Promoting Healthy Families:
A Canadian Evaluation of Two Evidence-based Parenting Programs

NCT04702191

Statistical analysis plan: 3 April 2024

Introduction and trial design

Promoting Healthy Families is a three-arm trial undertaken in the province of Ontario, in which Triple P and Circle of Security Parenting Program (COSP) are compared against a treatment as usual (TAU) control in four community-based agencies. Randomisation was undertaken stratified by agency using a block size of 6. The trial included baseline, post-intervention, six-month and 12-month assessment points. Two further waves at three and nine months focused on service use.

This statistical analysis plan (SAP) is an addendum to the original protocol for the trial, and includes an updated account of measures used to capture psychosocial outcomes. Details of the statistical analysis plan largely mirror those in the original protocol, but in the event of any differences, the current SAP supersedes the protocol. Outcomes relating to health economic assessment (service use, quality of life), parenting observations and psychophysiological assessments are not included in the SAP and will be addressed in future addenda.

Outcomes and baseline characteristics

Outcome measures. Outcomes in this trial are reflected below in Table 1. The primary outcomes for this trial are the Strengths and Difficulties Questionnaire (SDQ) total problem score, and the Overreactivity subscale of the Parenting Scale. An additional primary outcome relating to observed parenting is not addressed in this SAP. Additional outcomes are broadly grouped into parenting capacity, child outcomes and other parental outcomes. Of note is that at each assessment point, the SDQ was administered in an age-appropriate form (aged 2-3 years, aged 4 and above). Scores are interchangeable between forms of the SDQ and thus these were 'stacked' for analysis.

In the original trial protocol, the data analysis plan included Parenting Daily Hassles and Plotkin Child Vignettes. Due to issues with data collection, Parenting Daily Hassles data were not considered to be reliable and thus we dropped this outcome from the analysis plan. Plotkin Child Vignettes data were not collected due to the need to minimise time and frequency of home visits during COVID-19 lockdowns.

Baseline characteristics. A number of baseline family characteristics are also relevant to this analysis. These include:

- Child age at baseline
- Parental education (beyond high school vs not)
- Family income (moderate poverty vs no poverty)
- Family structure (single-parent vs other)

We will also consider total child behaviour problems (SDQ total problems score) and parental distress (CESD-10) as relevant baseline characteristics. We will include these baseline characteristics in relevant imputation and effect modification models.

Table 1. Outcome measures

Measure	Variable	BL	PT	6-mo	12-mo	Analysis strategy
Primary Outcomes						
Strengths and Difficulties Questionnaire (SDQ; 2-4 years; 4-18 years)	Child behavioural problems	X	X	X	X	Sum score
Overreactivity Scale	Parental harsh discipline	X	X	X	X	Sum score
Additional Outcomes						
Parenting capacity						
Composite Caregiving Questionnaire	Self-efficacy Empathy Helplessness Reflective function	X	X	X	X	Subscale scores
Difficulties in Emotion Regulation Scale (DERS-SF)	Emotion regulation	X		X	X	Sum score
Parent Cognition Scale	Parental attributions	X	X	X	X	Sum score
Laxness Scale	Dysfunctional discipline	X	X	X	X	Sum score
Brief Child Abuse Potential Inventory	Child abuse risk	X	X	X	X	Sum score
Child outcomes						
Parent-Child Conflict-Tactics Scale-Revised (CTSPC-R)	Parenting practices				X	Sum score
SDQ Prosocial Behaviour Subscale	Child prosocial behaviour	X	X	X	X	Sum score (of subscale)
Other parental outcomes						
Alcohol Use Disorders Identification Test Consumption (AUDIT-C)	Parental alcohol and substance use	X		X	X	Scale score (complex scoring)
Cannabis use	Parental alcohol and substance use	X		X	X	Single item
Center for Epidemiologic Studies Depression Scale (CESD-10)	Parental mental health	X	X	X	X	Sum score
Generalized Anxiety Disorder (GAD-7)	Parental mental health	X	X	X	X	Sum score

Marital Conflict Questionnaire	Partner conflict	X	X	X	X	Sum score
Dyadic Adjustment Scale (DAS-7)	Partner conflict	X	X	X	X	Sum score
Exploratory outcomes						
SDQ problem behaviour subscales	Child behavioural problems	X	X	X	X	Subscale scores
Other outcomes not in this SAP						
Child and Adolescent Service Use Schedule (CA-SUS)	Health service utilization (children)	X		X	X	
HUI3 proxy parental report and Health Status Classification System Preschool Version	Health-related quality of life (children)	X		X	X	
HUI3	Health-related quality of life (adult)	X		X	X	
Emotional Availability Scale (EAS)	Observed parenting (parental sensitivity)	X	X	X	X	
EAS (child scales)	Responsiveness	X	X	X	X	

BL, baseline; PT, post-treatment

Analysis principles

We will analyse this phase III effectiveness trial using the intention-to-treat principle in a superiority framework. We will use frequentist methods and a nominal significance level of 0.05 for two-sided tests. In all analyses, the primary comparisons will be Triple P vs TAU and COSP vs TAU. To control for false discovery, we will implement a Bonferroni correction within each primary outcome analysis used to conclude effectiveness (i.e. using p=0.025 for each pairwise comparison vs TAU). We will not use a Bonferroni correction for additional outcomes. All effect estimates will be accompanied by 95% confidence intervals.

Analysis strategy

Descriptive statistics. We will describe each study arm using means, standard deviations, proportions and quantiles with respect to key demographics for parents and children.

Missing data. It is unlikely that our analysis will accrue more than 10% missingness for any one outcome. However, to ensure that subsets of data can be used in later analyses, we will implement a multiple imputation strategy across all outcomes.

First, we will inspect individual scale responses at each assessment point will be inspected. In situations where participants have answered 75% or more of scales (or subscales where these are analysed separately) scored using simple sums, we will sum the available items and 'upweight' these to match the same possible range as for the full scale. This strategy is based on the generally high reliability and validity established for all study measures in this analysis, and is a conservative version of the rule originally proposed by Graham (2012). Where scales do not use simple sums (specifically, AUDIT-C), we will reserve individual items for imputation.

Second, we will specify an imputation model using fully conditional specifications. All scale variables will be analysed using predictive mean matching with five donors. Where items are imputed for AUDIT-C and cannabis use, we will use a conditional imputation model to account for the logic of these items' scoring and an ordered logit link. We will include as auxiliary variables all relevant baseline characteristics identified above, as well as the site-specific stratum variable. We will undertake 10 imputations separately for each trial arm.

It is likely that the number of variables to be included will be too large for one imputation model. If a single imputation model fails due to nonconvergence or multicollinearity, we will undertake three imputation models using the same set of auxiliary variables in each one:

- SDQ problem subscales (primary outcome) and prosocial behaviour subscale;
- Overreactivity (primary outcome) and parenting capacity outcomes; and
- other parental outcomes.

Both the Marital Conflict Questionnaire and DAS-7 pertain only to people who reported being partnered at a given assessment point. We will assess any changes in eligibility for these measures over assessment points and use conditional imputation to account for these variables. In addition, because CTSPC-R was only administered to a subset of participants for whom home visits were possible, we will not include CTSPC-R in any imputation models.

Effect estimation. The general framework for effect estimation will be generalised linear mixed-effects models, with measurements (level 1) nested within families (level 2). The general form of each model is to include terms at level 1 for categorical time and for the interaction of intervention by categorical time, and at level 2 for study site as the stratification factor. This is equivalent to a model where the baseline measure is constrained to be equal between groups. Consistent with guidance from Coffman et al. (2016), we will not include a level 2 intervention term as this is essentially uninterpretable in the context of randomisation.

For each outcome model, we will select an appropriate outcome distribution, preferring an identity link where possible. If necessary to accommodate distributional violations, we will use percentile-based confidence intervals, undertaking cluster-level bootstrapping within each imputation and pooling runs to generate confidence intervals. In situations where variables have a limited range and discrete values, we will consider use of an ordered logit model. We will use Rubin's rules to combine estimates from multiple imputations.

The test of effectiveness will focus on the six-month value of the outcome. Correspondingly, we will examine the intervention-by-time interaction for the six-month assessment point for each intervention, and undertake a Wald test for each interaction (or, if bootstrapped models are used, examine a 97.5% confidence interval).

Sensitivity analysis. We will re-estimate all outcome models using unimputed data.

Additional analyses

Effect modification analyses. We will explore moderation of intervention effects at the six-month and 12-month assessment point for primary outcomes, child outcomes and the Brief Child Abuse Potential Inventory. To simplify analysis, we will not include post-intervention estimates in these models. Each candidate moderator will be interacted with time and with intervention by time, and a Wald test will be computed on the relevant three-way interaction terms to establish the presence of effect modification. Candidate moderators include study site (given differences in TAU across each site), baseline characteristics described above, and baseline values of child conduct problems (SDQ total problems scale) and of parental distress (CESD-10).

To evaluate the presence of moderation by baseline value of the outcome, we will consider two strategies. The first is to estimate random slope models and compare models with intercept-slope covariances (i.e. with unstructured covariance matrices of random effects) against those without using standard inferential tests. However, this may pose challenges for convergence and interpretation. In the event that it does, we will use time-by-time analyses with an appropriate ANCOVA model with terms for intervention allocation, stratum and baseline value of the outcome as predictors, in addition to candidate moderators. Each candidate moderator will be interacted with intervention status, and a Wald test will be computed on the interaction term to establish the presence of effect modification.

Mediation analyses. We will use a fully longitudinal mediation model to explore causal pathways to intervention impacts where a) a plausible relationship exists between the intervention, a mediator and an outcome; b) the effect estimate of the mediator is significant at six months; and c) the effect estimate of the outcome is significant at 12 months (using a $p<0.05$ threshold). This model will include paths between intervention allocation, mediator and outcome, and will control for site and for mediator and outcome values at baseline.

Indirect effects will be estimated via percentile-based confidence intervals derived from bootstrapping to account for asymmetries in the confidence interval induced by multiplying two regression terms. Separate indirect treatment effects will be reported for each intervention compared to TAU.

Exploratory analyses. We will undertake three sets of exploratory analyses: Triple P vs COSP, analyses for the CTSPC-R, and trajectory-based analyses.

Triple P vs COSP. For all outcome models, we will re estimate analyses with COSP as the reference category to estimate

CTSPC-R. Because of COVID-19 restrictions, collection of CTSPC-R data via home visit was severely restricted. Thus, we will analyse this as an exploratory outcome. The CTSPC-R is only collected at the 12-month assessment point. For this outcome, we will estimate a single-level model with terms for intervention arm and study site, using an appropriate outcome distribution and link function.

Trajectory-based models. Given the multiple assessment points in this trial, we will also explore developing trajectory-based estimates of intervention impacts for primary outcomes and child outcomes. After considering the most likely functional form of the time relationship suggested in primary outcome models, we will treat time as continuous and describe the impacts of interventions on outcome trajectories using interaction terms for intervention status by linear or quadratic time trends.