SAFIR FAMILY TALK: a selective primary preventive intervention vs. service as usual for children of parents with mental illness

Statistical Analysis Plan
Version 1.2
11/07/2025

ClinicalTrials.gov ID: NCT05615324

Statistical analysis plan (SAP) for the randomized controlled clinical trial SAFIR FAMILY TALK – version 1.2

Lisbeth Juhl Mikkelsen[1], Signe Sofie Nielsen[1], Anne Thorup, Merete Nordentoft[1], Carsten Hjorthøj[1,2] and Anne Ranning[1,3]

1 Copenhagen Research Center for Mental Health [CORE], Mental Health Services Capital Region of Denmark, Gentofte Hospitalsvej 1, Opgang 15-4, DK-2900
2 University of Copenhagen, Department of Public Health, Section of Epidemiology

3 University of Copenhagen, Department of Psychology

1 Administrative information

1.1 Title and Trial Registration

This SAP is the detailed statistical analysis plan for the randomized clinical trial *"SAFIR FAMILY TALK: a selective primary preventive intervention vs. service as usual for children of parents with mental illness"* (ClinicalTrial.gov identifier: NCT05615324).

1.2. SAP version

This is the second version of the Statistical Analysis Plan (Version 1.2). The original plan was developed under the assumption that 12-month follow-up (12FU) data would not be available in time for the main outcome analyses. However, it later became feasible to include the 12FU data. Consequently, the statistical models were updated to accommodate repeated measures across all three time points (baseline, 4-month, and 12-month follow-up), and this revised version reflects those changes

2 INTRODUCTION: BACKGROUND, RATIONALE AND OBJECTIVES

The background, rationale, and objectives of the study are described thoroughly in the protocol article [1].

3 Study methods

3.1 Randomization

From the protocol[1]: The allocation between the two arms of the study is 1:1. Families who provide written informed consent are randomly allocated to either Family Talk or service as usual. Randomization will be stratified by site using REDCap software. REDCap is an electronic data capture tool hosted at the Center for IT and Medicotechnology (CIMT) in the Capital Region of Denmark. The randomization programme is set up by CH. When the baseline assessment is completed, their contact information is sent to the allocation team that will assign the participants to either intervention or service as usual. The allocation is randomized and computer-generated. The randomization cannot be influenced by the person making it or any other person. Personnel who are not blind to the treatment arm are responsible for the randomization process. When a family has been recruited for the study and baseline assessment has been conducted, the assessor informs the person responsible for conducting the randomization process via e-mail. Randomization is centralized and computerized with a concealed randomization. Block size will be unknown to the researchers and clinicians. The randomized intervention allocation is concealed until the statistical analyses of the resulting data have been completed and conclusions have been drawn. Randomization is carried out by a member of the research team at the Copenhagen Research Center for Mental Health who generates the allocation sequence and assigns participants to interventions.

3.2 Sample size and power calculation

Replicated from the protocol[1]: Power calculation for all three primary outcomes; The Children's Global Assessment Scale (C-GAS)[2], The Parenting Sense of Competence Scale (PSOC)[3], and The McMaster Family Assessment Device (FAD)[4] was performed before the start of inclusion. The power calculations detected that the study needed to include 143 children in each group to be able to measure a difference of 5 points on the C-GAS (a scale from 1-100) between the two groups with a power of 0.90, resulting in an ***n*=286**. Power calculations for the primary outcome PSOC showed that to measure a mean difference of 5 points, obtain a power of 0.90 and an error 1 rate of 0.05, the study needed to enroll a total of *n=135* participants. Power calculations for the primary outcome FAD showed that in order to obtain a power of 0.90 and an error 1 rate of 0.05, a total of *n=168* participants were needed.

3.3 STATISTICAL INTERIM ANALYSES AND STOPPING GUIDANCE

No interim analyses will be performed. We planned no stopping guidance.

3.4 Timing of final analysis

The researcher who will perform the analyses (CH) will be blinded from group allocation until all analyses have been performed and conclusions are drawn. The true randomization group is concealed and named groups A and B, reflecting group allocation in the blinded dataset. A co-worker will perform the randomization variable conversion from another project and will not assist with or perform any analysis.

At the time of publication of this SAP, none of the planned analyses have been performed.

4 Statistical principles

4.1 Confidence intervals and p-values

For all outcomes, the two randomization groups are pairwise compared (TAU vs. Family Talk intervention).

4.2 Analysis population

All analyses are performed as *intention-to-treat* (ITT) unless otherwise stated.

5 Trial population

5.1 Eligibility and recruitment

All potential participants were screened for eligibility criteria before randomization. However, due to the nature of the recruitment methods employed in this trial, it was not feasible to ascertain the exact number of individuals who were screened. Recruitment involved two primary methods: first, healthcare workers in both inpatient and outpatient clinics informed their patients about the study and encouraged them to contact the research team if interested. Second, patients meeting inclusion criteria within the Capital Region were identified through the Danish National Patient Register, and invitation letters were subsequently sent via the national electronic notification system (e-boks). These letters contained detailed information about the study and instructions on how to participate. Given that these approaches relied on patient-initiated responses and external registries, tracking the total number of families assessed for eligibility during the inclusion period was not possible. A detailed description of the recruitment process is available in the study protocol[1]. See CONSORT flow-diagram for screening, randomization, allocation and follow-up (Fig. 1).

The eligibility criteria were:

1. Parent(s) must have at least one ICD-10[5] psychiatric diagnosis by a psychiatrist
2. At least one point of contact with the secondary mental health system within the previous 2 years before the assessment day
3. Have at least one child between the ages of 7 and 17 on the day of the assessment

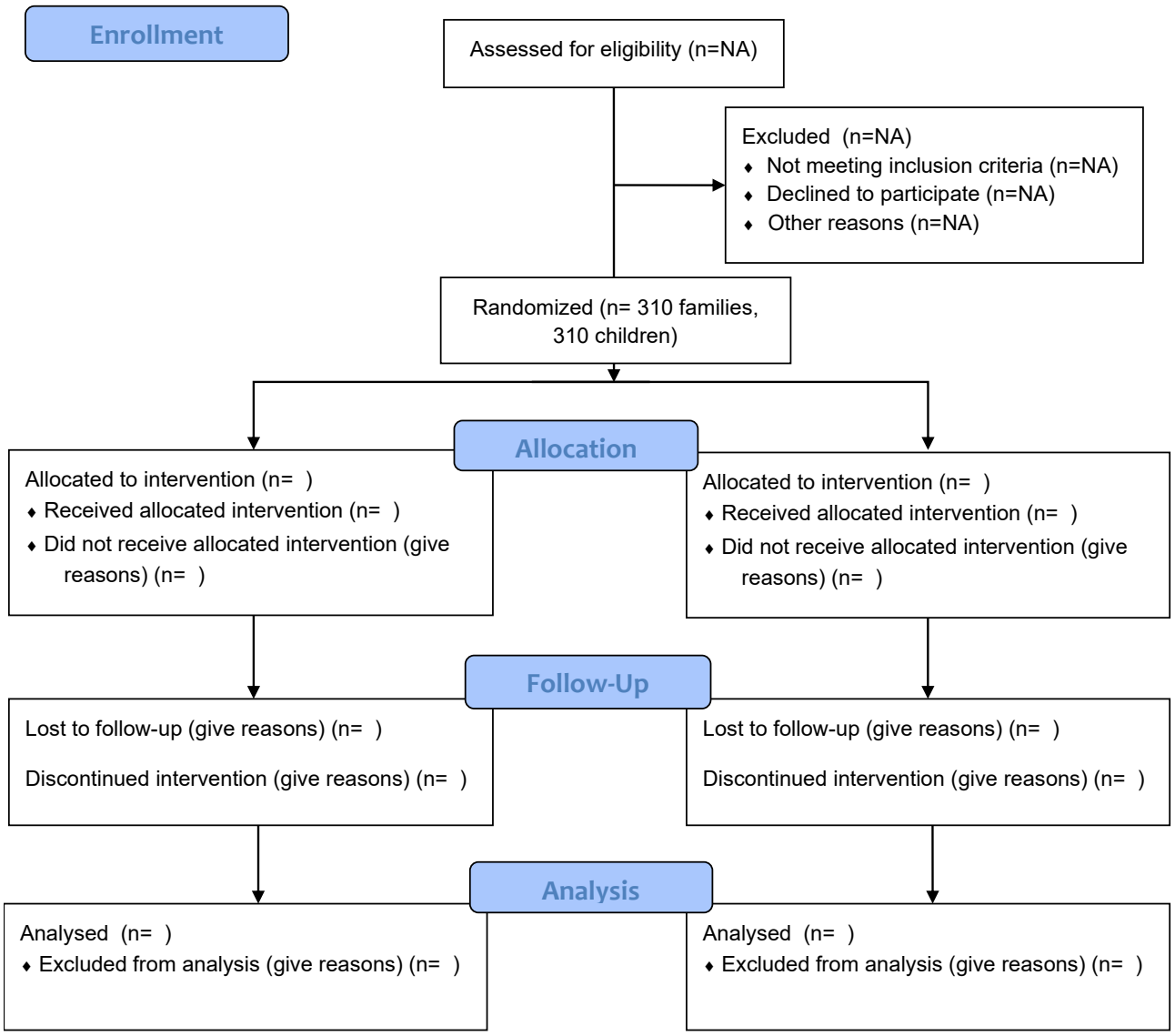
The exclusion criteria were:

1. Non-Danish or non-English speakers

5.2 Withdrawal and follow-up

Withdrawal from the study was continuously registered, and the reason for lost to follow-up was registered when possible.

Fig. 1 CONSORT 2010 Flow Diagram



6 Outcome measures

6.3 primary and secondary outcomes

Table 2. Primary and secondary outcomes

PRIMARY AND SECONDARY OUTCOMES					
Outcome class	**Data Source**	**Outcome**	**Baseline**	**4 month follow-up**	**12 month follow-up**
Primary	Interview w. parent+child and researcher rated	Children Global Assessment Scale (C-GAS)	X	X	X
	Questionnaire (parent)	Parental Sense of Competence (PSOC)	X	X	X
	Questionnaire (parent)	Family Assessment Device (FAD)	X	X	X
Secondary	Questionnaire (child)	Beck's Youth Inventories II (BYI-II)	X	X	X
	Questionnaire (child 7-12)	Parent-child communication (PCCS)	X	X	X
	Questionnaire (child 12-17)	Parent-adolescent communication (PACS)	X	X	X
	Questionnaire (parent)	Change in parental recovery (Brief INSPIRE-O)	X	X	X

6.4 Exploratory outcomes

Table 3. Exploratory outcomes

Outcome class	**Data Source**	**Outcome**	**Baseline**	**4 month follow-up**	**12 month follow-up**
Exploratory	Questionnaire (child 7-10)	Strengths and Difficulties Questionnaire (SDQ)	X	X	X
	Questionnaire (child 11+)	Strengths and Difficulties Questionnaire (SDQ)	X	X	X
	Questionnaire (parent)	Strengths and Difficulties Questionnaire (SDQ)	X	X	X
	Questionnaire (parent)	Parent-adolescent communication (PACS)	X	X	X
	Questionnaire (child)	Child and Youth Resilience Measurement (CYRM)	X	X	X
	Questionnaire (child)	Guilt and Shame Questionnaire (GSQ-AMPI)	X	X	X
	Interview parent and researcher rated	Personal and Social Performances Scale (PSP)	X	X	X
	Interview w. parent and researcher rated	Global Assessment of Functioning – Symptoms (GAF-S)	X	X	X
	Questionnaire (child)	Family Talk Evaluation Questionnaire		X	
	Questionnaire (parent)	Family Talk Evaluation Questionnaire		X	
	Questionnaire (parent)	Working Alliance Inventory Short (WAI-S)		X	
	Questionnaire (parent)	Client Satisfaction Questionnaire (CSQ-8)		X	
	Questionnaire (parent)	Negative Effects Questionnaire (NeQ)		X	

	Danish registers and self-report from parents	Child's number of days absent from school	X	X	X
	Questionnaire (children aged 12-17)	Family Assessment Device (FAD)	X	X	X
	Questionnaire (child)	Children of Parents with Mental Illness Questionnaire (COPMI-Q)	X	X	X
	Referrals to CAMHS (Family Talk health professional)	No.		X	
	Notification to social services/child protective services (Family Talk mental health professional)	No.		X	

7 Statistical analyses

This section describes general strategies applying to all analyses unless otherwise specifically stated.

7.1 Hypotheses

We hypothesize that *the Family Talk Intervention group is superior to Treatment As Usual (TAU) in improving all primary outcomes* (see table 2 for list of primary and secondary outcomes).

7.2 Baseline characteristics

The following will be reported, and standard deviations will be reported for all mean values of numeric variables. Distributions for the two allocation groups will be estimated as means (SD) for numerical data and percentages for categorical data.

Table 1. Baseline characteristics

Participant characteristics	**TAU n=**	**Family Talk n=**	**All participants n=**
PMI* gender (f/m/other, %)			
PMI age (mean, year)			
Parental mental illness (ICD-10 diagnosis)			
GAF-S score (mean)			
PSP score (mean)			
Child age (mean)			
Child gender (f/m/other, %)			
Child mental illness (yes/no, %)			
C-GAS score (mean)			
Other Parent gender (f/m/other, %)			
Other Parent age (mean, year)			
Mental illness (yes/no)			

GAF-S score (mean)			
PSP score (mean)			
Parents cohabitation status (together/apart/single parent headed household, living with new partner, %)			
Referral to CAMHS before allocation to TAU or FT (no.)			
Notification to social services before allocation to TAU or FT (no.)			

*PMI = parent with a mental illness

7.3 Primary and Key-Secondary Analyses

7.3.1 Objectives and Estimands

Primary estimand: The adjusted mean difference (Family Talk – TAU) in the outcome at 4month follow-up (4FU), analysed under the intentiontotreat (ITT) principle.

7.3.2 Statistical Model

All continuous outcomes will be analysed with a linear mixed-effects model for repeated measures (LMM; MMRM). The model will be fitted by restricted maximum likelihood (REML) and will include: Fixed effects for Treatment group (Family Talk vs. TAU), categorical time (baseline, 4FU, 12FU) and their Group × Time interaction.
A fixed effect for Recruitment site (stratification factor).
A random intercept for participant to account for intra-individual correlation.
The unstructured covariance matrix will be used initially; if the model fails to converge, the structure with the lowest BIC (e.g. heterogeneous compound symmetry) will be adopted.
For binary outcomes the same fixed-effect structure will be fitted using a generalised linear mixed-effects model (GLMM) with logit link and random participant intercept.
Baseline measurements are treated as *Time = Baseline* within the repeated-measures factor; no separate baseline covariate is added. If an outcome is recorded only post-baseline, baseline is included as a covariate instead.

7.3.4 Estimates and Reporting

We will report least-squares means (LS-means) by group at each post-baseline time point, the adjusted mean difference, 95 % confidence intervals and two-sided p-values. We will provide model-based plots of estimated mean trajectories with 95 % CI ribbons. For binary outcomes report odds ratios with 95 % CI.

Model assumptions will be checked via residual and Q–Q plots; if heteroscedasticity is present, robust (sandwich) standard errors will be reported.

7.3.5 Handling of Missing Data

Missing data will be handled according to intention-to-treat (ITT) principles. Multiple imputation (MI) using chained equations (MICE) will be performed to impute missing data prior to fitting linear or generalized linear mixed models (LMM/GLMM). Imputation models will incorporate baseline covariates, stratification variables, and additional variables identified as independent predictors of outcomes or predictors of

missingness (P < 0.05 in univariate analyses). A total of 100 imputed datasets will be generated using 20 iterations each, and the analyses will be conducted by pooling results across these datasets according to Rubin's rules.

Due to an error in the survey setup in REDCap, item no. 83 in the secondary outcome Beck Youth Inventories II (BYI-II)[6], was never included in the survey. As a result, no participants responded to this item at baseline or at the 4- and 12-month follow-up. According to the survey manual, if an item is missing, the clinician should calculate the mean of the subscale from which the item is missing and impute this value into the survey. This method is applicable only if no more than two items are missing in total. Following the manual's guidance, we will calculate an individual mean for each participant and impute this value into the total score of the survey. If more than two items are missing, we will follow the previous stated rules for missing data.

Due to an error in the survey in REDCap, item nr. 20 in the secondary outcome PACS[7], was never included in the survey given to participants. As a result, no participants responded to this item at any timepoint. We found no guidance from the authors of the survey on how to handle missing data, so we decided to handle it in the same way as BYI-II, and a mean score from the subscale will be calculated and imputed as the missing value for each respondent. If more than two items are missing we will follow the previous stated rules for missing data.

Due to an error in the survey setup, two different versions of the Child and Youth Resilience Measurement (CYRM)[8] were administered depending on child age. Children below the age of 12 received a version with a 3-point Likert response scale, while children aged 12 and above received a version with a 5-point Likert response scale. To enable combined analysis across all children, responses from the 5-point scale will be recoded to approximate the 3-point scale as follows: responses 1–2 will be recoded as 1, response 3 as 2, and responses 4–5 as 3. This recoding aims to retain the ordinal nature of the data while ensuring comparability between age groups.

Pattern-mixture / delta approach – We will shift imputed values by ± 0.5 SD to create "best-case" and "worst-case" scenarios, providing bounds on the treatment effect.

Concordant inference across these analyses will strengthen confidence in the mixed-model results; any discrepancies will be highlighted in the results section.

7.3.6 Software

Analyses will be performed Stata (commands *mixed*, *melogit*).

7.4 Timing of Final Analysis — Clarification

The final locked analysis dataset will be created once the last participant has completed the 12-month follow-up assessment. The primary endpoint, however, remains the 4-month contrast defined in Section 7.3. Interim analyses are not planned.

7.5 Subgroup Analyses — Unchanged Methods, LMM Specification

Subgroup effects will be investigated with the same LMM/GLMM framework described previously, adding a fixed interaction term Treatment × Subgroup and retaining the random participant intercept. The predefined subgroups remain:

Parental psychiatric symptom level (parent with lowest score) (GAF-S): <50 vs ≥50.

Parental social functioning (from parent with lowest score) (PSP): <50 vs ≥50.

Child daily functioning (C-GAS): <60 vs ≥60.

A significant interaction ($p < 0.05$, two-sided) will be interpreted as evidence that treatment effects differ by subgroup. Results will be presented as adjusted mean differences/odds ratios within each subgroup with 95 % CI.

7.6 Sensitivity Analyses — Robustness of Primary Conclusions

We will repeat the primary mixed-model analysis under the following scenarios to assess robustness:
Baseline-imbalance adjustment – add any baseline covariates that differ between groups at $p < 0.10$.
Per-protocol population – restrict to families who attended the Family Talk modules including the Family meeting (intervention) compared to same comparison group as in other analyses (SAU).
All sensitivity models will mirror the fixed-effect structure of Section 7.3. Point estimates, 95 % CI and *p*-values will be contrasted with the primary ITT results.

7.7 Exploratory Analyses — Predictors of Engagement in Family Talk

Within the Family Talk arm only, we will explore whether parental PSP or GAF-S at baseline predict:
Drop-out of treatment (yes/no) — logistic regression.
Session attendance (count or proportion) — Poisson or linear regression as appropriate.
Covariates: parent age, child age and sex, site. These results are descriptive and hypothesis-generating.

7.8 Exclusion of outcome measures

7.8.1 Children's response to parental mood

The measure *'Children's response to parental mood'* has been excluded from the analyses because the scale cannot be meaningfully summarized into a single sum-score or another continuous measurable indicator. The original development of the instrument was based on a multidimensional profile classification (e.g., active empathy, over-involvement, indifference, and avoidance). This profile-based approach would require extensive psychometric analyses and validation work, which is beyond the scope of the current study. Therefore, we decided to exclude this measure from the pre-specified analyses.
The collected data will, however, be retained, and a separate publication may be considered in the future using the methods originally described by the developers.

7.8.2 KIDSCREEN-27 (Danish version):

Although initially included to assess child-reported health-related quality of life, recent evidence has raised serious concerns about the psychometric validity of the Danish version of KIDSCREEN-27. Specifically, a Danish psychometric expert[10,11] has strongly criticised the instrument's dimensionality, scoring properties, and interpretability, concluding that it cannot be recommended for use in its current form. In light of these concerns, we have decided to exclude the KIDSCREEN-27 from all planned analyses.

8. Bibliography

1. Nielsen, S. S. *et al.* A study protocol for the randomized controlled trial SAFIR FAMILY TALK: a selective primary preventive intervention vs. service as usual for children of parents with mental illness. *Trials* **24**, (2023).
2. Shaffer, D. *et al.* A Children's Global Assessment Scale (CGAS). *Arch Gen Psychiatry* **40**, 1228–1231 (1983).

3. SELF-ESTEEM AND SITUATIONAL STRESS: FACTORS RELATED TO SENSE OF COMPETENCE IN NEW PARENTS. - ProQuest. https://www.proquest.com/docview/287957227.

4. Miller, I. W., Epstein, N. B., Bishop, D. S. & Keitner, G. I. THE McMASTER FAMILY ASSESSMENT DEVICE: RELIABILITY AND VALIDITY. *J Marital Fam Ther* **11**, 345–356 (1985).

5. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision ICD-10 : Tabular List. *World Health Organization* **1**, 1–1080 (2016).

6. Thastum, M., Ravn, K., Sommer, S. & Trillingsgaard, A. Reliability, validity and normative data for the Danish Beck Youth Inventories: Development and Aging. *Scand J Psychol* **50**, 47–54 (2009).

7. Barnes, H. L. & Olson, D. L. Parent-adolescent communication scale. in *Family inventories: Inventories used in a national survey of families across the family life cycle* (ed. D. H. Olson et al.) 33–48 (St Paul: Family Social Science, University of Minnesota, 1982).

8. Ungar, M. *et al.* The Study of Youth Resilience Across Cultures: Lessons from a Pilot Study of Measurement Development. *Res Hum Dev* **5**, 166–180 (2008).

9. Solantaus-Simula, T., Punamäki, R. L. & Beardslee, W. R. Children's Responses to Low Parental Mood. I: Balancing Between Active Empathy, Overinvolvement, Indifference, and Avoidance. *J Am Acad Child Adolesc Psychiatry* **41**, 278–286 (2002).

10. Tine Nielsen. Tine Nielsens 10 forbehold i forhold til fremtidig anvendelse af KIDSCREEN-27 i Danmark. *UCL: Erhvervskole og professionshøjskole* (2025).

11. Nielsen, T. *et al.* Validering af den danske udgave af KIDSCREEN-27 børn og unge selv-rapporteringsversion. Psykometriske egenskaber undersøgt ved hjælp af Rasch-analyser: Projektet "Validering af KIDSCREEN-27 (Tillægsundersøgelse ifm. Børns Vilkårs undersøgelse af skolebørns livskvalitet)" er gennemført af Tine Nielsen, UCL Erhvervsakademi og Professionshøjskole, Afdelingen for Anvendt Forskning i Pædagogik og Samfund, med datagrundlag fra Børns Vilkår og fondsstøtte fra Ole Kirk's Fond. *PLoS One* **18**, (2024).