and TAU groups adjusted for (other variables, clustering) in the final impact model. This estimate is obtained from the β_1 coefficient for the group identity (T_j) variable described in equation 2.0 above.

δ^2_{fam} and δ^2_{CYP} are both obtained from the null (empty) model described in equation 2.1 above. δ^2_{fam} is the variance in mothers SDQ-TD scores that is between-families and δ^2_{CYP} is the 'residual' variance in mothers SDQ-TD scores which is found within-families, between-CYP.

$\delta^2_{Fam} + \delta^2_{CYP}$ is the total variance in mothers SDQ-TD scores and $\sqrt{\delta^2_{Fam} + \delta^2_{CYP}}$ is the total standard deviation used to standardise the adjusted mean difference in equation 8.0.

Similarly, the upper and lower 95% confidence intervals for from the β_1 coefficient for the group identity (T_j) variable described in equation 2.0 above will be divided by $\sqrt{\delta^2_{Fam} + \delta^2_{CYP}}$ variable to provide 95% confidence interval estimates for the effect size.

References

Cavanagh, K. and Millings, A. (2013) '(Inter)personal Computing: The Role of the Therapeutic Relationship in E-mental Health', *Journal of Contemporary Psychotherapy*, 43, pp. 197–206. Available at: https://doi.org/10.1007/s10879-013-9242-z.

Dong, N. and Maynard, R.A. (2013) 'PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies', *Journal of Research on Educational Effectiveness*, 6(1), pp. 24–67. Available at: https://doi.org/10.1080/19345747.2012.67314b3.

Goodman, R. (1997) 'The Strengths and Difficulties Questionnaire: A research note', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38(5), pp. 581–586. Available at: http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x.

Hedges, L.V. and Hedberg, E.C. (2013) 'Intraclass correlations and covariate outcome correlations for planning two- and three-level cluster-randomized experiments in education', *Evaluation Review*, 7(6), pp. 445-489. Available at: https://doi.org/10.1177/0193841X14529126 .

Hedges, L.V. and Hedberg, E.C. (2007) 'Intraclass correlation values for planning group-randomized trials in education', *Educational Evaluation and Policy Analysis*, 29(1), pp. 60-87. Available at: https://doi.org/10.3102/0162373707299706.

Hertzmann, L., Target, M., Hewison, D., Casey, P., Fearon, P., & Lassri, D. (2016). Mentalization-based therapy for parents in entrenched conflict: A random allocation feasibility study. Psychotherapy, 53(4), pp. 388–401. https://doi.org/10.1037/pst0000092Hu, L.T. and Bentler, P.M. (1999) 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives', *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), pp. 1-55.

Kliem, S., Lohmann, A., Mößle, T., Foran, H.M., Hahlweg, K., Zenger, M. and Brähler, E. (2019) 'Development and validation of a parenting scale short form (PS-8) in a representative population sample', *Journal of Child and Family Studies*, 28(1), pp. 30–41. Available at: https://doi.org/10.1007/s10826-018-1257-3.

Kline, G.H., Wood, L.F. and Moore, S. (2003) 'Validation of Modified Family and Interparental Conflict Scales for Use with Young Adults from Divorced and Non-Divorced Families', *Journal of Divorce & Remarriage*, 39(3-4), pp. 125-142. Available at: https://doi.org/10.1300/J087v39n03_07.

Konstantopoulos, S. (2008) 'The power of the test for treatment effect in three-level cluster randomised designs', *Journal of Research on Educational Effectiveness*, 1(1), pp. 66-88. Available at: https://doi.org/10.1080/19345740701692522.

Liddle, I. and Carter, G.F. (2015) 'Emotional and psychological well-being in children: The development and validation of the Stirling Children's Well-being Scale', *Educational Psychology in Practice*, 31(2), pp. 174-185. Available at: https://doi.org/10.1080/02667363.2015.1008409.

Luyten, P., Mayes, L.C., Nijssens, L. and Fonagy, P. (2017) 'The parental reflective functioning questionnaire: Development and preliminary validation', *PLoS ONE*, 12(5), e0176218. Available at: https://doi.org/10.1371/journal.pone.0176218.

Millings, A., Hirst, S.L., Sirois, F. and Houlston, C. (2020) 'Emotional adaptation to relationship dissolution in parents and non-parents: A new conceptual model and measure', *PLoS ONE*, 15(10), e0239712.

Nederlof, A.F., Hovens, J.E., Muris, P. and Novaco, R.W. (2009) 'Psychometric Evaluation of a Dutch Version of the Dimensions of Anger Reactions', *Psychological Reports*, 105(2), pp. 585-592. Available at: https://doi.org/10.2466/PR0.105.2.585-592.

O'Rourke, H.P. and MacKinnon, D.P. (2018) 'Reasons for Testing Mediation in the Absence of an Intervention Effect: A Research Imperative in Prevention and Intervention Research', *Journal of Studies on Alcohol and Drugs*, 79(2), pp. 171–181. Available at: https://doi.org/10.15288/jsad.2018.79.171.

Porter, B. and O'Leary, K.D. (1980) 'Marital discord and childhood behavior problems', *Journal of Abnormal Child Psychology*, 8(3), pp. 287-295. Available at: https://doi.org/10.1007/BF00916376.

Putnick, D.L. and Bornstein, M.H. (2016) 'Measurement invariance conventions and reporting: The state of the art and future directions for psychological research', *Developmental Review*, 41, pp. 71-90.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

StataCorp (2023) *Stata 18 Multiple-Imputation Reference Manual*. College Station, TX: Stata Press. Available at: https://www.stata.com/manuals/mi.pdf.

van Eldik, W.M., de Haan, A.D., Parry, L.Q., Davies, P.T., Luijk, M.P., Arends, L.R. and Prinzie, P. (2020) 'The interparental relationship: Meta-analytic associations with children's maladjustment and responses to interparental conflict', *Psychological Bulletin*, 146(7), pp. 553-594. Available at: http://dx.doi.org/10.1037/bul0000233.

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413