Statistical Analysis Plan

Evaluation of the effectiveness of an online intervention for discharged childabuser

@myTabu

Studienprotokoll	
EudraCT-Nr.	—
Clinicaltrials.gov-Nr./DRKS - Nr.	DRKS00021256
Interne Protokoll-ID-Nr.	—
NCT-Nr.	—
SAP Version	0.1/13.09.2024
SAP-Revisionen	—
Entwicklungsphase	—
Sponsor	Universitätsmedizin Göttingen (UMG) Vertreten durch das Studienzentrum der UMG Robert-Koch-Str. 40 37075 Göttingen, Deutschland
Coordinating Investigator	Dr. Peter Fromberger Klinik für Psychiatrie und Psychotherapie Forensische Psychiatrie Universitätsmedizin Göttingen Rosdorfer Weg 70 37081 Göttingen, Deutschland

Approval of the Statistical Analysis Plan

@myTabu

Protocol version No.:

Koordinierender Prüfarzt
Dr. Peter Fromberger

18.11.24
Date

Signature

Biostatistiker
Dr. Andreas Leha

Date

Signature

Table of Conents

Approval of the Statistical Analysis Plan 2
Table of Conents 3
1 Introduction 5
2 Study Methods 7
3 Statistical Principles 9
4 Trial Population 9
5 Analysis 10
6 References 16

Abbreviations

BIS-15	Barratt Impulsiveness Scale – 15
BMS	Bumby Molest Scale
CI	Confidence intervals
CMC	Checklist of Behavioral Misconduct
CS	community supervision
CsaU	CS as Usual
CUSI	Coping Using Sex Inventory
CTE	Checklist of Treatment Effectiveness
CVTRQ	Corrections Victoria Treatment Readiness Questionnaire
CAs	Child abusers
DERS	Difficulties in Emotion Regulation Scale - Sub-scale Impulsivity
EKK-R	Emotional Congruence with Children – Revised
ESIQ	The Explicit Sexual Interest Questionnaire
F-SozU K-7	Social Support Questionnaire
HBI-19	Hypersexual Behavior Inventory – 19
IoD	Index of Desistance
IfSuFP	Institute for Sex Research and Forensic Psychiatry
ITT	intention-to-treat
NARQ	Negative Affect Repair Questionnaire
OQMPR	Optimized Questionnaire for the Measurement of Psychological Reactance
RCQ	Readiness to Change Questionnaire
SO	supervision officer
SOI-R	Outlet Inventory – Revised
SPSI-R	Social Problem-Solving Inventory Revised
SSIC	Specific self-efficacy for modifying Sexual Interest in Children
UCLA	University of California Los Angeles Loneliness Scale
UMG	University Medical Center Göttingen

1 Introduction

This document has been written based on the study protocol. A design paper has been published in Frontiers in Psychiatry; DOI: 10.3389/fpsyt.2020.575464

1.1 Background and Rationale

Treatment programs for child abusers are able to significantly reduce the risk to re-offend. In spite of some established treatment projects in primary prevention, treatment offers for child abusers conditionally released from prison are scarce – especially in rural areas. This is alarming, since most re-offences take place shortly after release from prison.

In addition to psychiatric or psychological treatment efforts, community supervision is effective in reducing re-offenses but only with high financial effort. Online interventions provide a cost-efficient additional treatment option and can be applied smoothly in rural areas.

1.2 Objectives and Endpoints

The current study aims to evaluate the effectiveness of (1) @myTabu in reducing dynamic risk factors compared to a placebo condition (2) the individual online modules within @myTabu with regard to reducing psychologically meaningful dynamic risk factors. It investigates the primary hypothesis that child abusers (CAs) who take part in @myTabu show lower values on dynamic risk factors (Index of Desistance, IoD) from baseline to inter-, and post-treatment, compared to an online-based placebo. The second hypothesis assumes that each individual online module within @myTabu is able to specifically address and reduce the dynamic risk factors which correspond to the psychotherapeutic content of the module. See table 1 for an overview of hypotheses.

Table 1 Objectives and related endpoints

	Objective	Endpoint
Primary	Evaluate the effectiveness of @myTabu in reducing dynamic risk factors compared to a placebo condition	Dynamic risk factors (Index of Desistance, IoD) from baseline to inter-, and post-treatment, compared to an online-based placebo
	Evaluate the effectiveness of @myTabu in reducing the number of officially recorded re-offenses compared to a placebo condition	Number of officially recorded re-offenses five years after last patient out
Secondary	Evaluate the Motivation Module individually with regard to its targeted risk factors low therapy motivation and low readiness for change	Differences in the Corrections Victoria Treatment Readiness Questionnaire, CVTRQ, pre and post Motivation Module
	Evaluate the Motivation Module individually with regard to its targeted risk factors low therapy motivation and missing readiness for change	Differences in the Readiness to Change Questionnaire, RCQ, pre and post Motivation Module
	Evaluate the Community-Supervision Module individually with regard to its	Differences in the Optimized Questionnaire for the Measurement of

	Objective	Endpoint
	targeted risk factors negative social influences and resistance to rules and supervision	Psychological Reactance, OQMPR, pre and post Community-Supervision Module
	Evaluate the Community-Supervision Module individually with regard to its targeted risk factors negative social influences and resistance to rules and supervision	Differences in the Social Support Questionnaire, F-SozU K-7, pre and post Community-Supervision Module
	Evaluate the Community-Supervision Module individually with regard to its targeted risk factors negative social influences and resistance to rules and supervision	Differences in the University of California Los Angeles Loneliness Scale, UCLA, pre and post Community-Supervision Module
	Evaluate the Emotion Management Module individually with regard to its targeted risk factors lifestyle impulsiveness and dysfunctional coping	Differences in the Difficulties in Emotion Regulation Scale - Sub-scale Impulsivity, DERS, pre and post the Emotion Management Module
	Evaluate the Emotion Management Module individually with regard to its targeted risk factors lifestyle impulsiveness and dysfunctional coping	Differences in the Negative Affect Repair Questionnaire, NARQ, pre and post the Emotion Management Module
	Evaluate the Emotion Management Module individually with regard to its targeted risk factors lifestyle impulsiveness and dysfunctional coping	Differences in the Barratt Impulsiveness Scale – 15, BIS-15, pre and post the Emotion Management Module
	Evaluate the Emotion Management Module individually with regard to its targeted risk factors lifestyle impulsiveness and dysfunctional coping	Differences in the Coping Using Sex Inventory, CUSI, pre and post the Emotion Management Module
	Evaluate the Problem Solving Module individually with regard to its targeted risk factor poor (social) problem solving	Differences in the Social Problem-Solving Inventory Revised, SPSI-R, pre and post the Problem Solving Module
	Evaluate the Offense-supportive Attitudes Module individually with regard to its targeted risk factor offense-supportive attitudes	Differences in the Bumby Molest Scale, BMS, pre and post the Offense-supportive Attitudes Module
	Evaluate the Sexuality Module individually with regard to its targeted risk factors sexual preoccupation, lack of emotionally intimate relationships with adults and emotional congruence with children	Differences in the Hypersexual Behavior Inventory – 19, HBI-19, pre and post the Sexuality Module
	Evaluate the Sexuality Module individually with regard to its targeted risk factors sexual preoccupation, lack of emotionally intimate relationships	Differences in the Specific self-efficacy for modifying Sexual Interest in Children, SSIC, pre and post the Sexuality Module

	Objective	Endpoint
	with adults and emotional congruence with children	
	Evaluate the Sexuality Module individually with regard to its targeted risk factors sexual preoccupation, lack of emotionally intimate relationships with adults and emotional congruence with children	Differences in the Outlet Inventory – Revised, SOI-R, pre and post the Sexuality Module
	Evaluate the Sexuality Module individually with regard to its targeted risk factors sexual preoccupation, lack of emotionally intimate relationships with adults and emotional congruence with children	Differences in Questionnaire on Emotional Congruence with Children – Revised, EKK-R, pre and post the Sexuality Module
	Evaluate the Sexuality Module individually with regard to its targeted risk factors sexual preoccupation, lack of emotionally intimate relationships with adults and emotional congruence with children	Differences in The Explicit Sexual Interest Questionnaire, ESIQ, pre and post the Sexuality Module
Safety	Assessment of safety	Number of violations of directives and re-offences

1.3 Primary objectives and endpoints

Officially recorded re-offenses in the experimental condition compared to the placebo condition are the first primary outcome. Officially recorded re-offenses are obtained five years after last patient out.

To evaluate the effectiveness of @myTabu in reducing dynamic risk factors compared to a placebo condition the Index of Desistance, IoD, is analyzed. The classical full IoD will be used. The IoD will be used in its reduced form without considering the CMC scale in an exploratory analysis.

1.4 Secondary objective and endpoints

Secondary endpoints are listed in Table 1.

1.5 Safety endpoints

Safety endpoint include violations of directives and re-offences.

2 Study Methods

2.1 Trial design

This trial is a concealed, prospective, parallel group, placebo-control trial with 1:1 randomization. Due to security reasons, blinding of supervision officers (SOs) and supporting therapists is not

possible. Nevertheless, subjects, outcome adjudicators and data analysts (observer) are blinded until the end of the data analysis.

2.2 Randomization

@myTabu comprises two parallel groups (online intervention group and online Placebo group). Subjects are randomly allocated to treatment using an allocation ratio of 1:1. The randomization lists are centrally generated using a computerized system stratified by the offense type, type of community supervision (CS), external treatment in addition to CS as Usual (CSaU), and risk to re-offend at baseline. At screening, each subject receives the next consecutive screening number. At randomization, each subject eligible for study participation receives the next consecutive randomization number according to his stratum from a block of randomization numbers (block size: 4). The randomization list is kept in safe and confidential custody at the "Stabstelle Klinische Studien" of the UMG.

2.3 Sample Size

@myTabu assumes that CAs who take part in @myTabu show significantly lower values on dynamic risk factors (IoD) from baseline to inter-, and post-treatment, compared to an online-based placebo condition. Within sex offender programs, higher levels of motivation and better treatment engagement were associated with decreased attrition. @myTabu addresses this in a direct manner (see sub-project 1). In addition, compliance of CAs in taking part of scientific studies is a critical point. Based on the experiences of the consortium, who have successfully finished numerous scientific studies with child abusers, @myTabu assumes a conservative rate of 50% of participants taking voluntarily part in the clinical trial. @myTabu tries to enhance the motivation to take part in the clinical trial by providing an adequate allowance. It is expected to see 1165 subjects eligible for participation during the recruitment period of which 50% (582) are expected to participate in the study. Taking an expected dropout of 26.4% into account, @myTabu expect to see 428 subjects completing the online intervention. With 214 subjects in both groups (treatment and placebo) it is possible to detect small effects ($d = .27$) with a power of 80% at a significance level $\alpha = 5\%$ (calculated using nQuery Advisor 7.0).

2.4 Framework

All endpoints are tested for superiority of the intervention over placebo control.

2.5 Stopping Guidance

The individual subject is excluded from the clinical trial if the responsible SO or the supporting therapist reports clear evidence for sexual recidivism or concrete preparations for any sexual recidivism. Stopping rules for participating centres are not applicable, since the UMG serves as the only participating centre.

2.6 Timing of the final analysis

The final analysis will take place when all outcomes have been collected and the database is cleaned and locked.

2.7 Timing of Outcome Assessments

Officially recorded re-offenses are obtained five years after last patient out.

The IoD is collected at the beginning of each module and at the end of the last module.

The secondary endpoints are collected at the beginning and at the end of the corresponding module (see Table 1).

2.8 Methods against Bias

@myTabu tries to minimize a potential volunteer-bias by providing an adequate allowance.

3 Statistical Principles

3.1 Confidence intervals and p-values

All tests will be performed two-sided with 5% significance level. Confidence intervals (CI) will be reported with 95% confidence level.

3.2 Adherence and protocol deviations

Feedback to subjects is automatized, elsewise manualized. Regular meetings between therapists and an experienced supervisor provided by IfSuFP will promote treatment consistency and regular supervision. Sessions are recorded and assessed.

3.3 Analysis populations

For all analyses on the primary endpoints the intention-to-treat (ITT) population will be used. The ITT population consists of all participants who have completed at least one module.

Analyses on the secondary endpoints will be carried out on the specific per-protocol (PP) population. The PP population for a specific module consists of all participants who have completed this module.

4 Trial Population

4.1 Screening Data

Screening data will be reported and described within a CONSORT flowchart.

4.2 Eligibility

All sexual offenders under CS (§§56, 57, 68 StGB), which are sentenced for at least one case of child abuse (§§176 StGB) or child pornography (§184b StGB) in Germany (besides the federal states Hamburg, Bremen and Saarland), are eligible for the trial. There exist no exclusion criteria regarding gender. Only persons who are older than 18 years are included. Exclusion criteria are (1) a probation period shorter than 6 months at enrollment, (2) no access to a PC, tablet or smart phone, (3) a severe acute psychiatric disorder (e.g., acute psychosis), (4) a severe cerebro-organic disorder, (5) a severe cognitive impairment, (6) withdrawal of the informed consent, and (7) no written informed consent. Due to a possible volunteer-effect, the representativeness of the sample for the entire population of CAs cannot be ensured.

4.3 Recruitment

Recruitment numbers will be summarized within a CONSORT flow diagram.

4.4 Drop-out/follow-up

Reasons for study drop-out will be documented using a CONSORT flowchart.

4.5 Baseline patient characteristics

All baseline patient characteristics will be summarized using descriptive statistics (e.g. mean, standard deviation, median, inter-quartile range, minimum and maximum value for continuous variables and frequencies (percentages) for categorical variables) and appropriate graphical methods (e.g. boxplots, bar plots) depending on the data type.

5 Analysis

5.1 Outcome definitions

A full list of outcomes and their timing is described in Section 2.7.

Officially recorded re-offenses Officially recorded re-offenses in the experimental condition compared to the placebo condition are the first primary outcome. Officially recorded re-offenses are obtained 5 years after last patient out.

Index of Desistance (IoD) The IoD represents the individual risk to re-offend assessed through self-report. This composite measurement reflects improvements with regard tomeaningful psychological risk factors, offense behaviors as well as offense-related behaviors that have not been registered or convicted yet. The IoD comprises three self-report instruments, which were developed and have been evaluated by members of the @myTabu-consortium (Nitsche et al., 2022): (1) The Acute-2007-SR is an online and self-report adaption of the Acute-2007 (Hanson & Harris 2007) risk-assessment tool in order to assess treatment induced short-term changes in recidivism risk factors. (2) The Checklist of Treatment Effectiveness (CTE) is developed to assess the differential treatment-induced changes in every individual treatment module. Thus, the CTE assesses the constructs that are targeted by the six modules of the intervention. (3) The Checklist

of Behavioral Misconduct (CMC) measures offense-related behavior as well as low- and high-level new offenses. Offense-related behavior includes behaviors which are relevant for individuals previously convicted of sexual offenses prior to a new offense (recidivistic behavior, e.g., if a person convicted of sexual abuse of minors visits a playground or a school). The classical IoD will provide the individual risk level to re-offend by summing up the scores of the Acute-2007-SR, CTE, and CMC. The reduced IoD is the sum of Acute-2007-SR and CTE.

Corrections Victoria Treatment Readiness Questionnaire (CVTRQ) The Corrections Victoria Treatment Readiness Questionnaire [CVTRQ; Casey et al., 2007] is a self-report measure designed to assess treatment readiness in offenders who have been referred to a cognitive skills program. It comprises 20 items and consists of four scales: (1) Attitudes and motivation, (2) emotional reaction, (3) offending beliefs, and (4) efficacy. Each item has to be answered on a five-point Likert-scale. The CVTRQ shows an acceptable convergent validity, discriminant validity as well as predictive validity (Casey et al., 2007). To the best of our knowledge, there is no validated German version of the CVTRQ. @myTabu translated the CVTRQ. The translation was checked and translated back into English. Resulting differences due to adaptation were inspected and discussed in terms of item content.

Readiness to Change Questionnaire (RCQ) The Readiness to Change Questionnaire [RCQ; Rollnick et al., 1992] is a 12-item questionnaire originally designed to identify the stage of change based on the transtheoretical model (Prochaska et al, 1992) reached by individuals who excessively drink alcohol. The German version [RCQ-D; Hannöver et al., 2002] was adapted within the project by changing the questions regarding drinking of alcohol in questions regarding the offenses of ISAC and ICCSEM. Responses are made on a five-point Likert-scale. The RCQ-D as well as the RCQ shows good psychometric properties (Hannöver et al., 2002). Psychometric properties of the version adapted for ISAC and ICCSEM are not available yet.

Optimized Questionnaire for the Measurement of Psychological Reactance (OQMPR) The Questionnaire for the Measurement of Psychological Reactance [QMPR; Merz 1983] is a questionnaire for the assessment of psychological reactance defined as the theory that people resist attempts to constrain either their thoughts or their behaviors. The Optimized Questionnaire for the Measurement of Psychological Reactance [OQMPR; Herzberg 2002] is a German variant of the QMPR. The OQMPR consists of 12 statements on a five-point Likert-scale. It has a test-retest-reliability of $rtt = 0.85$ (Herzberg 2002).

Social Support Questionnaire (F-SozU K-7) The seven-item short version of the Social Support Questionnaire (F-SozU K-7) is an efficient questionnaire to assess perceived social support (Dunkel et al., 2005). Each item comprises a five-point Likert-scale. Internal consistency of the original long version showed Cronbach's α between 0.81 and 0.93 (Sommer & Fydrich, 1989).

University of California Los Angeles Loneliness Scale (UCLA) The UCLA Loneliness Scale (Russell, 1996) consists of 20 items in order to assess subjective feelings of loneliness. The German short version (Bilsky & Hosser, 1998) consists of 12 items on a four-point Likert-scale. A German study with ISAC and ICCSEM demonstrated a high reliability of $\alpha = 0.92$ (Neutze et al, 2012).

Bumby Molest Scale (BMS) The Bumby Molest Scale [BMS; (Bumby, 1996)] is a self-report questionnaire which assesses cognitive distortion in ISAC [German Version: Feelgood et al., 2009]. It consists of 38 items on a four-point Likert-scale. The German version shows good construct validity, internal consistency, and test-retest reliability (Feelgood et al., 2009)..

Social Problem-Solving Inventory Revised (SPSI-R) The Social Problem-Solving Inventory Revised [SPSI-R; D'Zurilla et al., 2002] is a self-report questionnaire for the assessment of the five dimensions in the social problem-solving model: (1) positive problem orientation, (2) negative problem orientation, (3) rational problem solving, (4) impulsivity/carelessness style, and (5) avoidance style. The SPSI-R consists of 52 items on a five-point Likert-scale. It shows good psychometric properties for individuals who have sexually offended (Nezu et al., 2005). The validated German version is a short form consisting of 25 items (Graf, 2003).

Difficulties in Emotion Regulation Scale—Sub-scale Impulsivity (DERS) The Difficulties in Emotion Regulation Scale [DERS; Gratz & Roemer, 2004] is a self-report questionnaire to assess emotion dysregulation. The sub-scale "impulse control difficulties" is used as outcome measure and comprises five items on a five-point Likert-scale. The German version is used, which shows a good internal consistency, construct and predictive validity (Ehring et al., 2008).

Negative Affect Repair Questionnaire (NARQ) The Negative Affect Repair Questionnaire [NARQ; Scherer et al., 2013] is a self-report questionnaire to assess strategies to regulate negative affect in a systematic manner. It consists of 17 items on a five-point Likert-scale and provides a good construct validity. Reliability scores (Cronbach's α) for the three NARQ scales ranged between 0.71 and 0.80 (Scherer et al., 2013).

Barratt Impulsiveness Scale-15 (BIS-15) The Barratt Impulsiveness Scale [BIS; Patton et al., 1995] is a questionnaire developed to assess the personality/behavioral construct of impulsiveness. It consists of 30 items on a four-point Likert-scale. The German short version [BIS-15; Meule et al., 2011] consisting of 15 items is used in the clinical trial. The BIS-15 is an efficient measure of impulsiveness with good internal consistency [α = 0.81; Meule et al., 2011].

Coping Using Sex Inventory (CUSI) The Coping Using Sex Inventory [CUSI; Cortoni & Marshall, 2001] is a questionnaire to assess the presence of and the degree to which sex was used to deal with problematic situations. It consists of 16 items on a five-point Likert-scale with a satisfying internal consistency (Cortoni & Marshall, 2001). The translation was checked and translated back into English by the authors. Resulting differences due to adaptation were inspected and discussed in terms of item content. For another German translation, it has been shown that the CUSI is able to assess therapy-induced changes of ISAC (Beier et al., 2015).

Hypersexual Behavior Inventory-19 (HBI-19) Sexual preoccupation is assessed by the Hypersexual Behavior Inventory-19 [HBI-19; Klein et al., 2014]. The HBI-19 is a three-factor measure (coping, control, and consequences) developed to assess hypersexual behavior. The instrument consists of 19 items (e.g., "I use sex to forget sorrows of everyday life.") answered on a scale from 1 (never) to 5 (very often). The maximum score is 95, with higher scores indicating a higher level of sexual preoccupation. The questionnaire was shown to have good reliability (α = 0.90) and validity (Klein et al., 2014).

Questionnaire on Emotional Congruence with Children-Revised (EKK-R) Emotional congruence with children is assessed by the Questionnaire on Emotional Congruence with Children-Revised [EKK-R; Mack & Yundina, 2012] including three factors (special relationship to children, immaturity, and emotional closeness to children). Twenty items are answered on a four-point scale. The questionnaire demonstrated good reliability (α = 0.80) and validity (Mack & Yundina, 2012).

Specific self-efficacy for modifying Sexual Interest in Children (SSIC) Specific self-efficacy for modifying sexual interest in children (SSIC) is assessed by the Self-Efficacy for Modifying Sexual Interest in Children Scale [SSIC-Scale; Tozdan et al., 2015]. Six items on the participant's conviction regarding the ability to change their sexual interest in children (e.g., "I can succeed in reducing my sexual interest in children") were answered on a scale from 1 (do not agree at all) to 5 (totally agree). The maximum score is 30, with higher scores indicating a higher level of self-efficacy. The instrument was shown to have good reliability ($\alpha = 0.87$) and validity (Tozdan et al., 2015). Four variables that were shown to be related to the SSIC (Tozdan & Briken, 2015, Tozdan & Briken, 2019) are also assessed: flexibility of sexual interest in children (three items on previous experiences concerning changes in the participant's sexual interest in children), exclusiveness of sexual interest in children (single-item question), motivation to change sexual interest in children (single-item question), and age of onset of sexual interest in children (single-item question).

The Explicit Sexual Interest Questionnaire (ESIQ) The Explicit Sexual Interest Questionnaire [ESIQ; Banse et al., 2010] directly assesses pedophilic interest. It consists of two scales measuring sexual behavior (20 items, e.g., "I enjoyed orally stimulating a man.") and sexual fantasies (20 items, e.g., "I find it attractive to imagine a little boy sexually stimulating me."). All items are answered on a scale from 1 (totally disagree) to 5 (totally agree). The reliability of the instrument ranges between 0.86 and 0.97 (Banse et al., 2010).

Outlet Inventory-Revised (SOI-R) The second measurement for sexual interest in children is the Item 2a of the Sexual Outlet Inventory-Revised [SOI-R; Briken, 2010] which assesses the desire for sexual activity involving children on a visual analog scale from 0 (desire is absent) to 100 (I have to act to satisfy the desire). Higher values on the scale indicate a stronger sexual interest in children (Briken, 2010).

5.2 Analysis methods

All patients who discontinue from the study are identified and the extent of their participation in the study reported. If discontinuation is due to clear evidence for recidivism or concrete preparation for or intention to recidivism (assessed by responsible SO or guiding therapist) the subject receives the maximal value for IoD for the remaining measurements (assuming the worst case of a failed treatment). In all other cases of drop-out, the reason is recorded.

All variables corresponding to the endpoints will be summarized using descriptive statistics (e.g. mean, standard deviation, median, inter-quartile range, minimum and maximum value for continuous variables and frequencies (percentages) for categorical variables) and appropriate graphical methods (e.g. boxplots, bar plots) depending on the data type per time point.

The efficacy of the online intervention will be analyzed by means of Gaussian linear model for repeated measures (MMRM) with treatment group, time (six time-points; after each online intervention module), treatment-by-time interaction, and the randomization strata as factors (offense type, type of CS, external treatment in addition to CSaU), baseline measurements (at clearing up) of the outcome and potentially time span from baseline as covariate and the classical IoD as dependent variable. The error terms are assumed to follow a multivariate normal distribution with unstructured covariance if all parameters of the corresponding model can be estimated and Ante-dependence covariance structure otherwise. Least squares mean changes from baseline are reported for the treatment groups with 95% confidence interval (CI) as well as the difference between the least squares treatment group means with 95% CI and p-value testing

the null hypothesis of no treatment effect. The analysis is performed on the ITT population comprising all patients who have completed at least one module.

The analysis on the ITT population will be complemented by a sensitivity analysis on imputed values. Imputation will be done using multiple reference based imputation followed by ANCOVA models (Wolbers et al., 2022). Differences between the results of the two analyses will be discussed.

The effect of each module will additionally be analyzed using the Wilcoxon signed rank test for paired data if appropriate and the Brunner-Munzel test for paired data otherwise for both the IoD and the scores corresponding to the secondary endpoints comparing the pre- and post-module scores, adjusting the p-values using the Bonferroni-Holm's method to control the family-wise error rate.

The effect of each module on the scores corresponding to the secondary endpoints will be further analyzed by fitting linear models with treatment group and the randomization strata as factors (offense type, type of CS, external treatment in addition to CSaU), pre module scores as covariate and post module scores as dependent variable if appropriate. If necessary, the scores are transformed before this analysis. If the assumptions of a linear model are not fulfilled either an ordinal regression model will be fitted analogously instead or the score is dichotomised in order to fit a logistic regression model analogously.

The proportion of officially recorded re-offenses five years after last patient out is estimated using the Kaplan-Meier method with 95% confidence interval estimated using the beta product confidence procedure for right censored data (Fay, 2013) in both arms and compared between the arms using the beta product confidence procedure for right censored data (Fay, 2014). Depending on the number of events a Cox proportional hazard regression model with treatment group and an additional appropriate set of factors will be fit and resulting hazard ratios will be reported with 95% confidence intervals.

Economic analyses will be based on analytic modeling techniques, including data about recidivism rates based on dynamic risk factors (derived from the clinical trial), tangible benefits (based on existing literature), intangible benefits (based on surveys using the contingent valuation method) and intervention costs (based on the costs incurred per participant in the study). Tangible benefits are estimated on the basis of existing literature, such as costs occurring in the German society in the context of recidivism of child abusers. After reviewing the literature, a selection of relevant studies was made. Several criteria were taken into account in the final selection of relevant studies. Since the health economic evaluation was to be modeled over a 5-year period, only studies that allowed conclusions to be drawn about annual costs were taken into account. Furthermore, only studies were included in which the cost unit (e.g. per capita or per company) was clearly understandable. Furthermore, for studies that were based on the same dataset, only the study thematically most appropriate was included. For the intangible benefits, the contingent valuation method (CVM) was used to estimate the lower and upper bounds for intangible benefits related to reduced recidivism rates. Intangible costs of CA are non-measurable costs that arise around the suffering of the victims. CVM is a survey-based, hypothetical and direct method to value intangible costs, assessed by the willingness to pay. Participants are asked how much should be paid for a program which reduces the relapse rates of sexual child offenders and thus the victims' pain. Survey participants define the maximum money amount accounting for the indicated public good in an unbiased open-answer format. The percentage of prevented recurrence of CA is increased stepwise starting with 5% up to 100%, tightly spaced in the lower ranks, representing more likely potential outcomes of an intervention. The intervention costs include costs of developing the intervention (personnel costs for content and programming), costs associated with conducting the

study and maintaining the intervention (personnel costs for therapeutic support, personnel costs for maintenance and technical support, supervision costs, travel costs, personnel costs for study information and file analysis, compensation for participants) and costs associated with future implementation of the intervention (personnel costs for therapeutic support, personnel costs for maintenance and technical support, technology costs for the provision of appropriate servers and daily backups, supervision costs). We will perform a cost-benefit analysis. Results of the analysis will be presented in terms of a net benefit (NB) estimate and a benefit-cost ratio. The net benefit will be calculated as NB = tangible benefits + intangible benefits – intervention costs. The benefit-cost ratio will be calculated as benefits / costs.

Since many cost parameters occur after the trial period, the cost-benefit analysis of the intervention will be modelled for a 5-year period. Additionally, sensitivity analysis will be conducted to examine how the NB will be influenced by the recidivism rates, intangible benefits and discount rates. Tangible costs of the control group (no intervention) are estimated on the basis of existing literature. The per-person costs found in literature will be multiplied with German prevalence data, to get an estimate for the German society. In order to account for the costs caused by relapses, this data is multiplied by the probability of re-offending (without any treatment). The tangible costs of the intervention group will be estimated taking into account the risk to reoffend in both groups (IoD (IG) / IoD (KG)). The ratio of the risk to reoffend in the intervention and control group will be multiplied by the tangible costs of the control group, assuming that if the risk to reoffend is lower in the intervention group, the costs will be decreased respectively: tangible costs of intervention group = tangible costs of control group * (IoD (IG)/IoD (KG)). The difference in costs between groups represents the tangible benefits: tangible benefits = tangible costs control group – tangible costs intervention group. Intangible benefits will be calculated by analyzing the data from the CVM survey. Within the CVM survey, the question nearest to assessed risk to reoffend (IoD) will be used as the basis of data. Then, the medium of all answers will be calculated and used for the main analysis. Lower and upper bound will be calculated and used for sensitivity analysis. In order to consider the intervention costs from a societal perspective, the costs which occurred per person during the clinical trial will be multiplied by the potential number of convicted child sexual offenders that use the intervention.

5.3 Missing values

In the main analysis missing values will not be imputed. Reference based multiple imputation will be used in a sensitivity analysis for the primary endpoint IoD.

5.4 Additional analyses

Additional analyses will be conducted on an exploratory basis. In particular the analysis of the classical IoD will be complemented by analysis of the reduced IoD.

5.5 Harms – Safety endpoints

Safety data are summarized using descriptive statistics. In particular proportions of occurrences of reoffence and concrete preparation for or intention to recidivism are estimated with 95% confidence intervals for both groups using the beta product confidence procedure for right censored data (Fay, 2013) and differences in the proportions between both groups are estimated

with 95% confidence intervals using the beta product confidence procedure for right censored data (Fay, 2014).

5.6 Statistical Software

The statistical analysis is conducted with R (version 4.2.3 or higher).

6 References

Nitsche, K., Etzler, S., Tröger, A.-S. & Rettenberger, M. (2022). Erfassung akut- und stabildynamischer Risikofaktoren sowie devianten und delinquenten Verhaltens bei Pädosexualität: Manual zur Verwendung von Acute-SR, Stable-SR und CMC (Berichte und Materialien, Band 34). Wiesbaden: Kriminologische Zentralstelle (KrimZ). Verfügbar unter: https://www.krimz.de/fileadmin/dateiablage/E-Publikationen/BM-Online/bm-online34.pdf

Hanson RK, Harris A. *Acute-2007 Scoring Guide*. Ottawa, ON: Public Safety Canada (2007).

Casey S, Day A, Howells K, Ward T. Assessing suitability for offender rehabilitation: development and validation of the treatment readiness questionnaire. *Criminal Justice Behav*. (2007) 34:1427–40. doi: 10.1177/0093854807305827

Rollnick S, Heather N, Gold R, Hall W. Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. *Br J Addict*. (1992) 87:743–54. doi: 10.1111/j.1360-0443.1992.tb02720.x

Hannöver W, Thyrian JR, Hapke U, Rumpf HJ, Meyer C, John U. The readiness to change questionnaire in subjects with hazardous alcohol consumption, alcohol misuse and dependence in a general population survey. *Alcohol Alcohol*. (2002) 37:362–9. doi: 10.1093/alcalc/37.4.362

Merz J. A Questionnaire for the measurement of psychological reactance. *Diagnostica*. (1983) 29:75–82.

Herzberg PY. Zur psychometrischen Optimierung einer Reaktanzskala mittels klassischer und IRT-basierter Analysemethoden. *Diagnostica*. (2002) 48:163–71. doi: 10.1026//0012-1924.48.4.163

Dunkel D, Antretter E, Fröhlich-Walser S, Haring C. Evaluation of the short-form social support questionnaire (SOZU-K-22) in clinical and non-clinical samples. *Psychother Psychosom Med Psychol*. (2005) 55:266–77. doi: 10.1055/s-2004-834746

Sommer G, Fydrich T. Soziale unterstützung. *Diagnostik, Konzepte, Fragebogen F-SozU*. Tübingen: Deutsche Gesellschaft für Verhaltenstherapie (1989).

Russell DW. UCLA loneliness scale (version 3): reliability, validity, and factor structure. *J Pers Assess*. (1996) 66:20–40. doi: 10.1207/s15327752jpa6601_2

Bilsky W, Hosser D. Social support and loneliness: psychometric comparison of two scales based on a nationwide representative survey. *Z Differ Diagn Psychol*. (1998) 19:130–44.

Neutze J, Grundmann D, Scherner G, Beier KM. Undetected and detected child sexual abuse and child pornography offenders. *Int J Law Psychiatry*. (2012) 35:168–75. doi: 10.1016/j.ijlp.2012.02.004

Bumby KM. Assessing the cognitive distortions of child molesters and rapists: Development and validation of the MOLEST and RAPE scales. Sexual Abuse: A Journal of Research and Treatment 8.1, (1996) 37–54.

Feelgood S, Schaefer G, Hoyer J. *KV-M-Skala Zur Erfassung Kognitiver Verzerrungen Bei Missbrauchern* Dresden: Technische Universitä. (2009).

D'Zurilla TJ, Nezu AM, Maydeu-Oliveras A. *Manual for the Social Problem-Solving Inventory-Revised.* North Tonawanda, NY: Multi-Health Systems (2002).

Nezu CM, Nezu AM, Dudek JA, Peacock MA, Stoll JG. Social problem-solving correlates of sexual deviancy and aggression among adult child molesters. J Sexual Aggress. (2005) 11:27–36. doi: 10.1080/13552600512331329540

Graf A. Psychometrische Überprüfung einer deutschsprachigen Übersetzung des SPSI-R. *Z Differ Diagn Psychol.* (2003) 24:277–91. doi: 10.1024/0170-1789.24.4.277

Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. *J Psychopathol Behav Assess.* (2004) 26:41–54. doi: 10.1023/B:JOBA.0000007455.08539.94

Ehring T, Fischer S, Schnülle J, Bösterling A, Tuschen-Caffier B. Characteristics of emotion regulation in recovered depressed versus never depressed individuals. *Pers Indiv Differ.* (2008) 44:1574–84. doi: 10.1016/j.paid.2008.01.013

Scherer A, Eberle N, Boecker M, Vögele C, Gauggel S, Forkmann T. The negative affect repair questionnaire: factor analysis and psychometric evaluation in three samples. *BMC Psychiatry.* (2013) 13:16. doi: 10.1186/1471-244X-13-16

Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. *J Clin Psychol.* (1995) 51:768–74.

Prochaska, J., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist*, *47*(9), 1102-1114.

Meule A, Vögele C, Kübler A. Psychometrische Evaluation Der Deutschen Barratt Impulsiveness Scale–Kurzversion (BIS-15). *Diagnostica.* (2011) 57:126–33. doi: 10.1037/t63911-000

Cortoni F, Marshall WL. Sex as a coping strategy and its relationship to juvenile sexual history and intimacy in sexual offenders. *Sexual Abuse.* (2001) 13:27–43. doi: 10.1177/107906320101300104

Beier KM, Grundmann D, Kuhle LF, Scherner G, Konrad A, Amelung T. The German Dunkelfeld project: a pilot study to prevent child sexual abuse and the use of child abusive images. J Sexual Med. (2015) 12:529–42. doi: 10.1111/jsm.12785

Klein V, Rettenberger M, Boom KD, Briken P. Eine Validierungsstudie der deutschen Version des Hypersexual Behavior Inventory (HBI). *Psychother Psychosom Med Psychol.* (2014) 64:136–40. doi: 10.1055/s-0033-1357133

Mack C, Yundina E. "Emotionale Kongruenz mit der Kinderwelt als mögliches diagnostisches Merkmal von Pädophilie?" In Müller JM, Rösler M, Briken P, Fromberger P, Jordan K, editors. *EFPP Jahrbuch 2012 – Empirische Forschung in der Forensischen Psychiatrie, Psychologie und Psychotherapie*, (2012) p. 71–81.

Tozdan S, Jakob C, Schuhmann P, Budde M, Briken P. Spezifische Selbstwirksamkeit zur Beeinflussung des sexuellen Interesses an Kindern (SSIK): Konstruktion und Validierung eines

Messinstruments. *Psychother Psychosom Med Psychol.* (2015) 65:345–52. doi: 10.1055/s-0035-1548842

Tozdan S, Briken P. The earlier, the worse? Age of onset of sexual interest in children. *J Sexual Med.* (2015) 12:1602–8. doi: 10.1111/jsm.12927

Tozdan S, Briken P. Age of onset and its correlates in men with sexual interest in children. *Sexual Med.* (2019) 7:61–71. doi: 10.1016/j.esxm.2018.10.004

Banse R, Schmidt AF, Clarbour J. Indirect measures of sexual interest in child sex offenders: a multimethod approach. *Criminal Justice Behav.* (2010) 37:319–35. doi: 10.1177/0093854809357598

Briken P. Sexual Outlet Inventory Revised. *Institute for Sex Research and Forensic Psychiatry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*, (2010).

Wolbers, Marcel, Alessandro Noci, Paul Delmar, Craig Gower-Page, Sean Yiu, and Jonathan W Bartlett. 2022. “Standard and Reference-Based Conditional Mean Imputation.” Pharmaceutical Statistics 21 (6): 1246–57.

Fay MP, Brittain EH, Proschan MA. Pointwise confidence inten/als for a survival distribution with small samples or heavy censoring, Biostatistics. 2013;14:723-36.

Fay MP, Proschan MA, Brittain E. Combining one-sample confidence procedures for inference in the two-sample case. Biometrics. 2014;71:146-56