

Statistical Analysis Plan

The OPUS YOUNG Trial

Date: April 22, 2026

Title

Early intervention versus treatment as usual for adolescents with first-episode psychosis: Statistical analysis plan for the randomized controlled OPUS YOUNG trial

Registration

ClinicalTrials.gov: NCT04916626, (date of registration: 03-04-2021)

The Ethics Committee of Capital Region: H-20064085 (date of approval: 20-04-2021)

Danish Data Protection Agency: P-2020-1081 (date of approval: 20-11-2020)

Protocol Version

Version 6.1 July 2023

Roles and responsibilities

Project Leader: Anne Katrine Pagsberg, PhD, Clinical Professor

SAP Collaborators/authors: Mathilde Søndergaard Madsen, Anne Katrine Pagsberg, Naja Kirstine Andersen, Caroline Friis Nielsen, Martin Køster Røimvall, Natascha Larsen, Laura Pipaluk Pold Christensen, Thomas Lind Andersen, Marianne Melau, Merete Nordentoft, Jens Richardt M. Jepsen, Anne A.E. Thorup, Pia Jeppesen, Birgitte Fagerlund, Patrick McGorry, Swaran Singh, Ditte Lammers Vernal, Lene Halling Hastrup, Lis Raabæk Olsen, Jacob Rydkjær, Christoph U. Correll, Tobias Haugegaard, Robin Christensen.

Senior Biostatistician: Robin Christensen, BSc, MSc, PhD; Professor

Statistical Analyst: Tobias Haugegaard, MSc; Biostats Fellow

INTRODUCTION

Background

Mental health disorders are a leading cause of disease burden in youth worldwide[1]. Severe mental disorders can have substantial negative impact on many aspects of daily life and personal and social functioning. Psychotic disorders are associated with shortened life-expectancy[2] poorer educational attainment, unemployment, and psychiatric and somatic comorbidities[3], [4]. Furthermore, providing care for youths with psychosis often entails large psychological, social, and economic implications for the affected families[5]. Early Intervention Services (EIS) is an evidence-based treatment that has shown effective and feasible results in young adults with psychosis[6], and high quality research has shown superiority of EIS compared to standard care in young adults (18-35 years) with FEP regarding both symptoms, social functioning, treatment adherence, hospitalization, recovery and cost-effectiveness[7], but no study to date has examined the effect of a similar comprehensive and intensive psychosocial treatment in a population consisting exclusively of children and adolescents below age 18 years.

Aims and Objectives

The aim of the OPUS YOUNG trial is to examine the treatment outcomes for children and adolescents with first-episode psychosis (FEP) by investigating the effects of an EIS developed for this population. In a randomized controlled trial, we compare the effects of two years treatment with the OPUS YOUNG EIS intervention versus treatment as usual (TAU) in adolescents aged 12-17 years (both inclusive) with FEP. The hypothesis is that the OPUS YOUNG intervention has superior beneficial effect compared to TAU without posing an increased risk of harms.

The primary objective is to compare the effects of the OPUS YOUNG intervention relative to TAU on change in level of personal and social functioning after 24 months of treatment. Key secondary objectives are to compare the effects of the OPUS YOUNG intervention relative to TAU on the following secondary outcomes after 24 months of treatment: severity of psychotic-, negative-, and disorganized symptom dimensions; client satisfaction; and health-related quality of life. The hypothesis for the secondary outcomes is that the severity of psychosis psychopathology, satisfaction with treatment, and health-related quality of life will show superiority of the OPUS YOUNG intervention compared to TAU.

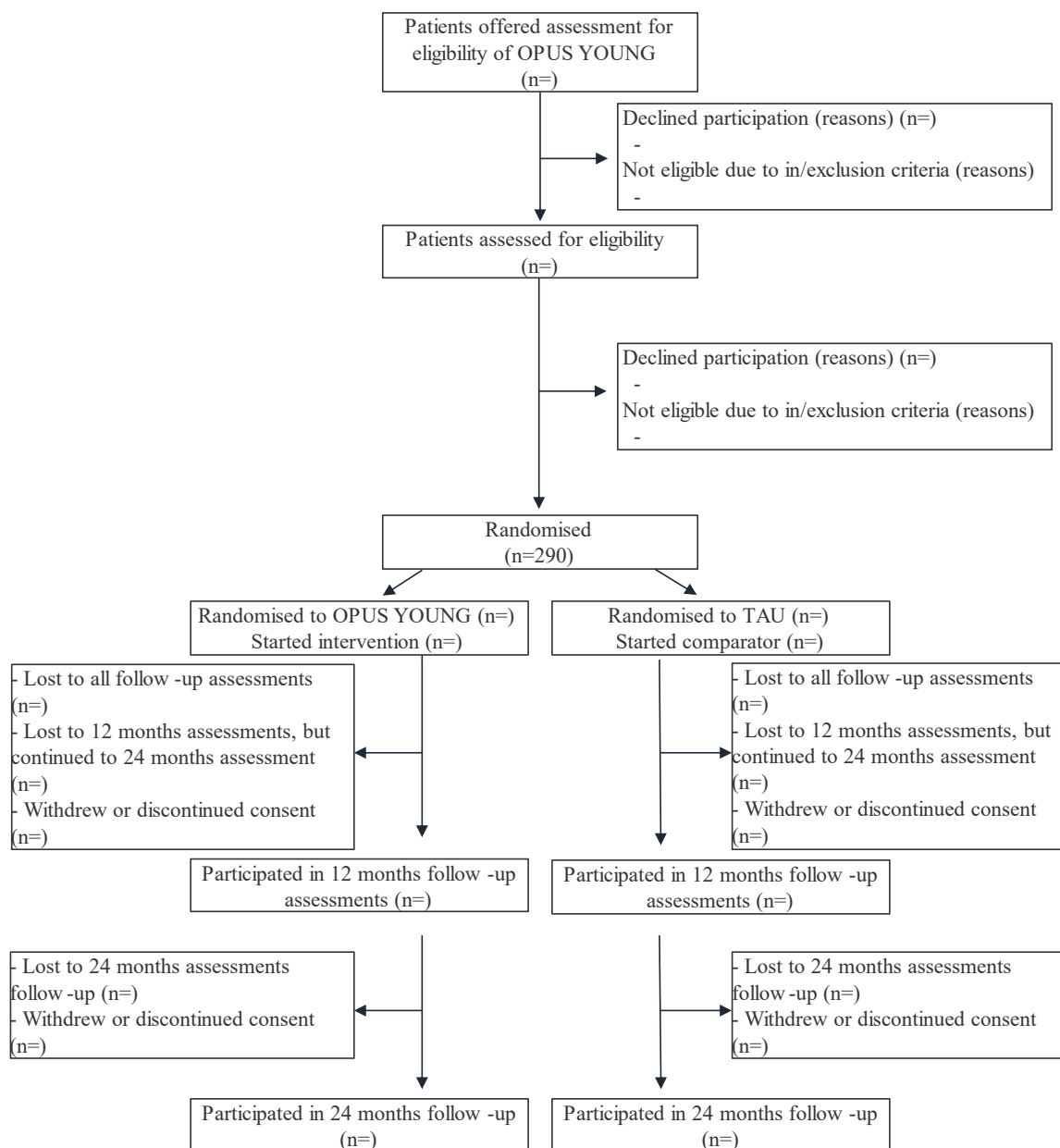
STUDY METHODS

Trial Design and Interventions

The OPUS YOUNG trial is an investigator-initiated, single-center, pragmatic RCT with blinded outcome assessments. We randomized 290 participants aged 12-17 years with FEP 1:1 to a two-year intervention with OPUS YOUNG versus TAU. We assessed treatment effects and potential harms after 12 and 24 months. 24 months was treatment endpoint, see trial flow in Figure 1. The trial protocol followed the SPIRIT recommendations[8]. Furthermore, the trial procedures follow the Good Clinical

Practice Guideline[9] principles. No commercial interests were involved in the trial. No forms of concomitant care were prohibited during the trial.

Figure 1. Trial flow, OPUS YOUNG trial



Experimental Intervention: OPUS YOUNG

The OPUS YOUNG intervention builds on OPUS, the Danish evidence-based EIS for young adults (aged 18-35 years) with FEP [10]. The OPUS YOUNG intervention is a manualized treatment and consists of the following elements: assertive community treatment, Cognitive Behavioral Case Management (CBCM) [11], psycho-educational family treatment including multiple family groups[12] and psychoeducational siblings groups[13], Social Cognition and Interaction Training (SCIT)[14], individual cognitive behavioral therapy (CBT) [11], and manual-based psychopharmacologic treatment building on the Recovery After an Initial Schizophrenia Episode – Early Treatment Program (RAISE-ETP) principles[15] and national Danish guidelines[16]. To further meet the common challenges for this patient group, the intervention is focused on 1) special transition support to bridge the gap from adolescent psychiatry and social care to adult services[17]; 2) individual school/employment support[18]; 3) prevention and treatment of substance abuse combining CBT and motivational enhancement therapy – the MOVE intervention[19]. Moreover, the OPUS YOUNG intervention follows the NICE guidelines concerning management of substance misuse[20], and actively intervenes to prevent initiation/continuation of tobacco smoking.

The intervention is delivered by a specialized multidisciplinary team of psychologists, nurses, social workers, physiotherapists, occupational therapists, and specialist consultants in child and adolescent psychiatry, with a caseload of 1 staff member to 10 patients. The team provides assertive outreach and individualized case management. In addition, the team adapts to the patient needs that arise during the treatment period[21].

We will assess the manual fidelity of the OPUS YOUNG treatment and structure during the trial period, using a special OPUS YOUNG modified version of The Danish Fidelity Scale for Specialized Early Intervention Team twice during the trial period[21], [22]. To lower the possibility of idiosyncratic judgement, two assessors have independently conducted the fidelity assessment, and their interrater reliability will be reported. Each component in the fidelity scale can be rated either 1 or 0 (1 = fulfilling the requirements for the components and 0 = the requirements were not met). The maximum score is a total of 20 points. Based on the literature [21], [22], [23], we chose five of the components to be of particular significance and mandatory for the concept of the Danish EIS, they are: 1) Independent management, 2) Multidisciplinary teams, 3) Low patient to case manager ratio, 4) Assertive outreach including home visits, and 5) Systematic engagement of family and relatives. Satisfactory fulfilment of the scale criteria for program-fidelity can be obtained at two levels. Elite level: Total score between 20-18, and presence of all five of the mandatory components. Adequate level: Total score of 17, and, at least, four of the five mandatory components present. The result of fidelity will be reported but is not an outcome.

Control Comparator: Treatment As Usual (TAU)

According to the Specialty Plan for Child and Adolescent Psychiatry by the Danish Health Authorities, 2019 [24], children and adolescents with psychosis are referred to a predefined level of care in the Danish health care system according to the severity and complexity of their psychotic illness. In essence, the individual patient is referred to the unit in child and adolescent mental health services (CAMHS)

that best fits the indicated level of care. The treatment follows national and local Danish guidelines but is not manualized. None of the levels of care in TAU are currently assertive as a standard but assertive approaches are in some cases offered after individual clinical evaluation. The unit responsible for the highly specialized function (level 3) in CAMHS manages the patients with the most severe and complicated psychotic illness and all cases of childhood-onset schizophrenia (onset < 12 years). In addition to standard treatment, this unit is specialized by providing a multidisciplinary team, case-management (no defined upper caseload), family groups and psychoeducation, in addition to psychopharmacological treatment following national guidelines. When possible, the unit offers counselling, social skills training adapted to illness severity and motivation of the patient, and CBT may be offered in relation to comorbidity. In general, office visits take place in outpatient clinics striving for a minimum of one visit per month. Depending on level of functioning and patient preference, patients with other ICD-10[25] F2 psychosis diagnoses than schizophrenia can be referred to treatment at general (level 1) or regional (level 2) functions. General and regional functions are both at a less intensive level of care than the highly specialized function. Standard treatment usually consists of elements of psychoeducation, antipsychotic treatment, and support to the families.

Randomization and Allocation Concealment

Participants in the study were randomized 1:1 to the experimental intervention OPUS YOUNG or TAU after eligibility criteria were confirmed, and baseline assessment had been performed. Randomization was centralized and computerized with a concealed randomization sequence carried out by the electronic data capture tool REDCap (Research Electronic Data Capture) [26] hosted by the Capital Region of Denmark. Sequence and block size was unknown to the principal investigator, investigators, and clinicians. To optimize comparability between treatment groups, randomization was stratified by age (12–14 vs. 15–17 years) and the primary outcome measure, i.e., level of social functioning (Personal and Social Performance Scale [PSP] total score 1–44 vs. 45–100).

Power and Sample Size Considerations

Based on the literature concerning the level of social dysfunction measured with PSP in young adults and adolescents with early onset psychosis (EOP)[27], [28], [29], [30], meaning onset before the age of 18 years, we expect a Standard Deviation (SD) on PSP around SD=13. There is currently no validated published minimal important difference (MID) for the PSP scale used specifically in adolescents. For adults, a 7-point threshold is considered a provisional benchmark [31]. Concerning the expectations on relevant group differences in improvement during two years of EIS [6], [10], we have evaluated the MCID. A metaanalysis on EIS interventions in seven studies including n=1005 participants aged 16–65 years (mean 27.5 years) examined functioning as outcome, measured with the Global Assessment of Functioning (GAF, scaled 1–100) [6]. Here a standardized mean difference (difference in mean outcome between groups divided by standard deviation in outcome among participants) of 0.21 (in the individual seven studies included ranging from 0.02 to 0.56) was found. With an SD of 13, this equals an expected difference in PSP scores between the intervention groups of around 2.73, ranging from 0.26 to 7.28. Based on these findings and evaluation of what is clinically meaningful when regarding the PSP 10-point intervals of functioning, a minimal difference of five points in outcome score at end-of-treatment on the total PSP scale between the two interventions will be considered clinically relevant. We consider

this difference meaningful since both intervention groups are in active treatment, and both, if assessed eligible, are exposed to pharmacological treatment.

An a priori sample size calculation was performed using a two-sided, two-sample t-test to detect a mean difference of five units with a standard deviation of 13, assuming equal group sizes and a significance level of 0.05. To achieve statistical power of 90%, 85%, and 80%, the required total sample sizes were 288, 246, and 216 participants, respectively. To ensure adequate power, the target sample size was set at 290 participants (145 per group), with a minimum enrolment goal of 284. At this sample size, the estimated power to detect the specified difference was 90.4%. Calculations were conducted using the POWER procedure in SAS (SAS Institute Inc).

Hypothesis testing framework

The study examines superiority of the intervention OPUS YOUNG compared to TAU regarding improvement in personal and social functioning and key secondary outcomes of severity of psychosis psychopathology, satisfaction with treatment, and health-related quality of life. The hypotheses will be tested in hierarchical order using gatekeeping principles.

Statistical Interim Analyses and Stopping Guidance

An interim analysis to check for clear early superior efficacy, futility or safety issues enabling timely decisions to stop or modify the trial, was evaluated not to be needed during the trial due to the intervention being composed of several evidence-based elements all considered safe. Hence, the trial had no prespecified guidelines for determining stopping rules due to safety concerns.

Outcomes

Timing of the outcomes assessments

The last participant's 24-month assessment took place on the September 11th, 2025. The analyses based on primary and key secondary repeated measures from baseline to the 24-month assessment, comparing OPUS YOUNG and TAU, are anticipated to begin in Spring 2026. The 12-month follow-up assessment data will be prepared and analyzed in the same process but will be conducted as a secondary analysis in the main publication. See BOX 1 for timing of the outcome assessments.

BOX 1: Timing of Outcome Assessments

Variables	Time in months		
	0	12	24*
Primary outcome			
Social function: Personal and Social Performance Scale (PSP)	x	x	x
Key secondary outcomes			
Psychotic dimension: Scale for Assessing Positive Symptoms (SAPS)	x	x	x
Negative dimension: Scale for Assessing Negative Symptoms (SANS)	x	x	x
Disorganized dimension: Combination of items from SANS and SAPS	x	x	x
Client satisfaction: Client satisfaction questionnaire (CSQ)		x	x
Health-related quality of life: Health-Related Quality of Life Questionnaire for Children and Young People and their Parents (KIDSCREEN-10)	x	x	x

Note. *Primary Treatment Endpoint.

Data analysis for other outcomes and 30-month follow-up will be performed at later stages and be addressed in separate publications.

Outcome measures

The primary outcome of personal and social functioning is measured with the Personal and Social Performance Scale (PSP)[32] after 24 months of treatment. Through an interview with the participant PSP examines social functioning within the last month. Based on the evaluation of the four subdomains Socially useful activities, Personal and social relationships, Self-care, and Disturbing and aggressive behavior, the scale provides both a total score and sub-scores for each subdomain. The total score (range 1-100) is the primary outcome.

Key secondary outcomes include severity of psychotic, negative, and disorganized symptom dimensions measured with Scale for Assessing Positive Symptoms (SAPS)[33] and Scale for Assessing Negative Symptoms (SANS)[34]; client satisfaction with care measured with Client Satisfaction Questionnaire (CSQ)[35]; and health-related quality of life measured with Health-Related Quality of Life Questionnaire for Children and Young People and their Parents (KIDSCREEN-10)[36] for detailed description of outcomes measures see BOX 2. Since initiation of the trial, a study on the psychometric properties of the Danish instrument of KIDSCREEN-10 has been published [37]. The study concludes that the Danish version of KIDSCREEN-10 was unfit for population-level studies and small-scale intervention studies. This makes the examination of the results of superiority on this outcome more questionable, and results should be interpreted with caution.

Table 1 presents an overview of variables for stratification, demographic characteristics, psychopathology, and the primary and secondary outcomes collected at baseline and their distribution in the two intervention groups.

Table 1. Baseline Characteristics of the Participants in the Intention to Treat (ITT) Population

Variable	OPUS YOUNG N=?	TAU N=?	Total N= 290
Stratification variables			
Low age (12-14 years), n (%)			
Low PSP total score (1-44), n (%)			
Demographics and Psychopathology			
Sex, female, n (%)			
Gender, cis-gender, n (%)			
Age, years			
Living in institution, n (%)			
Siblings in the household, n (%)			
Attending school/work/special school/full-time/part-time, n (%)			
Days of absence from school/work/special school, last three months, days			
DUP, weeks			
Full scale IQ (range 50-160)			
Family predisposition first-degree relatives (schizophrenia/other psychosis), n (%)			
Stressful life events with current negative effects (lifetime), n (%)			
Exposed to traumatic life event (lifetime), n (%)			
Inclusion diagnosis (ICD-10)			
- Substance-Induced psychosis (F1X.5), n (%)			
- Schizophrenia (F20.X), n (%)			
- Schizotypal disorder (F21), n (%)			
- Schizoaffective disorder (F25.x), n (%)			
- Other/unspecified non-organic psychosis (F28/F29), n (%)			
- Severe depression with psychotic symptoms (F32.3), n (%)			
Psychiatric comorbidities*, n (%)			
- 0			
- 1			
- ≥ 2			
Comorbid autism spectrum disorder, n (%)			
Chronic somatic illness (asthma, diabetes, eczema, epilepsy, other), n (%)			
In antipsychotic treatment, n (%)			
Outcome measures			
<i>Primary outcome measure</i>			
PSP total score, range 0 – 100			
<i>Key secondary outcome measure</i>			
Psychotic symptom dimension (SAPS); range 0 – 5			
Negative symptom dimension (SANS); range 0 – 5			
Disorganized symptom dimension; range 0 – 5			
KIDSCREEN-10, T-score, range 1-100			

Notes. Values will be reported as Mean and Standard Deviation unless otherwise indicated. Abbreviations: PSP: Personal and Social Performance Scale, DUP: Duration of Untreated Psychosis, IQ: Intelligence Quotient. ICD-10: International Classification of Diseases, 10th Revision. KIDSCREEN-10: Health-Related Quality of Life Questionnaire for Children and Young People and their Parents. *including autism spectrum disorders.

Table 2 presents an overview of the primary and key secondary outcomes used in the main analysis and furthermore shows the defined treatment response and clinical remission measures.

Table 2. The primary and key secondary outcomes at 24 months (The ITT Population)

End point	OPUS YOUNG N=?	TAU N=?	Difference between groups (95% CI)	P Value
Primary outcome measure				
Change in social function (PSP), range 1-100				
Key secondary outcome measure				
Change in psychotic dimension (SAPS), range 0-5				
Change in negative dimension (SANS), range 0-5				
Change in disorganized dimension (SAPS/SANS), range 0-5				
Client satisfaction (CSQ), range 8-32				
Change in health-related quality of life (KIDSCREEN-10), T-score, range 1-100				
Treatment response and remission				
Response (CGI-I) ^a , n (%)				
Clinical remission (SANS/SAPS) ^b , n (%)				
Antipsychotic treatment				
Prescribed antipsychotic medication, n (%)				

Notes. Results will be reported as least-squares means with standard errors, and between-group differences will be expressed with 95% confidence intervals. Responder outcomes will be analyzed using logistic regression and reported as odds ratios with 95% confidence intervals. Missing continuous outcome data will be handled using mixed-effects models for repeated measures, which accommodate data missing at random. Missing data for responder outcomes will be handled conservatively using non-responder imputation.

Abbreviations: PSP: Personal and Social Performance Scale, SAPS: Scale for Assessing Positive Symptoms SANS: Scale for Assessing Negative Symptoms, CSQ: Client Satisfaction Questionnaire, KIDSCREEN-10: Health-Related Quality of Life Questionnaire for Children and Young People and Their Parents, CGI-I: The Clinical Global Impression Scale–Improvement.

^a: Treatment response/non-response at 24 months is defined as a CGI-I score of 1, 2 or 3 (range from “very much improved” to “minimally improved”) and CGI-I scores from 4-7 (range from “no change” to “very much worse”) respectively.

^b: Clinical remission at 24 months is defined as a global score of maximum 2 (mild) on all global ratings of symptoms on the positive, negative, and disorganized dimensions (SAPS/SANS) within the last three months[38], [39].

Measures of potential harms

We will report and compare potential harms in each study arm during the trial period. See Table 3: Harms experienced from baseline to 24 months, for an overview of harms outcomes.

Serious adverse events (SAE) are any untoward medical occurrence in a clinical trial participant which does not necessarily have a causal relationship with the study interventions. SAEs are events that (according to ICH Harmonized Tripartite Guideline [9]): 1) result in death, 2) are life-threatening, 3) require inpatient hospitalization or prolongation of existing hospitalization, 4) result in persistent or significant disability or incapacity, 5) result in congenital anomaly/birth defect (in a child of a participant exposed during pregnancy), or 6) qualifies as “other serious important medical events”, i.e.

that may jeopardize the patient or may require intervention to prevent one of the other five outcomes (1 to 5).

Participants are registered as research participants in the digital medical record system, and reports of potential SAEs are sent automatically by the system to the project coordinator when registered in the medical records. SAEs are then systematically collected and continuously reported to The Ethics Committee of the Capital Region of Denmark during the 24-month treatment period. We will categorize the SAEs in two steps. First, we will categorize each event according to the six ICH definitions stated above. As a second step, we will further categorize in more detail the study defined clinical event type associated with the SAE: 1) non-suicidal self-injury (NSSI), 2) aggravated psychiatric symptomatology; 3) suicidal ideation (thoughts/impulses), 4) suicide attempt, 5) homicidal thoughts/impulses, and 6) other.

In addition to reporting the untoward events described above we also collect data on patient-rated suicidal ideation from baseline to 24 months assessed by interviews at baseline and at 12- and 24-month assessments using the semi-structured interview Kiddie Schedule for Affective Disorders and Schizophrenia – Present and Lifetime Version (K-SADS-PL)[40] item on Suicidal Ideation. Scores of both subthreshold “*occasional thoughts of suicide*” and threshold “*often thinks of suicide*” is categorized as suicidal ideation. Furthermore, patient-reported suicide attempts and self-harm from baseline to 24 months is assessed using the K-SADS-PL item on Suicide Acts and Self-harm at the three timepoints. Threshold scores based on actions with *definite suicidal intent* is categorized as suicide attempts. Number of attempts during the trial period will be divided into 0, 1-3, ≥ 4 attempts, or number of attempts unknown in a patient who has had at least one suicide attempt over the complete trial period. Any case of self-reported self-harm during the complete trial period is also reported.

Since psychotherapeutic approaches such as CBCM is part of the treatment offered in the OPUS YOUNG intervention, and CBT often is a part of treatment in both the OPUS YOUNG intervention and TAU (see description of interventions in Study Methods) we will report negative effects of psychological treatment using the Negative Effects Questionnaire (NEQ) [41]. The NEQ self-report is divided into six factors (symptoms, quality, dependency, stigma, hopelessness, and failure) containing a total of 32 items that are scored on a five-point Likert-scale (0-4) with higher scores meaning more negative impact. NEQ differentiates between negative effects that are attributed to treatment and those possibly caused by other circumstances, as well as one open-ended question. We will use the mean score and SD for the NEQ from both participants and parents for the 24 months treatment period. For parents’ responses about their child’s negative effects, we will report a composite score if both parents have filled out the questionnaire.

Information on side-effects to psychotropic medication is collected at baseline, 12- and 24-month assessments with the UKU Side Effect Rating Scale[42] focusing on the experience of specific side-effects within the past three days, except for reports on weight gain and weight loss which are based on the past month, epileptic seizures are based on the last six months and menstrual cycle related bleeding disorders are based on the last three months. The UKU is divided into four categories of typical side-effects of antipsychotic medication: psychological, neurological, autonomous, and

other. We will report information on these categories for all participants receiving psychotropic medication except for melatonin: items that score at least 1(mild) on UKU and possible (related to psychotropic medication) or probable (related to psychotropic medication) will count as a report of side-effects within one of the four categories.

Table 3. Harms experienced from baseline to 24 months

	OPUS YOUNG N=?	TAU N=?	Difference (95%CI) N=?
SAE categories (ICH definitions):			
SAEs total, n (%)			
Deaths, n (%)			
Life-threatening, n (%)			
Inpatient psychiatric hospitalization (>24 hours) or prolongation of existing hospitalization, n (%)			
Inpatient somatic hospitalization (>24 hours) or prolongation of existing hospitalization, n (%)			
Persistent or significant disability or incapacity, n (%)			
Congenital anomaly/birth defect, n (%)			
Other serious important events, n (%)			
SAEs study defined clinical categories:			
Non-suicidal self-injury (NSSI), n s (%)			
Aggravated psychiatric symptomatology, n (%)			
Suicidal ideation (thoughts/impulses), n (%)			
Suicide attempt, n (%)			
Homicidal thoughts/impulses, n (%)			
Other, n (%)			
SAE categories (ICH definitions), number of events:			
SAEs total, events (rate)			
Life-threatening, events (rate)			
Inpatient psychiatric hospitalization (>24 hours) or prolongation of existing hospitalization, events (rate)			
Inpatient somatic hospitalization (>24 hours) or prolongation of existing hospitalization, events (rate)			
Persistent or significant disability or incapacity, events (rate)			
Congenital anomaly/birth defect, events (rate)			
Other serious important events, events (rate)			
SAEs study defined clinical categories, number of events:			
Non-suicidal self-injury (NSSI), events (rate)			
Aggravated psychiatric symptomatology, events (rate)			
Suicidal ideation (thoughts/impulses), events (rate)			
Suicide attempt, events (rate)			
Homicidal thoughts/impulses, events (rate)			
Other, events (rate)			
Distribution of number of SAE events among participants:			
0 events, n (%)			
1-5 events, n (%)			
6-10 events, n (%)			
≥11 events, n (%)			

Table 3. Harms experienced from baseline to 24 months

	OPUS YOUNG N=?	TAU N=?	Difference (95%CI) N=?
Number of participants with suicidal and NSSI behavior (K-SADS-PL) baseline to 24 months*:			
Suicidal ideation, n (%)			
Suicide attempt, n (%)			
0 attempts, n (%)			
1-3 attempts, n (%)			
≥4 attempts, n (%)			
no. of attempts unknown, n (%)			
NSSI behavior, n (%)			
Negative effects of psychotherapy (NEQ) baseline to 24 months			
Negative effects of psychological treatment (NEQ), patient-rated, mean score (SD)			
Negative effects of psychological treatment (NEQ), parent-rated, mean score (SD)			
Number of patients with side-effects to psychopharmacological treatment (UKU). At baseline, 12 months follow-up, and/or 24 months follow-up:			
Psychiatric, n (%)			
Neurological, n (%)			
Autonomous, n (%)			
Other, n (%)			
Withdrawals due to SAE's **, n (%):			
Withdrawals due to SAEs, n (%)			

Abbreviations: TAU: treatment as usual; SAE: serious adverse event; K-SADS-PL: Kiddies Schedule for Affective Disorders and Schizophrenia – Present and Lifetime Version.

Notes: *The number of participants with suicidal and NSSI behavior (K-SADS-PL) baseline to 24 months is based on self-report and will as a rule of thumb also include the suicidal and NSSI events also reported in the SAE events section (thereby the reports are not mutually exclusive, but the assessment of the self-reported data aims to cover all events) **Withdrawal due to SAE's: Because of blinding of the randomization outcome, assessments did not include inquiry of withdrawal from treatment (OPUS YOUNG or TAU) due to harmful effects. Instead, an unblinded employee withdrew data from medical records of reasons for terminating treatment.

Statistical Principles

Confidence intervals and p-values

All applicable statistical tests and confidence intervals will be two-sided and will be performed using a 5% statistical significance level. The 95% confidence intervals will not be adjusted for multiplicity and should not be used in place of hypothesis testing. To account for multiplicity and preserve the overall type I error for the numerous secondary endpoints, a prioritized analysis strategy will be applied. Key secondary outcomes assessed at 24 months will be evaluated using a prespecified hierarchical (serial gatekeeping) testing strategy to control the family-wise type I error rate, i.e. they will be tested sequentially in the following order: (1) severity of psychotic symptoms assessed using the Scale for the Assessment of Positive Symptoms (SAPS) defined as the mean of the global scores for hallucinations and delusions; (2) severity of negative symptoms assessed using the Scale for the Assessment of Negative Symptoms (SANS) defined as the mean of the global scores for anhedonia, avolition, affective flattening and alogia; (3) severity of disorganized symptoms, defined as the mean of the SAPS global scores for formal thought disorder and bizarre behavior and the SANS item for inappropriate affect; (4) patient satisfaction assessed using the Client Satisfaction Questionnaire (CSQ); and (5) health-related quality of life assessed using the KIDSCREEN-10 questionnaire. Each outcome will be formally tested only if the preceding endpoint in the hierarchy demonstrates a statistically significant between-group difference. If a nonsignificant result is observed at any step, all subsequent outcomes will be considered exploratory.

Adherence and protocol deviations

The flow diagram in Figure 1 shows patient adherence to the research assessments at baseline, 12 months and 24 months.

For both groups non-adherence to treatment is registered if they discontinue treatment for any reason.

Protocol deviations include outcomes reported by proxy by parents in cases where a participant is unavailable for assessment; assessments performed later than three months after the scheduled time of assessment; and active or passive refusal or inability to participate in clinician rated outcomes or questionnaires. In cases where participants agreed to a shortened assessment, a prioritized list of outcome measures were assessed. All deviations were registered and will be reported in supplementary material.

Analysis populations

The main analyses will be based on the Intention-to-Treat (ITT) population. Thus, all randomized participants will be followed, assessed, and analyzed according to their originally assigned treatment group, independent of withdrawals and cross-over phenomena. The main analysis population represents the treatment policy estimand corresponding to the average treatment effect among all randomly assigned participants. This population will be used for the main inferential measures.

We will make corresponding analysis based on the per protocol population. The per-protocol set is defined as participants with available data at baseline and 24 months, adhering to treatment for the full

trial period and not transitioning during the trial period to adult OPUS intervention had they been randomized to TAU. See Appendix Table 1.

Secondly, to explore our findings under different missing data assumptions we will redo the analysis of primary and key secondary outcomes using non-responder imputations, where missing data is replaced with baseline values. See Appendix Table 2.

Furthermore, we will make corresponding analysis using the hypothetical estimand, the “what if”- population analyzed under the theoretical scenario that all participants adhered to their assigned treatment for the complete 24 months. See Appendix Table 3.

Lastly, we will make an additional analysis corresponding to the main analysis on the ITT population but at the 12-month assessments (mid-trial). See Appendix Table 4.

Trial Population

Screening data

The screening data of the referred patients evaluated for eligibility will be used to describe the representativeness of the sample and to show the procedure of recruitment and assessment of eligibility. An overview of the patient flow including eligibility criteria is shown in Figure 1, and the in- and exclusion criteria are shown below. Ineligible participants were excluded before randomization and therefore not allocated to treatment.

Eligibility

Patient inclusion criteria

- Age 12 to 17 years (both inclusive).
- First-episode psychosis within F2 spectrum (F20 schizophrenia, F21 Schizotypal disorder, F22 delusional disorder, F23 acute and transient psychotic disorders, F25 schizoaffective disorders, F28/29 other or un-specified non-organic psychosis) or F32.3, F33.3 depression with psychotic symptoms or F1X.5 substance-induced psychosis according to ICD-10[25]
- Maximum of 12 months since first prescription of antipsychotic treatment on the indication of psychosis.
- Speaks and understands Danish.
- Written informed consent from parents/legal caretakers. Adolescents between 15 and 17 years sign a written informed consent in addition to the parental consent.

Patient exclusion criteria

- An IQ of 49 or below (ICD-10: DF71, DF72, DF73).
- Current compulsory admission or compulsory treatment according to Danish legislation.

Diagnostic evaluation was done with the semi-structured interview, K-SADS-PL, supplemented with the psychosis sections (sections 17,18,19) of Schedule for Clinical Assessment in Neuropsychiatry Version 2.1 (SCAN 2.1)[43]. To enable evaluation of external validity, we registered sex, age and diagnoses of patients fulfilling the inclusion criteria, who did not wish to participate in the trial.

Recruitment procedure

The Central Visitation Unit and all in- and outpatient services at CAMHS referred patients with suspected first-episode psychosis to an assessment team located in the research unit, within the CAMHS. The health care professionals in the OPUS YOUNG assessment team conducted a comprehensive multidimensional assessment, including a structured psychiatric diagnostic interview with the K-SADS - PL and the psychosis sections of SCAN 2.1. Upon the assessments, a clinical diagnostic conference was conducted in collaboration between the clinical team in CAMHS and the relevant assessor from the assessment team. The conference was led by a consultant in child- and adolescent psychiatry or a psychologist specialized in child and adolescent psychiatry from the outpatient clinic. Clinical patients already diagnosed with first-episode psychosis could also be referred to the trial and after an affirmation of eligibility and adapted assessment be invited to participate.

If a patient was eligible for the OPUS YOUNG trial according to the in- and exclusion criteria, the patient and their parents/legal caretakers were invited to participate. The child and parents/legal caretakers received age-appropriate verbal and written information about the trial. Adolescents between 15 and 17 years of age signed a written informed consent in addition to the parental consent. Participants who turned 18-years of age during the trial were asked to give their own independent informed consent to continue their study participation. If the consent for some reason was not obtained, no data was registered from the date of the participants 18th birthday.

To assess relevant information concerning experience of parenting a child with a mental illness, each parent/legal caretaker received their own written and verbal participant information and informed consent form regarding parental participation. If the parents did not give informed consent to their personal participation in the trial, their child could still be included in the trial.

Withdrawal/follow-up

For level, timing, and reason of withdrawal from the research or treatments groups and lost to follow-up, see Figure 1 that shows adherence to the research assessments.

Baseline patient characteristics

See Table 1

Analysis

Outcome definitions

BOX 2 defines each outcome in detail with description and units. Primary and key secondary variables are clearly identified, and the outcomes are presented in hierarchical order (gatekeeping)

BOX 2. Primary and key secondary outcome description

		Time in months			
		Format	0	12	24*
Primary Outcome	Primary Outcome Measure				
Social Functioning	PSP: Personal and Social Performance Scale. PSP provides a total score (last month) on a scale from 1-100, with lower scores indicating lower personal and social functioning. The reported total score is based on evaluation of four subdomains (A. Socially useful activities, B. Personal and social relationships, C. Self-care, and D. Disturbing and aggressive behavior), rated on a 6-point Likert scale (1=absent to 6=very severe). The scoring of PSP is based on all available information and concerns the patient's daily level of functioning in the family, in school, and during leisure time. Researchers interview the participants prior to the scoring of PSP using a semi-structured interview guide. Assessed at baseline, 12- and 24 months.	I	x	x	x
Key Secondary Outcomes	Key Secondary Outcome Measures				
Severity of psychotic symptom dimensions	Scale for Assessing Positive Symptoms (SAPS) is used to assess the phenomenological presentation of psychotic symptoms in schizophrenia. The score is based on the last (maximum three) month. The scale is divided into four sub-scales: hallucinations, delusions, bizarre behavior and formal thought disorder. Each is rated from 0 (absent) to 5 (severe) and each domain is also given a global symptom severity score. The reported psychotic symptom dimension is the mean of global scores of hallucinations and delusions (score range 0-5). Higher score indicates worse level of symptoms. Assessed at baseline, 12-and 24 months.	I	x	x	x
Severity of negative symptom dimension	Scale for Assessing Negative Symptoms (SANS) is used to assess the phenomenological presentation of negative symptoms in schizophrenia. The score is based on the last (maximum three) month. The scale is divided into five sub-scales: affective flattening, alogia, avolition/apathy, anhedonia/antisociality, attention. Each is rated from 0 (absent) to 5 (severe) and each domain is also given a global symptom severity score. The reported negative symptom dimension is the mean of	I	x	x	x

BOX 2. Primary and key secondary outcome description

	global scores of anhedonia, avolition, affective flattening and alogia (score range 0-5). Higher score indicates worse level of symptoms. Assessed at baseline, 12- and 24 months.				
Severity of disorganized symptom dimension	Based on items from SAPS and SANS. The score is based on the last (maximum three) month. The reported disorganized symptom dimension is the mean of the global scores of formal thought disorder and bizarre behaviors from SAPS and the score of inappropriate affect (item within the subscale for affective flattening) in SANS (score range 0-5). Higher score indicates worse level of symptoms. Assessed at baseline, 12- and 24 months.	I	x	x	x
Client satisfaction, patient-rated	CSQ: Client Satisfaction Questionnaire. The questionnaire has eight question-items rated by patients using a 4-point Likert scale. Their responses are scored from 1 to 4 (total scores range 8-32). Higher scores indicate greater satisfaction. Assessed at 12- and 24months.	Q		x	x
Health-related quality of life, patient-rated	KIDSCREEN-10: Health-Related Quality of Life Questionnaire for Children and Young People and their Parents. The score is based on the last week. The questionnaire consists of ten items. Participants respond using a 5-point Likert scale. Answer categories item 1 and 9: not at all-slightly-moderately-very-extremely; all other items: never-seldom-quite often-very often-always. Responses are coded so that higher values indicate better quality of life. Total score reported ranges 10-50. The raw scores are transformed into a T-score following scoring instructions. Assessed at baseline, 12- and 24 months.	Q	x	x	x

Note. *Treatment endpoint. I: Clinical interview; Q: Questionnaire.

Analysis methods

Continuous outcome measures will be analyzed as the change from baseline to 24 months using repeated measures mixed-effects linear models. The participant identification number will be included as a random effect, with the baseline score, treatment group (OPUS YOUNG or TAU [2 levels]), time (baseline, 12 months, or 24 months [3 levels]), the two stratification factors age group (12–14 vs 15–17 years), and baseline level of personal and social functioning (Personal and Social Performance Scale score 1–44 vs 45–100), and interactions between treatment group and time as fixed effects.

Model diagnostics will be conducted to evaluate the assumptions underlying the repeated measures mixed-effects linear models. Residual plots and Q-Q plots will be used to assess normality and homoscedasticity of residuals, while partial residual plots will be examined to assess linearity of fixed effects. Influential observations will be identified using Cook's distance. The distribution of random effects will be evaluated using empirical Bayes estimates and Q-Q plots, and intraclass correlation coefficients will be reported to quantify between-participant variability. We will report estimated means for each treatment group using least squares means (LSMeans), which represent model-based, adjusted group means that account for covariates included in the analysis. Unlike raw means, LSMs reflect the structure and assumptions of the fitted mixed-effects model and are appropriate for inference in analyses of repeated measures or covariate-adjusted outcomes. The primary and key secondary objectives based on continuous outcome measures will be assessed using differences between LSMs and reported with 95% confidence intervals.

Categorical efficacy outcomes will be analyzed at each time point individually using logistic regression models that include the same fixed-effect factors and covariates as the repeated measures mixed-effects linear models mentioned above. Categorical outcomes will be reported as counts and proportions for each group, with between-group contrasts initially expressed as odds ratios and corresponding 95% confidence intervals estimated from the logistic regression models. This approach provides both descriptive clarity and model-based inference regarding the relative likelihood of outcomes across treatment groups.

Stratified Analyses – Exploring potentially important contextual factors

Patients enrolled in randomized clinical trials are inherently heterogeneous with respect to clinical history, demographic characteristics, and other baseline features. Consequently, it is appropriate to examine whether the average treatment effect observed in the overall trial population applies consistently across clinically relevant subgroups or whether treatment effects vary according to prespecified baseline characteristics. Such subgroup analyses are not intended to undermine the primary trial result but to assess the extent to which treatment benefits and harms are modified by patient-level factors that may influence prognosis or response to treatment.

When conducted within a prespecified analytical framework, stratified analyses can provide important complementary insights beyond the overall treatment effect. They allow assessment of effect modification, identifying whether the magnitude or direction of treatment effects differs across levels of baseline risk or key clinical characteristics. Such analyses can inform clinical applicability and decision-making, particularly when baseline characteristics are strongly associated with outcome risk or

treatment tolerance. Finally, transparent, hypothesis-driven subgroup analyses that adhere to credibility criteria, such as those operationalized in the ICEMAN framework[44], help distinguish plausible effect modification from spurious findings driven by chance, multiplicity, or post hoc data exploration. When interpreted cautiously and in conjunction with the primary analysis, stratified analyses therefore enhance the clinical interpretability of trial results without compromising inferential integrity.

We will perform prespecified stratified analyses to examine whether selected baseline characteristics modify the effect of the intervention. Because randomization was stratified by age group (12–14 vs 15–17 years) and baseline level of personal and social functioning (Personal and Social Performance Scale score 1–44 vs 45–100), corresponding stratified analyses will be conducted to assess whether treatment effects are consistent across these strata or whether clinically meaningful effect modification is present. Evaluating treatment effects within strata defined a priori by the randomization scheme supports coherent interpretation of trial findings, maintains alignment between trial design and analysis, and informs the applicability of the overall treatment effect across clinically distinct subgroups. In addition, exploratory analyses will assess whether the following dichotomously defined baseline characteristics may act as effect modifiers: (i) prescription of antipsychotic medication at baseline (yes vs no), (ii) biological sex (male vs female), (iii) gender identity (cis vs non-cis), (iv) educational or occupational attendance at baseline (attending school/work/full-time/part-time vs. not attending), (v) psychosis diagnosis at baseline (schizophrenia vs. other psychosis diagnosis) and (vi) comorbid autism diagnosis at baseline (autism vs. no autism).

To evaluate whether the treatment effect on the primary outcome varies across the prespecified subgroups, stratified analyses will be conducted using the approach previously outlined by Christensen et al[45]. For each of the outlined candidate baseline characteristics, an interaction term between randomized treatment group and the subgroup variable will be added to the primary linear model for change in Personal and Social Performance Scale score. Treatment effects will be estimated separately within each stratum and presented with 95% confidence intervals to facilitate clinical interpretation. The outcome of the (number) outlined stratified analyses will focus on comparing the difference in net treatment benefit between subgroups using formal tests for interaction, rather than on within-subgroup statistical significance.

Missing outcome data

Missing data for continuous outcomes will be handled indirectly with the use of the mixed-effects linear-models approach. A valid approach assuming data are missing at random (MAR), in accordance with the working assumption underlying these models[46].

For the main analysis of categorical outcomes, missing data will be handled simplistically using non-response imputation, whereby participants who withdraw or do not provide a response will be classified as non-responders.

Additional analyses

Four additional analyses will be conducted. First, a per protocol analysis, where the main analysis of primary and key secondary outcomes is repeated in the per protocol population (PP). This analysis estimates the treatment effect among participants who adhered to the protocol, assessing whether deviations from protocol influenced the primary findings. Second, an analysis using non-responder imputation on missing outcome data in the primary and key secondary outcomes. This analysis assesses the robustness of the results under a conservative missing data assumption in which participants with missing outcome data are classified as non-responders. Third, we will conduct an analysis of the hypothetical estimand. This analysis targets a hypothetical estimand corresponding to the treatment effect that would have been observed if all participants had adhered to their assigned treatment for the full 24-month period. The three analyses will test the sensitivity of the results to different population scenarios and assumption. See Appendix Table 1-3. A fourth analysis will examine the primary, key secondary outcomes, treatment response and remission and status of prescribes antipsychotic treatment at 12-month assessments (mid-trial) in the ITT populations.

Appendix Table 1: The primary and key secondary outcomes at 24 months (The PP population).

Appendix Table 2: The primary and key secondary outcomes at 24 months (The non-responder population using imputation with baseline values to replace missing outcome data).

Appendix Table 3: The primary and key secondary outcomes at 24 months (The population based on hypothetical estimand)

Appendix Table 4: The primary and key secondary outcomes at 12 months (The ITT Population)

Harms (adverse events: coding, categorization, analysis)

All serious adverse events (SAEs) are registered and reported to The Ethics Committee of the Capital Region of Denmark. It is stated which variables are collected from medical records and which are patient reported.

Information on harms will be analyzed using descriptive methods with added Risk Difference with a 95% confidence interval.

Statistical software

To enhance analytical transparency and robustness, the primary analyses will be independently implemented in both SAS and R. Qualitative comparison of effect estimates, confidence intervals, and inferential conclusions across software platforms will allow assessment of the stability of the findings with respect to differences in numerical algorithms, default model specifications, and variance estimation procedures that may vary between statistical environments. Concordant results across SAS and R would strengthen confidence that the observed treatment effects reflect the underlying data rather than

software-specific implementation choices, whereas any discrepancies would prompt targeted investigation to confirm correct model specification and interpretation. This dual-implementation approach supports reproducibility, facilitates external verification, and aligns with best-practice recommendations for transparent statistical reporting.

Analyses will be performed using R version 4.4.2 or newer (packages nlme, emmeans, and tidyverse; R Project for Statistical Computing) and SAS Studio (SAS Institute Inc.), using SAS/STAT® software, version 9.4.

Data management

Data are entered directly into a REDCap[26] database and examined carefully to identify errors in data entry. Algorithms on randomization outcome are kept in special files. Data is exported to relevant software packages

References

- [1] C. Kieling *et al.*, “Worldwide Prevalence and Disability From Mental Disorders Across Childhood and Adolescence,” *JAMA Psychiatry*, vol. 81, no. 4, p. 347, Apr. 2024, doi: 10.1001/jamapsychiatry.2023.5051.
- [2] C. Hjorthøj, A. E. Stürup, J. J. McGrath, and M. Nordentoft, “Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis,” *Lancet Psychiatry*, vol. 4, no. 4, pp. 295–301, Apr. 2017, doi: 10.1016/S2215-0366(17)30078-0.
- [3] H. Jin and I. Moswewu, “The Societal Cost of Schizophrenia: A Systematic Review,” *Pharmacoeconomics*, vol. 35, no. 1, pp. 25–42, Jan. 2017, doi: 10.1007/s40273-016-0444-6.
- [4] C. Hakulinen *et al.*, “The association between early-onset schizophrenia with employment, income, education, and cohabitation status: nationwide study with 35 years of follow-up,” *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 54, no. 11, pp. 1343–1351, Nov. 2019, doi: 10.1007/s00127-019-01756-0.
- [5] L. Martens and J. Addington, “The psychological well-being of family members of individuals with schizophrenia,” *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 36, no. 3, pp. 128–133, May 2001, doi: 10.1007/s001270050301.
- [6] C. U. Correll *et al.*, “Comparison of Early Intervention Services vs Treatment as Usual for Early-Phase Psychosis,” *JAMA Psychiatry*, vol. 75, no. 6, p. 555, Jun. 2018, doi: 10.1001/jamapsychiatry.2018.0623.

- [7] L. H. Hastrup *et al.*, “Cost-effectiveness of early intervention in first-episode psychosis: economic evaluation of a randomised controlled trial (the OPUS study),” *British Journal of Psychiatry*, vol. 202, no. 1, pp. 35–41, Jan. 2013, doi: 10.1192/bjp.bp.112.112300.
- [8] A.-W. Chan *et al.*, “SPIRIT 2025 statement: updated guideline for protocols of randomized trials,” *Nat. Med.*, vol. 31, no. 6, pp. 1784–1792, Jun. 2025, doi: 10.1038/s41591-025-03668-w.
- [9] ICH International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH, “ICH Harmonised Guideline Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6 (R2),” 2016.
- [10] L. Petersen *et al.*, “A randomised multicentre trial of integrated versus standard treatment for patients with a first episode of psychotic illness,” *BMJ*, vol. 331, no. 7517, p. 602, Sep. 2005, doi: 10.1136/bmj.38565.415000.E01.
- [11] J. A. Hartmann *et al.*, “Opening the Black Box of Cognitive-Behavioural Case Management in Clients with Ultra-High Risk for Psychosis,” *Psychother. Psychosom.*, vol. 86, no. 5, pp. 292–299, 2017, doi: 10.1159/000477551.
- [12] W. R. McFarlane, “Multiple-Family Groups and Psychoeducation in the Treatment of Schizophrenia,” *Arch. Gen. Psychiatry*, vol. 52, no. 8, p. 679, Aug. 1995, doi: 10.1001/archpsyc.1995.03950200069016.
- [13] S. Bowman, M. Alvarez-Jimenez, D. Wade, L. Howie, and P. McGorry, “The Positive and Negative Experiences of Caregiving for Siblings of Young People with First Episode Psychosis,” *Front. Psychol.*, vol. 8, May 2017, doi: 10.3389/fpsyg.2017.00730.
- [14] D. L. Penn, D. L. Roberts, D. Combs, and A. Sterne, “Best Practices: The Development of the Social Cognition and Interaction Training Program for Schizophrenia Spectrum Disorders,” *Schizophr. Bull.*, vol. 58, no. 4, pp. 449–451, 2007, doi: 10.1176/appi.ps.58.4.449.
- [15] D. G. Robinson *et al.*, “Psychopharmacological Treatment in the RAISE-ETP Study: Outcomes of a Manual and Computer Decision Support System Based Intervention,” *American Journal of Psychiatry*, vol. 175, no. 2, pp. 169–179, Feb. 2018, doi: 10.1176/appi.ajp.2017.16080919.
- [16] Danish Medicines Council, “Guideline: <https://medicinraadet.dk/anbefalinger-og-vejledninger/behandlingsvejledninger-og-laegemiddelrekommandationer/antipsykotika-til-born-og-unge>,” ©Medicinrådet.
- [17] S. P. Singh, B. Anderson, K. Liabo, and T. Ganeshamoorthy, “Supporting young people in their transition to adults’ services: summary of NICE guidance,” *BMJ*, p. i2225, May 2016, doi: 10.1136/bmj.i2225.
- [18] National Institute for Health and Care Excellence, “NICE: Psychosis and schizophrenia in children and young people: recognition and management.

2016.<https://www.nice.org.uk/guidance/cg155/chapter/Recommendations#promoting-recovery-and-providing-possible-future-care-in-secondary-care>.”

- [19] M. U. Pedersen, S. Jones, K. B. E. Skov, K. S. Frederiksen, and M. M. Pedersen, *MOVE. Struktureret, forstærkende rusmiddelbehandling: Manual. Aarhus Denmark*. Aarhus, 2019.
- [20] National Institute for Health and Care Excellence, “<https://www.nice.org.uk/guidance/ng58>.”
- [21] M. Melau, N. Albert, and M. Nordentoft, “Development of a fidelity scale for Danish specialized early interventions service,” *Early Interv. Psychiatry*, vol. 13, no. 3, pp. 568–573, Jun. 2019, doi: 10.1111/eip.12523.
- [22] M. Melau, N. Albert, and M. Nordentoft, “Programme fidelity of specialized early intervention in Denmark,” *Early Interv. Psychiatry*, vol. 13, no. 3, pp. 627–632, Jun. 2019, doi: 10.1111/eip.12549.
- [23] D. Addington *et al.*, “Fidelity scales and performance measures to support implementation and quality assurance for first episode psychosis services,” *Early Interv. Psychiatry*, vol. 12, no. 6, pp. 1235–1242, Dec. 2018, doi: 10.1111/eip.12684.
- [24] Sundhedsstyrelsen, “Sundhedsstyrelsen. Specialevejledning for Børne- og ungdomspsykiatri.” [Online]. Available: https://www.sst.dk/-/media/Viden/Specialplaner/Specialeplan-for-børne-og-ungdomspsykiatri/SST_Specialevejledning-for-boerne-og-ungdomspsykiatri-11-april-2019.ashx?la=da&hash=BCF7C83B6D6DB5291B621A6805C841CADD2EA137
- [25] World Health Organisation, “ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines,” 1992, *World Health Organisation, Geneva*.
- [26] P. A. Harris *et al.*, “The REDCap consortium: Building an international community of software platform partners,” *J. Biomed. Inform.*, vol. 95, p. 103208, Jul. 2019, doi: 10.1016/J.JBI.2019.103208.
- [27] N. Albert *et al.*, “Five years of specialised early intervention versus two years of specialised early intervention followed by three years of standard treatment for patients with a first episode psychosis: randomised, superiority, parallel group trial in Denmark (OPUS II),” *BMJ*, p. i6681, Jan. 2017, doi: 10.1136/bmj.i6681.
- [28] R. E. Ulloa, R. Apiquian, G. Victoria, S. Arce, N. González, and L. Palacios, “Validity and reliability of the Spanish version of the Personal and Social Performance scale in adolescents with schizophrenia,” *Schizophr. Res.*, vol. 164, no. 1–3, pp. 176–180, May 2015, doi: 10.1016/J.SCHRES.2015.02.010.
- [29] R. Pérez-Garza, G. Victoria-Figueroa, and R. E. Ulloa-Flores, “Sex Differences in Severity, Social Functioning, Adherence to Treatment, and Cognition of Adolescents with

- Schizophrenia,” *Schizophr. Res. Treatment*, vol. 2016, pp. 1–7, 2016, doi: 10.1155/2016/1928747.
- [30] D. Petrić, V. Rački, N. Gačo, A. Kaštelan, and M. Graovac, “Retrospective Analysis of the Effectiveness and Tolerability of Long-Acting Paliperidone Palmitate Antipsychotic in Adolescent First-Episode Schizophrenia Patients,” *J. Child Adolesc. Psychopharmacol.*, vol. 29, no. 3, pp. 197–204, Apr. 2019, doi: 10.1089/cap.2018.0044.
 - [31] H. Nasrallah, P. Morosini, and D. D. Gagnon, “Reliability, validity and ability to detect change of the Personal and Social Performance scale in patients with stable schizophrenia,” *Psychiatry Res.*, vol. 161, no. 2, pp. 213–224, Nov. 2008, doi: 10.1016/j.psychres.2007.11.012.
 - [32] P. L. Morosini, L. Magliano, L. Brambilla, S. Ugolini, and R. Pioli, “Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning,” *Acta Psychiatr. Scand.*, vol. 101, no. 4, pp. 323–9, Apr. 2000.
 - [33] N. C. Andreasen, “Symptoms of Schizophrenia,” *Arch. Gen. Psychiatry*, vol. 52, no. 5, p. 341, May 1995, doi: 10.1001/archpsyc.1995.03950170015003.
 - [34] N. C. Andreasen, “The Scale for the Assessment of Negative Symptoms (SANS): conceptual and theoretical foundations,” *Br. J. Psychiatry Suppl.*, no. 7, pp. 49–58, Nov. 1989.
 - [35] C. C. Attkisson and R. Zwick, “The client satisfaction questionnaire,” *Eval. Program Plann.*, vol. 5, no. 3, pp. 233–237, Jan. 1982, doi: 10.1016/0149-7189(82)90074-X.
 - [36] U. Ravens-Sieberer *et al.*, “Reliability, construct and criterion validity of the KID-SCREEN-10 score: a short measure for children and adolescents’ well-being and health-related quality of life,” *Quality of Life Research*, vol. 19, no. 10, pp. 1487–1500, Dec. 2010, doi: 10.1007/s11136-010-9706-5.
 - [37] T. Nielsen, M. Pontoppidan, M. Pettersson, C. H. Donstrup, S. Kreiner, and S. B. Rayce, “Measuring child and adolescent well-being in Denmark: Validation and norming of the Danish KIDSCREEN-10 child/adolescent version in a national representative sample of school pupils in grades five through eight,” *PLoS One*, vol. 18, no. 9, p. e0291420, Sep. 2023, doi: 10.1371/journal.pone.0291420.
 - [38] C. M. Cassidy, R. Norman, R. Manchanda, N. Schmitz, and A. Malla, “Testing Definitions of Symptom Remission in First-Episode Psychosis for Prediction of Functional Outcome at 2 Years,” *Schizophr. Bull.*, vol. 36, no. 5, pp. 1001–1008, Sep. 2010, doi: 10.1093/schbul/sbp007.
 - [39] H. G. Hansen *et al.*, “Clinical Recovery Among Individuals With a First-Episode Schizophrenia an Updated Systematic Review and Meta-Analysis,” *Schizophr. Bull.*, vol. 49, no. 2, pp. 297–308, Mar. 2023, doi: 10.1093/schbul/sbac103.

- [40] J. Kaufman *et al.*, “Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data,” *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 36, no. 7, pp. 980–988, Jul. 1997, doi: 10.1097/00004583-199707000-00021.
- [41] A. Rozental, A. Kottorp, J. Boettcher, G. Andersson, and P. Carlbring, “Negative Effects of Psychological Treatments: An Exploratory Factor Analysis of the Negative Effects Questionnaire for Monitoring and Reporting Adverse and Unwanted Events,” *PLoS One*, vol. 11, no. 6, p. e0157503, Jun. 2016, doi: 10.1371/journal.pone.0157503.
- [42] O. Lingjærde, U. G. Ahlfors, P. Bech, S. J. Dencker, and K. Elgen, “The UKU side effect rating scale: A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients,” *Acta Psychiatr. Scand.*, vol. 76, no. s334, pp. 1–100, Jun. 1987, doi: 10.1111/j.1600-0447.1987.tb10566.x.
- [43] “World Health Organization. Division of Mental Health. (1994). Schedules for clinical assessment in neuropsychiatry: version 2. American Psychiatric Press. <https://iris.who.int/handle/10665/40356>.”
- [44] S. Schandelmaier *et al.*, “Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses,” *CMAJ*, vol. 192, no. 32, pp. E901–E906, Aug. 2020, doi: 10.1503/cmaj.200077.
- [45] R. Christensen, M. J. L. Bours, and S. M. Nielsen, “Effect Modifiers and Statistical Tests for Interaction in Randomized Trials,” *J. Clin. Epidemiol.*, vol. 134, pp. 174–177, Jun. 2021, doi: 10.1016/j.jclinepi.2021.03.009.
- [46] M. A. Detry and Y. Ma, “Analyzing Repeated Measurements Using Mixed Models,” *JAMA*, vol. 315, no. 4, pp. 407–8, Jan. 2016, doi: 10.1001/jama.2015.19394.

APPENDICES:

Appendix Table 1. Per protocol analysis. The primary and key secondary outcomes at 24 months (The PP population)

End point	OPUS YOUNG N=?	TAU N=?	Difference between groups (95% CI)	P Value
Primary outcome measure				
Change in social function (PSP), range 1-100				
Key secondary outcome measure				
Change in psychotic dimension (SAPS), range 0-5				
Change in negative dimension (SANS), range 0-5				
Change in disorganized dimension ^a (SAPS/SANS), range 0-5				
Client satisfaction (CSQ), range 8-32				
Change in health-related quality of life (KIDSCREEN-10), T-score				

Note: Definition of per-protocol population: 1) including participants with data from Baseline and 24-month follow-up. 2) Including participants adhering to treatment. 3) Excluding participants randomized to TAU who were provided clinical OPUS adult treatment during the trial period.

Results will be reported as least-squares means with standard errors, and between-group differences will be expressed with 95% confidence intervals. Responder outcomes will be analyzed using logistic regression and reported as odds ratios with 95% confidence intervals. Missing continuous outcome data will be handled using mixed-effects models for repeated measures, which accommodate data missing at random.

Abbreviations: PSP: Personal and Social Performance Scale, SAPS: Scale for Assessing Positive Symptoms SANS: Scale for Assessing Negative Symptoms, CSQ: Client Satisfaction Questionnaire, KIDSCREEN-10: Health-Related Quality of Life Questionnaire for Children and Young People and Their Parents.

^a: the mean of the global scores of formal thought disorder and bizarre behaviors from SAPS and the score of inappropriate affect (item within the subscale for affective flattening) in SANS

Appendix Table 2. Non-responder imputations. The primary and key secondary outcomes at 24 months.

End point	OPUS YOUNG N=?	TAU N=?	Difference between groups (95% CI)	P Value
Primary outcome measure				
Change in social function (PSP), range 1-100				
Key secondary outcome measure				
Change in psychotic dimension (SAPS), range 0-5				
Change in negative dimension (SANS), range 0-5				
Change in disorganized dimension ^a (SAPS/SANS), range 0-5				
Client satisfaction (CSQ), range 8-32				
Change in health-related quality of life (KIDSCREEN-10), T-score				

Note: Results will be reported as least-squares means with standard errors, and between-group differences will be expressed with 95% confidence intervals. Responder outcomes will be analyzed using logistic regression and reported as odds ratios with 95% confidence intervals. Missing continuous outcome data will be handled using mixed-effects models for repeated measures, which accommodate data missing at random. Missing data for responder outcomes will be handled conservatively using non-responder imputation.

Abbreviations: PSP: Personal and Social Performance Scale, SAPS: Scale for Assessing Positive Symptoms SANS: Scale for Assessing Negative Symptoms, CSQ: Client Satisfaction Questionnaire, KIDSCREEN-10: Health-Related Quality of Life Questionnaire for Children and Young People and Their Parents.

^a: the mean of the global scores of formal thought disorder and bizarre behaviors from SAPS and the score of inappropriate affect (item within the subscale for affective flattening) in SANS

Appendix Table 3. Analysis of the hypothetical estimand. The primary and key secondary outcomes at 24 months (hypothetical estimand population)

End point	OPUS YOUNG N=?	TAU N=?	Difference between groups (95% CI)	P Value
Primary outcome measure				
Change in social function (PSP), range 1-100				
Key secondary outcome measure				
Change in psychotic dimension (SAPS), range 0-5				
Change in negative dimension (SANS), range 0-5				
Change in disorganized dimension ^a (SAPS/SANS), range 0-5				
Client satisfaction (CSQ), range 8-32				
Change in health-related quality of life (KIDSCREEN-10), T-score				

Note: Results will be reported as least-squares means with standard errors, and between-group differences will be expressed with 95% confidence intervals. Responder outcomes will be analyzed using logistic regression and reported as odds ratios with 95% confidence intervals. Missing continuous outcome data will be handled using mixed-effects models for repeated measures, which accommodate data missing at random.

Abbreviations: PSP: Personal and Social Performance Scale, SAPS: Scale for Assessing Positive Symptoms SANS: Scale for Assessing Negative Symptoms, CSQ: Client Satisfaction Questionnaire, KIDSCREEN-10: Health-Related Quality of Life Questionnaire for Children and Young People and Their Parents.

^a: the mean of the global scores of formal thought disorder and bizarre behaviors from SAPS and the score of inappropriate affect (item within the subscale for affective flattening) in SANS

Appendix Table 4. The primary and key secondary outcomes at 12 months (The ITT Population)

End point	OPUS YOUNG N= ??	TAU N= ??	Difference between groups (95% CI)	P Value
Primary outcome measure				
Change in social function (PSP), range 1-100				
Key secondary outcome measure				
Change in psychotic dimension (SAPS), range 0-5				
Change in negative dimension (SANS), range 0-5				
Change in disorganized dimension (SAPS/SANS), range 0-5				
Client satisfaction (CSQ), range 8-32				
Change in health-related quality of life (KIDSCREEN-10), T-score, range 1-100				
Treatment response and remission				
Response (CGI-I) ^a , n (%)				
Remission (SANS/SAPS) ^b , n (%)				
Antipsychotic treatment				
Prescribed antipsychotic medication, n (%)				

Note: Results will be reported as least-squares means with standard errors, and between-group differences will be expressed with 95% confidence intervals. Responder outcomes will be analyzed using logistic regression and reported as odds ratios with 95% confidence intervals. Missing continuous outcome data will be handled using mixed-effects models for repeated measures, which accommodate data missing at random. Missing data for responder outcomes will be handled conservatively using non-responder imputation.

Abbreviations: PSP: Personal and Social Performance Scale, SAPS: Scale for Assessing Positive Symptoms SANS: Scale for Assessing Negative Symptoms, CSQ: Client Satisfaction Questionnaire, KIDSCREEN-10: Health-Related Quality of Life Questionnaire for Children and Young People and Their Parents, CGI-I: The Clinical Global Impression Scale– Improvement.

^a: Treatment response/non-response at 12 months is defined as a CGI-I score of 1, 2 or 3 (range from “very much improved” to “minimally improved”) and CGI-I scores from 4-7 (range from “no change” to “very much worse”) respectively.

^b: Remission at 12 months is defined as a global score of maximum 2 (mild) on the positive, negative, and disorganized dimension (SAPS/SANS) within the last three months[38], [39].

