Scaling Up Maternal Mental healthcare by Increasing access to Treatment (SUMMIT): Statistical Analysis Plan

Version Draft 2.3 June 16, 2023

Funding. This study is funded as a Pragmatic Clinical Study (PCS) by the Patient-Centered Outcomes Research Institute (PCORI), PCS-2018C1-10621.

ClinicalTrials.gov ID. NCT04153864

Trial Sponsor. Sinai Health
Contact Name: **Daisy R. Singla**
600 University Ave., #914A
Toronto, Ontario
M5G 1X5 Canada
Telephone: (416) 562-4438
Email: daisy.singla@utoronto.ca

CONTENTS

COMMON ACRONYMS

AE	Adverse Event
BA	Behavioral Activation
EPDS	Edinburgh Postnatal Depression Scale
GAD-7	General Anxiety Disorder – 7
IP	In Person
NIM	Non-inferiority Margin
NSP	Non-Specialist Provider
PCL-6	Abbreviated PTSD Checklist
PCORI	Patient-Centered Outcomes Research Institute
PHQ-9	Patient Health Questionnaire – 9
Q-HAP	Quality of Healthy Activity Program
SAE	Serious Adverse Event
SD	Standard Deviation
SP	Specialist Provider
SUMMIT	Scaling Up Maternal Mental health care by Increasing access to Treatment
TM	Telemedicine

1. Description of the Trial

The overarching goal of the Scaling Up Maternal Mental health care by Increasing access to Treatment (SUMMIT) trial is to examine the scalability of patient-centered provision of brief, evidence-based psychological treatments for perinatal depression and anxiety (N=1,226)[1]. Specifically, and through a multi-site, randomized, non-inferiority trial, the trial examines whether a brief, behavioral activation (BA) treatment delivered via telemedicine is as effective as the same treatment delivered in-person; and whether BA delivered by non-specialist providers (nurses, midwives, etc. with no previous mental health training) with appropriate training is as effective as when delivered by specialist providers (psychiatrists, psychologists and social workers) in reducing perinatal depressive and anxiety symptoms. The study is being conducted in Toronto, Chicago and Chapel Hill. The trial will also identify relevant underlying implementation processes and determine whether, and to what extent, these strategies work differentially for certain women compared to others.

1.1. Principal Research Objectives

The primary objectives of this trial are to:

- Examine if a brief, BA psychological treatment delivered by non-specialist providers (NSP) is as effective in treating perinatal depressive symptoms as specialist-delivered treatment* (*Primary Aim 1*); and
- Examine if a brief BA psychological treatment delivered through telemedicine (TM) is as effective in treating perinatal depressive symptoms as in-person treatment (IP; *Primary Aim 2*)*.

*Note: 'as effective' is the language that we used in the PCORI submission and what PCORI would like to see when they review this document. After consultations with several statistical experts, we will cater the language accordingly for the audience e.g., use 'non-inferior' for academic audiences and the current language for lay audiences.

The primary hypotheses are that among mothers with depressive and anxiety symptoms, psychological treatment delivered by NSPs will be as effective as treatment delivered by specialist providers. In addition, psychological treatment delivered via TM will be as effective as in-person treatment at 3-months, post randomization.

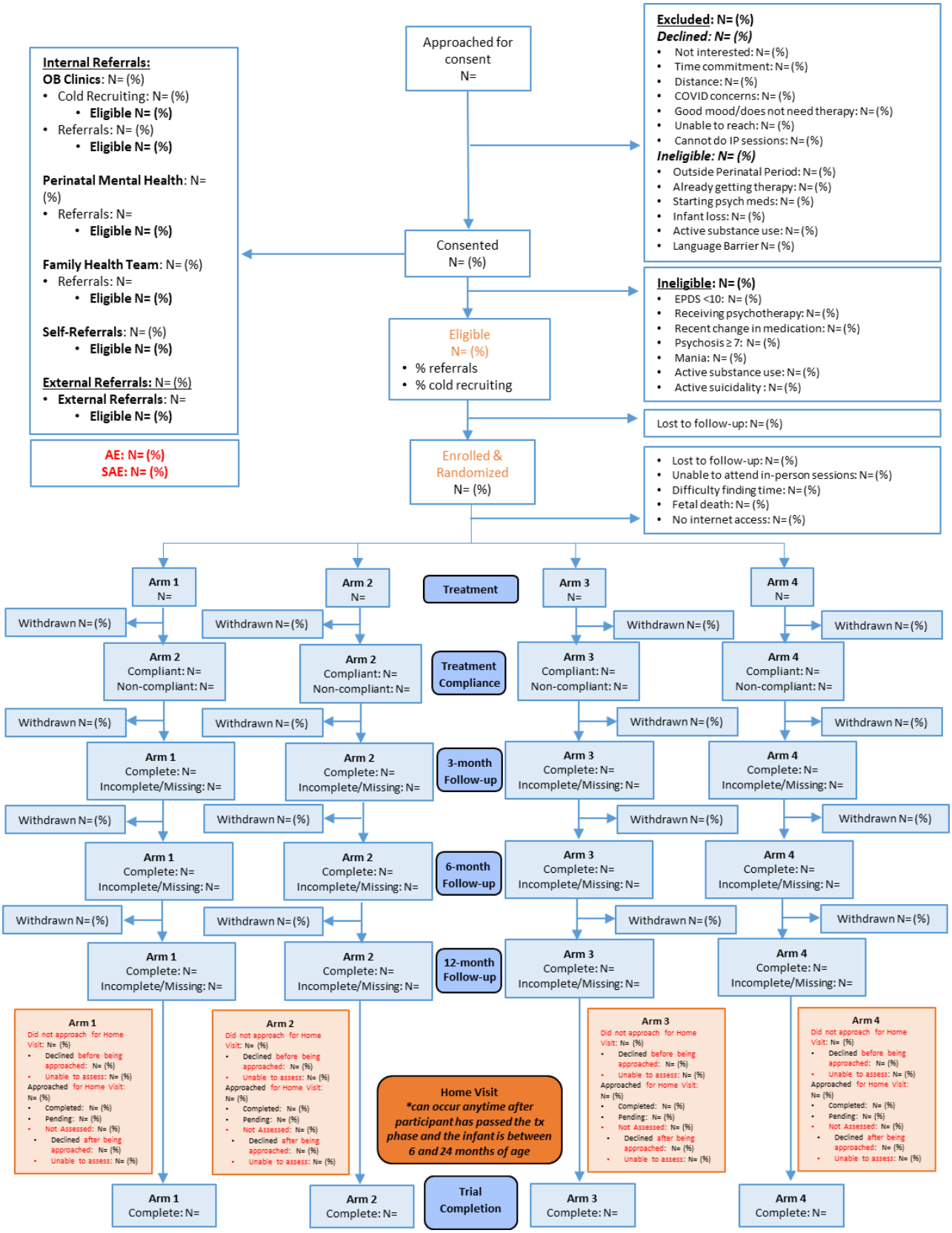
The secondary objectives are to:

- Examine the aforementioned questions for anxiety symptoms at 3-months post randomization (*Secondary Aim 1*);
- Assess moderating effects of clinical severity (mild, moderate and severe) on the comparative effectiveness of the two delivery modes on depressive and anxiety symptoms at 3-, 6- and 12-months post randomization (*Secondary Aim 2*);
- Explore whether the timing of the treatment (antenatal vs. postnatal) influences depressive and anxiety symptoms at 12-months post-randomization, and separately, on child mental development at 6 to 24 months post childbirth (*Secondary Aim 3*); and
- Conduct a process evaluation, i.e., identify the underlying processes related to delivery and scalability of a brief psychological treatment for perinatal depressive and anxiety symptoms from a multi-stakeholder perspective, including relevant barriers and facilitators (*Secondary Aim 4*).

1.2. Trial Design

This is a multi-site, randomized, non-inferiority trial examining the delivery mode of a brief evidence-based BA for perinatal depressive and anxiety symptoms, and to determine the underlying processes related to delivery and scalability of the psychological treatment from a multi-stakeholder perspective.

Figure 1 illustrates the participant recruitment and follow-up assessments. All analyses will consider site as a potential co-variate.



*Note: The trial flow chart may be modified in the final publication versions, including the format and the reasons for exclusion and ineligibility.

1.3. Eligibility Criteria

1.3.1. Trial Participants

Pregnant and postpartum women[†] are recruited from three study Hubs (Toronto, Chapel Hill, Chicago) through their networks of clinics. In Toronto, we are recruiting from three sites: Sinai Health, Women's College Hospital and St. Michael's Hospital (referrals-only site). In the US, we are recruiting from the University of North Carolina in Chapel Hill, and NorthShore University HealthSystem in Evanston and surrounding areas including Chicago. Inclusion and exclusion criteria are detailed in Table 1.

Table 1: Inclusion and exclusion criteria

INCLUSION	EXCLUSION
• EPDS≥10 • ≥18 years • Pregnant up to 36 weeks or 4 to 30 weeks postpartum • Speaks English or (US sites) Spanish	• Active suicidal intent (ideation and plan), active symptoms of psychosis or mania • Psychotropic medication dose or medication change within two weeks of enrollment or beginning treatment • Ongoing psychotherapy (no more than once every 8 weeks or during the duration of the intervention) • Active substance abuse or dependence • Severe fetal anomalies, stillbirth or infant death at time of enrollment for index pregnancy • Non-English, non-Spanish speakers

1.3.2. Non-specialist Providers (NSPs)

NSPs are healthcare workers with general health care professional skills (assessed during recruitment) but without formal training in mental health care or previous experience delivering psychological treatments.

1.3.3. Specialist Providers (SPs)

SPs are individuals with formal training in mental health care delivery (e.g., psychiatrists, psychologists and social workers) with experience in treating perinatal mental illness and a minimum of 5 years of experience delivering psychological treatments.

1.4. Outcome Assessments

We used the following definitions to define outcomes[2, 3]:

- **Primary Outcome:** the main outcomes of interest for the SUMMIT trial.
- **Secondary Outcome:** any outcome that is not a primary outcome but supports the primary outcome by corroborating results or by explaining a mechanism (mediators and moderators).
- **Exploratory Outcome:** variables of interest which can serve as a basis for new directions but may not be relevant to the primary outcome or differ between the treatment arms.

[†]Pregnant and postpartum women refer to women and other persons who are pregnant or postpartum.

Outcome data will be collected at 3-, 6- and 12-months post randomization and 6-24 months post childbirth. Table 2 lists primary and secondary outcomes and their assessment points. All measures proposed in the current study have been previously used and validated in one or more of the investigators' trials and selected because of their potential role in the presumed causal pathway (Figure 3). Our emphasis on mothers' self-reported data adheres to PCORI's methodology standards[4] that emphasize that the patient population is the best source of information.

Additionally, to address the aim of conducting a process evaluation, qualitative data is collected throughout the phases of the trial from perinatal participants, significant others, treatment providers, clinical leads, and stakeholders.

Due to the COVID-19 pandemic, and in line with our published study protocol[1], several measures were added at various assessment points following the start of data collection. These include:

- ***treatment preference*** (TM or IP; added April 6, 2020, n=30 missed* the measure at baseline; 2.4% of final sample size),
- ***trauma symptoms*** (PCL-6; added April 6, 2020, n=30 missed* the measure at baseline; 2.4% of final sample size),
- ***quality of life*** (EQ-5D5L; added April 6, 2020, n=30 missed* the measure at baseline; 2.4% of final sample size), and
- ***COVID-19 exposure*** (added June 26, 2020, n=170 missed* the measure at baseline; 13.9% of final sample size).

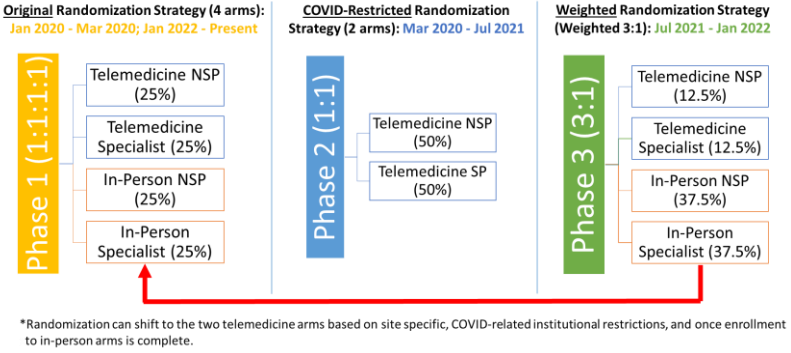
*Note: Multiple imputation methods will be used to impute missing data and compare model results to the complete case analysis. See Missing Data and Imputation.

In addition, treatment providers and a random subset of perinatal participants and stakeholders were asked to participate in a one-time qualitative interview to examine their experiences during COVID-19, along with perceived barriers and facilitators related to resuming IP treatment sessions[5].

1.5. Randomization

All SUMMIT participants are stratified by perinatal period (antenatal vs. postnatal) and were initially (Phase 1) randomized within site to one of four arms (see Figure 2). However, during the early part of the COVID-19 pandemic (Phase 2) when in-person care was prohibited, participants were randomized to only one of two arms: TM NSP and TM SP (Figure 2). All sites resumed randomization to all 4 arms 16 months later (Phase 3). Given that Phase 2 consisted of randomizing participants exclusively to TM, there was an imbalance in participant assignment between the IP and TM arms. To address this, a weighted randomization approach (3:1, favoring IP) was used in Phase 3 in which a reduced percentage of

Figure 2. Randomization strategy



participants were randomized to the TM arms to ensure that the study's sample size targets will be met, and the final sample has an equal number of IP and TM participants. Due to the Omicron wave in December 2021 and January 2022, participants were randomized to TM-only again and randomization switched back to a 1:1 ratio. However, sites were permitted to switch between a 1:1 TM-only (2 arm) and a 1:1:1:1 both TM and IP (4 arm) randomization scheme based on site-specific COVID restrictions and Infection Prevention and Control (IPAC) recommendations. Sites began resuming Phase 1 randomization to IP in the 1:1:1:1 ratio in January 2022, and all sites returned to Phase 1 randomization by April 2022. Randomization to all 4 arms will be followed until in-person arms are fully enrolled after which randomization 1:1 to the two telemedicine arms will be followed.

Table 2: SUMMIT Outcomes (A more detailed list can be found in Appendix 1).

Study Variable	Instrument	Outcome (Range)
Maternal: Measured at Baseline and 3*-, 6*- and 12*-months post-randomization		
Maternal Characteristics**	**Trial Baseline Questionnaire[6, 7]**	Self-reported age, education level, gender identity and sexual orientation, marital status, immigrant status and ethnicity, clinical history with depression or anxiety (severity, chronicity, number of prior episodes, and age at first episode), occupational status, number of children, pregnancy intention, pregnancy history, delivery and birth.
Depressive Symptoms¥	**Edinburgh Postnatal Depression Scale (EPDS)[8, 9]**	Mean continuous score of a 10-item scale (0-30). Severity ranges: none or minimal depression (0-9), mild (10-11), moderate (12-19), severe (>19)
Anxiety Symptoms	**Generalized Anxiety Disorder Scale (GAD-7)[10, 11]**	Mean continuous score of a 7-item scale (0-21). Cut-off value (10)
Response & Remission	**Patient Health Questionnaire-9 (PHQ-9)[12]**	Response: PHQ<10 Remission is defined as PHQ<5
Perceived Support	**Multidimensional Scale of Perceived Social Support (MSPSS)[13]**	Mean continuous score of a 12-item scale (1-84)
Disability Assessment	**World Health Organization Disability Assessment Schedule (WHODAS)[14]**	Mean continuous score of a 12-item scale (0 – 48)
Quality of Life Assessment	**EQ5D-5 Level (EQ5D-5L)[15]**	Mean continuous score of a 5-item scale (1-25)
Trauma Symptoms	**Abbreviated PTSD Checklist (PCL-6)[16-18]**	Mean continuous score of a 6-item scale (1-30). Cut-off value [14 (threshold) and 8 (subthreshold)]
Patient-Reported Activation	**Premium Abbreviated Activation Scale[19, 20]**	Mean continuous score of a 5-item scale (0-20)
Patient Satisfaction***	**Client Satisfaction Questionnaire-8 (CSQ-8)[21]**	Mean continuous score of an 8-item scale (1-32)
Therapeutic Alliance***	**Working Alliance Inventory-Short Revise (WAI-SR)[22]**	Mean continuous score of a 12-item scale (1-60)
Health Service Utilization	**Health Service Utilization Questionnaire (HSU-Q)[23]**	Total score of a 16-item scale (0-32)
Treatment Preference**	**Delivery of treatment and treatment provider preference**	Score of 0 or 1
COVID-19 Exposure¥	**1-item question on COVID-19 exposure**	Self-reported
Health Benefits Access and Use	**2-item question on access and use of health benefits**	Self-reported
Treatment: Measured at every session during treatment, unless otherwise indicated		
Dosage	**Treatment Log[24]**	Number of Sessions Attended
Therapy Quality****	**Quality of Healthy Activity Program (Q-HAP)[25]**	Mean continuous score of treatment-specific BA skills (0-4) and general counselling skills (0-4)
Session Depressive / Anxiety	**Session-by-session EPDS[8] and GAD-7[10] scores**	Mean continuous score of a 10-item scale (0-30) on EPDS and of a 7-item scale (0-21) on GAD-7
Homework Adherence	**Treatment Log[24]**	Mean continuous score of a 1-item question (0-2)
Adverse or Serious AEs	**Anytime an Adverse Event (AE) or Serious AE (SAE) occurs**	Any event that represents a serious threat to the safety of the mother or her child (see *Appendix D of protocol*)
Health Service Utilization	**Health Service Utilization Questionnaire (HSU-Q)[23]**	Total score of a 16-item scale (0-32)
List of Medications	**List of Medications**	Self-reported list of medications
Child: Measured at 6 to 24-months post-childbirth unless otherwise indicated		
Birth Weight & Length	**Retrieved from hospital chart or self-report†**	Assessed at birth
Breastfeeding	**Whether breastfeeding and if stopped age stopped[7]**	Total Number of Months
Psychosocial Stimulation	**Home Observation Measurement Evaluation (HOME)[26]**	Total score of a 45-item checklist
Child Mental Development	**Bayley Child Mental Development Scales IV[27]**	Mean continuous score of cognitive, receptive and expressive language development

* Assessment period will be extended to account for post-treatment outcomes when there are perinatal-related interruptions to treatment (e.g., giving birth, obstetrical complications, COVID-19); **Only at baseline; ***Measured at 3-months post-randomization only; ****Randomly selected for supervision, rated by self, peers, expert supervisor; ¥Also measured during treatment; †Self-report will be used when hospital charts are external to the recruiting site.

1.6. Primary Sample Size Calculations

The hardships of the COVID pandemic and its subsequent impact on in-person patient care resulted in our revision of our initial sample size approach. Specifically: 1) based on recent guidelines[28, 29], we will not adjust our two primary hypotheses for multiplicity as they do not involve different endpoints (i.e., both hypotheses test EPDS scores). Past references[28, 29] have indicated when multiple hypotheses test the similar underlying outcomes, no adjustment for multiplicity is required. Unlike superiority analyses which determine success based on p-values, non-inferiority analyses determine success based on the pre-specified non-inferiority margin and the confidence interval around the difference (outlined in 3.4.1); and 2) The non-inferiority margin for in-person compared to telemedicine has been changed to 13% (1.6.2.1).

1.6.1. Primary Aim 1 (Specialist vs. Non-Specialist) Comparison

The primary outcome measure is an EPDS mean score at 3 months post-randomization. The sample size calculation is based on an EPDS mean estimate of 7.93 (SD=4.68)[30]. Using a non-inferiority margin of 10% (i.e., EPDS score of 0.79 in relation to the mean), and an alpha=0.05, we require 431 participants in each of the two groups (SP, NSP) to provide greater than 80% power (Table 3). To account for 10% drop out, the sample size is inflated to **N=958** (479 per group, SP vs. NSP). All sample size calculations were run using PASS Version 12[31].

Table 3: Power Analysis of a Non-Inferiority Test of the Difference of Two Means

Power	N1/N2	Non-Inferiority Margin (NIM)	Actual Difference	Significance Level (Alpha)	Beta	Standard Deviation 1 (SD1)	Standard Deviation 2 (SD2)
0.80022	**431/431**	**0.793**	**0.000**	**0.05000**	**0.19978**	**4.680**	**4.680**

1.6.2. Primary Aim 2 (Telemedicine vs. In-Person) Comparison

The primary outcome measure is an EPDS mean score at 3 months post-randomization. The sample size calculation is based on an EPDS mean estimate of 7.93 (SD=4.68)[30]. Using a non-inferiority margin of 13% (i.e., EPDS score of 1.03 in relation to the mean), and an alpha=0.05, we require an additional 268 IP participants (Table 4).

Table 4: Non-inferiority margin in relation to sample size for Phase 3 comparisons

Non-inferiority margin	Power	Sample size per primary question	Total sample size + 10% dropout*	Total sample size + 20% dropout
10%	80%	431	958	1078
13%	**80%**	**241**	**536 (268 per group)**	**604**
14%	80%	221	492 (246 per group)	556

*Current dropout rates are <10% at primary outcome of 3-months post-randomization

When combined with the 958 TM participants, this yields a total study sample size of N=1,226 which accounts for 10% loss to follow-up (Table 5).

Table 5: Primary aims and required sample sizes

Primary Aim¥	Non-inferiority Margin	Power	Required sample size + 10% drop out
SP vs. NSP¥	**10%**	**88%**	**958 telemedicine**
IP vs. Telemedicine	**13%**	**80%**	**268 in-person***
	Total		**N=1226***

¥We acknowledge that the in-person sample is contained in the first hypothesis. This will result in the SP vs NSP comparison having up to 88% power. *We will over-recruit to account for protocol deviations where participants were switched from in-person to telemedicine due to the COVID-19 pandemic.

1.6.2.1 Justification of Revised Non-inferiority Margin

Pandemic-related disruptions have resulted in difficulties with recruitment to the IP arms. To account for this, the non-inferiority margin for the TM vs. IP comparison was increased from 10% to 13%, thereby reducing the required IP sample size. This increase is justified by the literature. We only identified one non-inferiority trial that used the EPDS and a non-inferiority margin of 15%[32]. We did find that a non-inferiority psychotherapy trial using a PHQ-9 of 1.9[33], which is comparable to the EPDS[34], with similar sensitivity among numerous perinatal populations. The justification of a clinically meaningful non-inferiority margin is based on the following approach by Richards' et al COBRA Trial (2017)[33]:

1. Start off with a comparison of superiority of the one treatment to control in the past literature. They looked at a meta-analysis of their treatment of interest versus control.
2. Take half of this effect size, as they indicate with the sentence: "Previous research has suggested that non-inferiority margins should be half of the mean controlled effect size from historical trials."
3. From their meta-analysis they conclude that treatment is superior to controls by 0.7 SD units.
4. Convert this 0.7 SD units to an actual PHQ-9 value (= 3.8 PHQ-9 units).
5. Take half of this as suggested in step 2 - this is how they arrive at 1.9 PHQ-9 units.

So, if we follow this logic and apply it to our EPDS mean estimate of 7.93 (SD=4.68):

1. 0.7 SD units translates to 3.28.
2. Half of 3.28 is EPDS=1.64.

Applying this approach, our 13% margin of non-inferiority (which is 1.03) falls under this value.

2. Variables

A detailed list of variables can be found in Appendix 1. Additionally, the Data Dictionary Codebook is available from the Data team.

3. Data analysis plan

3.1. Recruitment and Representativeness of Recruited Participants

Initial analyses will include examining the number of:

- patients that declined to consent and reasons,
- participants that consented, and
- eligible participants.

Analyses will also be conducted by intervention arm, examining the number of:

- participants allocated to each arm,
- participants that completed treatment and treatment compliance, and
- participants that completed the 3-, 6- and 12-months post-treatment assessments.

Recruitment data and representativeness of recruited participants are collected in the CONSORT flow chart. Participants are consented and randomized to one of the four arms at each of the 5 sites [Sinai Health, Women's College Hospital, St. Michael's Hospital (referral-site only), UNC, NorthShore] within the 3 Hubs (Toronto, Chicago, Chapel Hill). However, given the availability of treatment providers, patients may be referred to another site to enroll and randomize (in

Toronto, only). In this event, the collected data will be adjusted for site correlations due to potential site differences.

3.2. Withdrawals or Dropouts and Other Missing Data

The number and proportion of participants that are withdrawn from the trial or discontinue treatment will be reported overall and across arms at the following time points: screening, enrollment, treatment, and post-treatment at 3, 6 and 12 months. The reasons for withdrawal from the trial and discontinuing treatment will also be documented.

3.3. Adverse and Serious Adverse Event Reporting

Adverse events (AE) and serious adverse events (SAE) will be summarized (proportion of individuals with each type of AE/SAE, and total number of AEs/SAEs) by arm. Based on the final number of AEs/SAEs, the risks and 95% confidence intervals (CIs) may be reported, and the risks will be compared across intervention arms.

3.4. Outcome analysis

3.4.1. Primary Analysis

All analyses will be run as both intent-to-treat (ITT, i.e., group that the participant was randomized to) and per protocol (i.e., group participant actually participated in*). SAS Version 9.4[35] or later will be used for all analyses. For non-inferiority analyses, a successful result will be based on the upper bound of the 95% confidence interval being less than or equal to the non-inferiority margin. Based on recent guidelines[28, 29], we will not adjust our two primary hypotheses for multiplicity as they do not involve different endpoints (i.e. both hypotheses test EPDS scores). For secondary superiority hypotheses, a two-sided significance level of $p<0.05$ will be used to denote statistical significance. Descriptive statistics will be calculated for all variables of interest. Continuous measures such as age and depressive symptoms will be summarized using means and standard deviations, whereas categorical measures will be summarized using counts and percentages.

*Note: The trial has n=21 protocol deviations. This number has been consistent since the last major COVID wave (March 2022).

Demographic (e.g., ethnicity, age, marital status) and other baseline variables (e.g., severity and chronicity) will be compared for potential differences between study groups (TM vs IP, and SP vs NSP) using two sample two sided t-tests for continuous variables (or Wilcoxon rank sum test in the case of non-normal data), and chi-square analyses for categorical variables (Fisher's exact tests in the case of low expected cell sizes). Those who withdrew from the trial will be compared to those who continued on baseline indices including maternal education and occupation.

The primary outcome of non-inferiority in EPDS scores will be compared between NSP vs. SP and TM vs. IP groups at 3-months using a t-test with a 10% and 13% margins of non-inferiority, respectively.

Primary Aim 1: Examine if a brief, BA psychological treatment delivered by non-specialist providers (NSP) is as effective in treating perinatal depressive symptoms as specialist-delivered treatment.

Aim 1 is to show that NSP is non-inferior to SP (N=479 per group, SPs vs NSPs, Table 3).

Primary Aim 2: Examine if a brief BA psychological treatment delivered through telemedicine (TM) is as effective in treating perinatal depressive symptoms as in-person treatment (IP).
Aim 2 is to show that TM is non-inferior to in-person (N=268 per group, IP vs TM, Table 4).
Aim 1 and Aim 2 will initially be analyzed using a non-inferiority t-test (under the assumption that randomization will serve to balance out potential confounders). Each t-test will look at the confidence interval around the difference in EPDS scores (say between TM and IP) and see if the upper bound contains the non-inferiority margin (upper bound for the case in which we take the difference to be TM minus IP). In addition, we will run a linear regression model with mode, agent, and a mode by agent interaction term to assess whether an interaction between mode and agent exists. In the unlikely event that randomization was not able to reduce bias by balancing out confounders, we will adjust for variables found to be imbalanced at baseline.

Rules on the number of variables allowed in a multivariable model will be followed. The rule on number of variables in a **linear** regression is to take the total number of observations divided by 10. Given our trial sample size, we do not have any concerns with overfitting our linear model for our **covariates of interest** (education level, marital status, ethnicity, baseline severity and chronicity, timing of treatment, compliance, perceived support, and medication). For the **logistic** regression analyses, the number of predictors will be based on the smaller of the two outcome categories divided by 10[36].

3.4.1.1. COVID-related Adjustments (Phase Analysis)
In order to determine whether the participants recruited in Phase 2 can be combined with those in Phase 3 in the larger trial, we will examine a potential interaction between group (SP, NSP) and phase (2 and 3) in relation to change in EPDS and GAD scores. To examine this, a linear mixed model will be run for the outcome EPDS and another for the outcome GAD. Each model will contain a group (SP, NSP), Phase (2 and 3), and a group by phase interaction term. A comparison of Phase 1 (pre COVID phase) to Phase 3 (return to IP phase) will not be carried out as the sample size in the first phase (n=23) is too small to detect statistical differences; however, we will conduct a sensitivity analysis on our primary outcome (EPDS scores at 3-month post-randomization) which includes this portion of the sample.

The study analyses will be carried out as originally planned on the entire sample of participants across all 3 phases. Should we find a group by phase interaction, we will also run separate analyses by phase (2 and 3) and compare their results to the model on the entire sample.

3.4.2. Secondary Analysis

3.4.2.1. Analysis for Secondary Aim 1: Anxiety Symptoms
In our secondary analysis, we are interested in **examining the primary questions for anxiety symptoms at 3-months post randomization.**

This aim is to show that NSP is non-inferior to SP. One t-test will be run to assess this and compare agents (is NSP non-inferior to SP). This aim is to also show that TM is non-inferior to in-person. Another t-test will be run to assess this and compare modes (is TM non-inferior to IP). We will require a sample of 774 participants and use a NIM of 10% for NSP compared to SP which translates to a mean EPDS difference of 0.871 based on the mean of 8.71 (SD=4.61)[37] and 10% dropout. On the other hand, a sample of 460 participants and a NIM of 13% will be required for TM compared to IP which translates to a mean difference of 1.142 based on the

mean of 8.71 (SD=4.61)[37] and 10% dropout. In addition, we will run a linear regression model with mode, agent, and a mode by agent interaction term to assess whether an interaction between mode and agent exists.

No interim analyses will be carried out.

3.4.2.2. Trajectory of Depressive and Anxiety Symptoms over Time

We will also examine the trajectory in depressive and anxiety scores over time (baseline, 3-, 6- and 12-months post-randomization), using the EPDS and GAD-7, respectively. To assess change in EPDS scores over time between antenatal and postnatal, a linear mixed model, including the interaction between group and time, will be run. Similar models will be run to compare anxiety scores between the antenatal and postnatal groups as were carried out for depression.

3.4.2.3. Analysis for Secondary Aim 2: Clinical Severity

We are also interested in **assessing moderating effects of clinical severity (mild, moderate and severe;** see Appendix 2 for cut-off values and severity classification for outcomes) on the comparative effectiveness of the two delivery modes on depressive and anxiety symptoms at 3-, 6- and 12-months post randomization**.**

This will involve the use of linear mixed models. Using a non-inferiority margin of 10%, 10% dropout, and an alpha of 0.05, we require 44 participants[38]. Mothers will be taken as a random effect and the models will include a treatment-by-time interaction term. Other potential moderators will be explored such as age, ante or postnatal enrollment, patient preference, white/BIPOC, and trauma symptoms. The same set of analyses as above will be conducted for anxiety symptoms (GAD-7), disability scores, quality of life, and client satisfaction.

3.4.2.4. Analysis for Secondary Aim 3: Perinatal Period and Child Development

Another secondary aim is to **explore whether the timing of the treatment (antenatal vs. postnatal) influences child mental development at 6 to 24 months post childbirth.**

All child outcomes, including child mental development and the provision of psychosocial stimulation by the mother, will be compared between two groups (antenatal vs. postnatal) at 6-24 months post childbirth using a two-sample two-sided t-test. We will require a sample of 393 participants in each of the two groups to detect a mean clinically-significant change of 3.0 units with 80% power and an alpha of 0.05. Similar comparisons will be run between the primary comparison arms (SP vs NSP; in-person vs telemedicine). Sensitivity analyses will be carried out excluding those who dropped out prior and after first session to see if the results are comparable to the entire study group. We will adjust for time since treatment by including it as a variable in the regression model comparing antenatal vs. postnatal.

We will examine the hypothesis of whether the subset of mothers (up to 75% of the sample) who receive the antenatal treatment will benefit more in terms of improved child outcomes at 6-24 months than mothers who receive postnatal treatment. A 2-group comparison of 393 per group (antenatal vs. postnatal) and an assumed mean on any Bayley IV[27] raw subscale score of 100, SD=15 to assess child mental development will provide 80% power, with an alpha of 0.05 to detect a mean clinically-significant change of 3.0 units. Raw and standardized Bayley IV scores of

the two groups' children (antenatal vs. postnatal) at 12 months will be compared using a two-sided, two-sample t-test.

3.4.2.5. Analysis for Secondary Aim 3: Perinatal Period and Perinatal Depression and Anxiety

We are also **exploring whether the timing of the treatment (antenatal vs. postnatal) influences depressive and anxiety symptoms at 12-months post-randomization.**

This model will also test whether expectant mothers who receive antenatal treatment benefit more in terms of reduced depressive symptoms than mothers who receive postnatal treatment at 12-months post-randomization. A 2-group comparison with at least 200 per group and an assumed mean of EPDS score of 7.93 (SD=4.68)[30] will provide 90% power, with an alpha of 0.05, to detect a mean change of 1.5 points. This allows us to detect a drop corresponding to a small effect size of 0.3 (i.e., a drop to mean 6.43 on the EPDS). The two groups' (antenatal vs. postnatal) EPDS scores at 3-months will be compared using a two-sided two-sample t-test. An exploratory analysis will be carried out on this set of patients to test for a group by treatment interaction at 12 months.

3.4.3. Missing Data and Imputation

Multiple imputation methods will be used to impute missing data and compare model results to the complete case analysis. This process will involve assessing the distribution of the original data. Fully conditional specification (FCS) methods will be carried out using SAS's Proc MI and Proc MIANALYZE. The procedure will create five imputed datasets and the model results averaged across the five iterations[39]. If needed, we will also align the number of imputations to be on par with the percent missingness (i.e., if 10% is missing, 10 imputations), as per the recommendation from PCORI's methodological consultant[40, 41]. Linear mixed models will be used to assess repeated measures outcomes. These models use maximum likelihood estimation methods that retain participants who do not have complete data across all time points. Reasons for dropout will be ascertained and we will interview a subset of the participants who dropped out and reasons will be coded accordingly. Sensitivity analyses will be carried out should missing data lead to the use of multiple imputation methods. These analyses will compare the results of the models on the imputed data to the ones with the actual missing data included.

3.4.4. Outliers

All the collected data will be checked for any potential outliers or errors or invalidity after importing to SAS[35]. Any outliers that are acceptable values will be included in the final analysis, but we will consider a possible sensitivity analysis to examine the impact of the outliers.

3.4.5. Treatment Providers

Competency measures including multiple choice exam and role play scores from the training sessions among both types of treatment providers (SP and NSP) will be compared with participant outcomes to determine if competency scores during training predict participant outcomes. In addition, key variables related to therapy quality will include treatment-specific skills (range mean score of 0-4) and general skills (range mean score of 0-4) as measured by the Q-HAP and a total score (range 0-35) of a 35-item multiple choice exam (range total score of 0-35). Means, SDs and ranges of all competency measures for each individual provider and across providers will be calculated. This will be followed by estimating the relations between competency measures by calculating the Pearson correlation. Finally, multiple regression analyses will be used to estimate whether competency measures can predict patient outcome

scores of EPDS and GAD-7 scores post-treatment. Covariates including baseline EPDS scores, and treatment provider will be utilized to account for potential baseline heterogeneity.

3.4.6. Qualitative Data

Qualitative interviews with participants, significant others, providers, clinical leads, and stakeholders will be conducted to address the following secondary aim 4: **To conduct a process evaluation, i.e., identify the underlying processes related to delivery and scalability of a brief psychological treatment for perinatal depressive and anxiety symptoms from a multi-stakeholder perspective including relevant barriers and facilitators.**

All qualitative data will be analyzed using NVivo™, a qualitative data analysis software package. We will use content analysis with data analysis (coding) conducted by multiple independent raters, for whom inter-rater reliability will be calculated using Kappa (κ) scores. A coding index will be developed and finalized, and the data will be coded in a stepwise fashion to facilitate iterative revision. Specifically, there will be a process of first independently coding and then discussing a minimum of 3 cases per stakeholder group to achieve a kappa (κ) score of κ=0.75 or higher (defined as substantial to almost perfect agreement). The qualitative data will then be quantified and triangulated across stakeholder groups using our previously established methods.

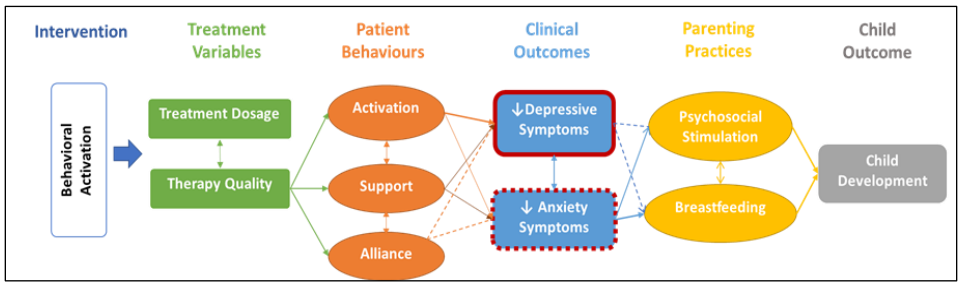
3.4.7. Scalability: An examination of what works (and did not work) for whom

Methods. Once follow-up data (6- and 12-months) have been collected in the SUMMIT Trial, we will attempt to answer the question of 'what works for whom' in order to ultimately identify the optimal intervention strategy for each mother (with the ultimate goal of truly **personalized, patient-centered care**)[42]. To do this, we will develop clinical prediction models (CPMs) using validated machine learning approaches[43-45]. Although the clinical utility of precision treatment rules (PTRs) informed by CPMs was recently demonstrated in a prospective randomized controlled trial[46], the approach has never been tested to identify optimal delivery strategy options for perinatal populations with depressive and anxiety symptoms.

Analysis. Developing PTRs for treatment selection will involve the generation of statistical models that capture both prognostic and prescriptive information to predict expected treatment response in two or more conditions. We will use SuperLearner, an ensemble machine learning method[47, 48] that assigns weights to a set of selected algorithms to develop a consolidated predictive algorithm which optimizes cross-validated *MSE*. Model validation will be performed using bootstrapping, which is the recommended procedure for assessing honest model performance[49]. Recent approaches[50, 51] will be used to evaluate the benefit that would be expected if the final algorithm were to be used to guide treatment selection.

3.4.8. Mediation: An examination of the conceptual causal model behind the workings of BA

Figure 3. Conceptual causal model behind the workings of BA



Methods. We will examine the above conceptual model to show several variables of interest within a causal pathway. Because the current study proposes a non-inferiority trial where the same treatment (BA) will be used in all four arms, we anticipate similar pathways in each arm, irrespective of who or how the treatment will be delivered.

This conceptual model involving treatment and patient variables and clinical outcomes has been used in the psychological treatment literature for depression and anxiety, as well as other disorders[52]. In short, this temporal model proposes that active treatment ingredients (processes that occur during intervention delivery) influence patient behaviors (in this case, improvements in patient-reported activation, perceived support and therapeutic alliance), which in turn influence clinical outcomes (reduced depressive and anxiety symptoms). The proposed model also extends the existing psychological treatment literature to include parenting practices and child development outcomes.

This model is supported by empirical literature demonstrating the mediating effects of activation levels[53, 54], interpersonal supports[55, 56] and therapeutic alliance of BA and BA-based interventions on patient clinical outcomes of depressive and anxiety symptoms. Our own analyses of two recent parallel BA-based treatments demonstrated that both patient activation and interpersonal supports mediated the effects of one BA-based intervention on reduced maternal depressive symptoms[57]. In addition, growing evidence suggests strong positive associations between maternal mental health and parenting practices[58, 59], including the provision of psychosocial stimulation[60] and breastfeeding[61, 62]. We have also found that improved psychosocial stimulation mediated the effects of an integrated parenting and psychological treatment intervention on child mental development[63].

Analysis. We will extend our existing analysis plan to assess the proposed model quantitatively. Using Monte Carlo Methods for Assessing Mediation and structural equation modelling, we will estimate individual and multiple mediating pathways on patient and child outcomes. Unlike the majority of the prior literature, our analysis will follow key guidelines[64-66] such as the assessment of multiple potential mediators, the use of a temporal design with hypothesized mediators being assessed at distinct time-points, a comparison of several active treatment groups with corresponding large sample sizes and adjusting for key variables at baseline (e.g., symptom severity).

3.4.9. Summary List of Potential Sensitivity Analyses

The trial has incorporated numerous sensitivity analyses to account for differences in participants' baseline characteristics, data collection time points and outcome variables such as depression and anxiety scores. Table 6 encompasses a few examples of the sensitivity analyses for this trial.

Table 6. Planned sensitivity analyses

Sensitivity Analysis	Purpose
Starting a new medication or change in medication between randomization and treatment initiation	To compare if the participants that started a new or changed medication were different from the larger sample
Accounting for the GAD-7 variable update from 'not at all sure' to 'not sure'	To determine whether participants that answered to GAD-7 with the 'not at all sure' option differed significantly from the larger sample
Inclusion of participants with SAE after randomization but before starting the treatment in the trial	To determine whether participants that had an SAE before treatment were different from the larger sample
Participants who were lost to follow up at randomization and did not start treatment but were asked to complete the follow up assessments	To include the lost to follow up after randomization before treatment data at 3,6 and 12 month follow up as part of intent-to-treat analysis
Considering number of days after baseline for completion of 3, 6 and 12 month data	To account for the time variability in completion of follow up data
Individuals who indicated yes to marijuana consumption but did not answer the follow up drug screening questions	To determine if participants who were using marijuana prior to updating the follow up marijuana screening question have significantly different outcomes than the remaining sample
Existing differences in Canada, Illinois and North Carolina jurisdiction in legalization of marijuana may impact individuals self-report of marijuana use at screening	To determine if there are differences between North Carolina and other sites in the outcome
Multiple imputation methods will be used to impute missing data and compare model results to the complete case analysis.	Should missing data lead to the use of multiple imputation methods, sensitivity analyses will compare the results of the models on the imputed data to the ones with the actual missing data included.
COVID-related Analyses: The starting models to address COVID-related Aim 1 will compare mean baseline EPDS scores and mean GAD scores between the time periods (Phase 2 vs. Phase 3). The model to address COVID-related Aim 2 will include potential moderators. Using our originally-stated mean on the EPDS=7.93 (SD=4.68)[30], our power calculation indicates that a sample size of n=87 in each time period (N=174 across the entire period; or 184 including a 5% drop-out rate) is required to detect a clinically-meaningful difference of two or more points on the EPDS. This sample size will be used for the same number of participants enrolled during Phase 2 and Phase 3 of the trial (addressing Aim 1). To address COVID-related Aim 2, assuming a correlation of 0.1 between covariates, the sample size will be inflated to 204 (102 per time period). To address COVID-related Aim 3, qualitative analyses using the same methods discussed in section 3.4.7 will be performed. Qualitative data will be analyzed to identify key themes related to barriers and facilitators using content analysis. Both barriers and facilitators will be categorized into internal, interpersonal and structural themes related to intervention delivery, impact and motivation, content and training and supervision.	<u>COVID-related aim 1:</u> Examining whether perinatal participants experience higher levels of depressive and anxiety symptoms during Phase 2 compared to those in Phase 3, whereby; 'Phase 2' is defined as period when IP allocation is treated as TM and 'Phase 3' is defined as the period when participants are randomized to IP and TM allocations by weighted randomization approach. <u>COVID-related aim 2:</u> Examining potential moderators (e.g., treatment preference, trauma symptoms) of COVID-19 on treatment response among perinatal participants. <u>COVID-related aim 3:</u> Exploring barriers and facilitators related to resuming IP treatment sessions from a multi-stakeholder perspective.

References

1. Singla DR, Meltzer-Brody SE, Silver RK, et al. Scaling Up Maternal Mental healthcare by Increasing access to Treatment (SUMMIT) through non-specialist providers and telemedicine: a study protocol for a non-inferiority randomized controlled trial. *Trials* 2021; 22: 186. DOI: 10.1186/s13063-021-05075-1.
2. Vetter TR and EJ M. Defining the Primary Outcomes and Justifying Secondary Outcomes of a Study: Usually, the Fewer, the Better. *Anesth Analg* 2017; 125: 678-681. DOI: 10.1213/ANE.0000000000002224.
3. Kostan H, Chen K and Cheng C. Selecting Useful Outcome Measures. 2020.
4. Institute P-COR. PCORI Methodology Standards, https://www.pcori.org/research/about-our-research/research-methodology/pcori-methodology-standards (2019, accessed June 7, 2023).
5. Andrejek N, Hossain S, Schoueri-Mychasiw N, et al. Barriers and Facilitators to Resuming In-Person Psychotherapy with Perinatal Patients amid the COVID-19 Pandemic: A Multistakeholder Perspective. *International journal of environmental research and public health* 2021; 18: 12234.
6. Dennis C-L, Hodnett E, Kenton L, et al. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. *Bmj* 2009; 338: a3064.
7. Dennis C-L, Ravitz P, Grigoriadis S, et al. The effect of telephone-based interpersonal psychotherapy for the treatment of postpartum depression: study protocol for a randomized controlled trial. *Trials* 2012; 13: 38.
8. Cox JL, Holden JM and Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British journal of psychiatry* 1987; 150: 782-786.
9. McCabe-Beane JE, Segre LS, Perkhounkova Y, et al. The identification of severity ranges for the Edinburgh Postnatal Depression Scale. *Journal of Reproductive and Infant Psychology* 2016; 34: 293-303.
10. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 2006; 166: 1092-1097.
11. Christensen H, Batterham PJ, Grant JB, et al. A population study comparing screening performance of prototypes for depression and anxiety with standard scales. *BMC Med Res Methodol* 2011; 11: 154. 2011/11/23. DOI: 10.1186/1471-2288-11-154.
12. Spitzer RL, Kroenke K, Williams JB, et al. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Jama* 1999; 282: 1737-1744.
13. Zimet GD, Dahlem NW, Zimet SG, et al. The multidimensional scale of perceived social support. *Journal of personality assessment* 1988; 52: 30-41.
14. Üstün TB, Chatterji S, Kostanjsek N, et al. Developing the World Health Organization disability assessment schedule 2.0. *Bulletin of the World Health Organization* 2010; 88: 815-823.
15. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research* 2011; 20: 1727-1736.
16. Lang AJ and Stein MB. An abbreviated PTSD checklist for use as a screening instrument in primary care. *Behaviour research and therapy* 2005; 43: 585-594.
17. Lang AJ, Wilkins K, Roy-Byrne PP, et al. Abbreviated PTSD Checklist (PCL) as a guide to clinical response. *Gen Hosp Psychiatry* 2012; 34: 332-338. 2012/03/31. DOI: 10.1016/j.genhosppsych.2012.02.003.
18. Han B, Wong EC, Mao Z, et al. Validation of a brief PTSD screener for underserved patients in federally qualified health centers. *Gen Hosp Psychiatry* 2016; 38: 84-88. 2015/09/21. DOI: 10.1016/j.genhosppsych.2015.07.009.

19.Weobong B, Weiss HA, McDaid D, et al. Sustained effectiveness and cost-effectiveness of the Healthy Activity Program, a brief psychological treatment for depression delivered by lay counsellors in primary care: twelve-month follow-up of a randomised controlled trial. *PLoS Medicine* 2017; 17.
20.Singla DR, MacKinnon DP, Fuhr DC, et al. Multiple mediation analysis of the peer-delivered Thinking Healthy Programme for perinatal depression: findings from two parallel, randomised controlled trials. *The British Journal of Psychiatry* 2019: 1-8.
21.Attkisson CC and Zwick R. The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and program planning* 1982; 5: 233-237.
22.Munder T, Wilmers F, Leonhart R, et al. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin Psychol Psychother* 2010; 17: 231-239. 2009/12/17. DOI: 10.1002/cpp.658.
23.Brown G. Health service utilization and cost of care questionnaire. *Health and Social Service Utilization Research Unit Hamilton, ON: McMaster University* 2001.
24.Fuhr DC, Calvert C, Ronsmans C, et al. Contribution of suicide and injuries to pregnancy-related mortality in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Psychiatry* 2014; 1: 213-225. 2015/09/12. DOI: 10.1016/s2215-0366(14)70282-2.
25.Singla DR, Weobong B, Nadkarni A, et al. Improving the scalability of psychological treatments in developing countries: an evaluation of peer-led therapy quality assessment in Goa, India. *Behaviour research and therapy* 2014; 60: 53-59.
26.Bradley RH and Caldwell BM. The HOME Inventory and family demographics. *Developmental Psychology* 1984; 20: 315.
27.Bayley N and Aylward G. *Bayley scales of infant and toddler development*. 4 edn ed.: MN: NCS Pearson, Inc, 2019.
28.GUIDANCE D. Multiple endpoints in clinical trials guidance for industry. *Center for Biologics Evaluation and Research (CBER)* 2017.
29.Use EoMfH. Points to consider on multiplicity issues in clinical trials. In: Products TEAftEoM, (ed.). 2002.
30.Dennis C-L, Hodnett E, Kenton L, et al. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. *British Medical Journal* 2009; 338: a3064. DOI: 10.1136/bmj.a3064.
31.Hintze J. PASS Version 12. Kaysville, Utah, USA: NCSS, LLC, 2014.
32.Molenaar NM, Brouwer ME, Bockting CLH, et al. Stop or go? Preventive cognitive therapy with guided tapering of antidepressants during pregnancy: study protocol of a pragmatic multicentre non-inferiority randomized controlled trial. *BMC Psychiatry* 2016; 16. DOI: 10.1186/s12888-016-0752-6.
33.Richards DA, Ekers D, Mcmillan D, et al. Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial. *The Lancet* 2016; 388: 871-880. DOI: 10.1016/s0140-6736(16)31140-0.
34.Yawn BP, Pace W, Wollan PC, et al. Concordance of Edinburgh Postnatal Depression Scale (EPDS) and Patient Health Questionnaire (PHQ-9) to Assess Increased Risk of Depression among Postpartum Women. *The Journal of the American Board of Family Medicine* 2009; 22: 483-491. DOI: 10.3122/jabfm.2009.05.080155.
35.SAS/ACCESS® 9.4 Interface to ADABAS. Cary, NC, USA: SAS Institute Inc 2013.
36.Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer, 2001.

37.O'Mahen HA RD, Woodford J, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. *Psychological Medicine* 2014; 44: 1675-1689.
38.Consortium. PDATCaTP. Heterogeneity of postpartum depression: a latent class analysis. *The Lancet Psychiatry* 2015; 2: 59-67.
39.Rubin DB. Multiple imputation after 18+ years. *Journal of the American statistical Association* 1996; 91: 473-489.
40.Royston P and White IR. Multiple imputation by chained equations (MICE): Implementation in Stata. *J Stat Softw* 2011; 45. DOI: 10.18637/jss.v045.i04.
41.Bodner TE. What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal* 2008; 15: 651-675.
42.Cohen ZD and DeRubeis RJ. Treatment selection in depression. *Annual review of clinical psychology* 2018.
43.Cohen ZD, DeRubeis RJ, Hayes R, et al. The Development and Internal Evaluation of a Predictive Model to Identify for Whom Mindfulness-Based Cognitive Therapy Offers Superior Relapse Prevention for Recurrent Depression Versus Maintenance Antidepressant Medication. *Clinical Psychological Science* 2023; 11: 59-76. DOI: 10.1177/21677026221076832.
44.Stirman SW, Cohen ZD, Lunney CA, et al. A personalized index to inform selection of a trauma-focused or non-trauma-focused treatment for PTSD. *Behaviour Research and Therapy* 2021; 142: 103872. DOI: https://doi.org/10.1016/j.brat.2021.103872.
45.Schwartz B, Cohen ZD, Rubel JA, et al. Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research* 2021; 31: 33-51.
46.Delgadillo J, Ali S, Fleck K, et al. Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA psychiatry* 2022; 79: 101-108.
47.Polley E, LeDell E and van der Laan M. SuperLearner: Super Learner Prediction [computer program]. R package version 2.0-21: The Comprehensive R Archive Network. 2016.
48.Webb CA, Cohen ZD, Beard C, et al. Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *Journal of Consulting and Clinical Psychology* 2020; 88: 25.
49.Steyerberg EW and Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. *Journal of clinical epidemiology* 2016; 69: 245-247.
50.Kapelner A, Bleich J, Levine A, et al. Evaluating the effectiveness of personalized medicine with software. *Frontiers in big Data* 2021; 4: 572532.
51.Efthimiou O, Hoogland J, Debray TP, et al. Measuring the performance of prediction models to personalize treatment choice. *Statistics in medicine* 2023; 42: 1188-1206.
52.Magill M, Kiluk BD, McCrady BS, et al. Active ingredients of treatment and client mechanisms of change in behavioral treatments for alcohol use disorders: Progress 10 years later. *Alcoholism: Clinical and Experimental Research* 2015; 39: 1852-1862.
53.Weobong B, Weiss H, McDaid D, et al. Sustained effectiveness and cost-effectiveness of the Healthy Activity Program, a brief psychological treatment for depression delivered by lay counsellors in primary care: a randomised controlled trial. . *PLoS Medicine* 2017.
54.Gaynor ST and Harris A. Single-participant assessment of treatment mediators: strategy description and examples from a behavioral activation intervention for depressed adolescents. *Behavior modification* 2008; 32: 372-402. DOI: 10.1177/0145445507309028.
55.Singla DR, Kumbakumba E and Aboud FE. Effects of a parenting intervention to address maternal psychological wellbeing and child development and growth in rural Uganda: a

community-based, cluster randomised trial. *Lancet Glob Health* 2015; 3: e458-e469. DOI: 10.1016/S2214-109X(15)00099-6.
56.Toth SL, Rogosch FA, Oshri A, et al. The efficacy of interpersonal psychotherapy for depression among economically disadvantaged mothers. *Development and psychopathology* 2013; 25: 1065-1078. DOI: 10.1017/S0954579413000370.
57.Singla DR. Examining the 'why' of the Thinking Healthy Programme by Peers in Goa, India and Rawalpindi, Pakistan: mediation and underlying processes. In: *Marce International* Bangalore, India 2018.
58.Walker SP, Wachs TD, Grantham-McGregor S, et al. Inequality in early childhood: risk and protective factors for early child development. *Lancet* 2011; 378: 1325-1338. 2011/09/29. DOI: 10.1016/s0140-6736(11)60555-2.
59.Cox A, Puckering C, Pound A, et al. The impact of maternal depression in young children. *Journal of Child Psychology and Psychiatry* 1987; 28: 917-928.
60.Cooper PJ, Tomlinson M, Swartz L, et al. Improving quality of mother-infant relationship and infant attachment in socioeconomically deprived community in South Africa: randomised controlled trial. *Bmj* 2009; 338: b974.
61.Figueiredo B, Canario C and Field T. Breastfeeding is negatively affected by prenatal depression and reduces postpartum depression. *Psychol Med* 2014; 44: 927-936. 2013/07/05. DOI: 10.1017/s0033291713001530.
62.Dennis C-L and McQueen K. The relationship between infant-feeding outcomes and postpartum depression: a qualitative systematic review. *Pediatrics* 2009; 123: e736-e751.
63.Singla DR, Kumbakumba E and Aboud FE. Effects of a parenting intervention to address maternal psychological wellbeing and child development and growth in rural Uganda: a community-based, cluster-randomised trial. *The Lancet Global Health* 2015; 3: e458-e469.
64.Lemmens LH, Muller VN, Arntz A, et al. Mechanisms of change in psychotherapy for depression: An empirical update and evaluation of research aimed at identifying psychological mediators. *Clin Psychol Rev* 2016; 50: 95-107. DOI: 10.1016/j.cpr.2016.09.004.
65.Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annual review of clinical psychology* 2007; 3: 1-27. DOI: 10.1146/annurev.clinpsy.3.022806.091432.
66.VanderWeele T and Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* 2009; 2: 457-468.
67.Levis B, Negeri Z, Sun Y, et al. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. *BMJ* 2020: m4022. DOI: 10.1136/bmj.m4022.
68.V; P, B; W, HA; W, et al. The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomized controlled trial. *Lancet* 217; 389: 176-185.
69.K; k, RL; S and JB W. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16: 606-613.
70.Lang AJ and Stein MB. An abbreviated PTSD checklist for use as a screening instrument in primary care. *Behav Res Ther* 2005; 43: 585-594. DOI: 10.1016/j.brat.2004.04.005.
71.Katajapuu N, Heinonen A and Saltychev M. Minimal clinically important difference and minimal detectable change of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) amongst patients with chronic musculoskeletal pain. *Clinical Rehabilitation* 2020; 34: 1506-1511. DOI: 10.1177/0269215520942573.

Appendix 1. List of Variables

VARIABLE	MEASURE	TIME OF ASSESSMENT						
		SCREENING	BASELINE	SESSION-WISE	3 MONTHS	6 MONTHS	12 MONTHS	HOME VISIT
PRIMARY OUTCOME								
Depressive symptoms	Edinburgh Postnatal Depression Scale (EPDS)	☐	☐	☐	☒	☐	☐	☐
SECONDARY OUTCOMES								
Depressive symptoms	Edinburgh Postnatal Depression Scale (EPDS)	☒	☐	☒	☐	☒	☒	☐
Anxiety Symptoms	General Anxiety Disorder-7 (GAD-7)	☒	☐	☒	☒	☒	☒	☐
Child Mental Development	Bayley-IV	☐	☐	☐	☐	☐	☐	☒
Psychosocial Stimulation	Home Observation Measurement Evaluation (HOME)	☐	☐	☐	☐	☐	☐	☒
EXPLORATORY OUTCOMES								
Response & Remission	Patient Health Questionnaire (PHQ-9)	☐	☒	☐	☒	☒	☒	☐
Suicide Risk*	Columbia Suicide Severity Risk Screener (C-SSRS)	☒	☐	☐	☐	☐	☐	☐
Patient-Reported Activation	Premium Abbreviated Activation Scale (PAAS)	☐	☒	☐	☒	☒	☒	☐
Quality of Life Assessment	EQ5D-5-level (EQ5D-5L)	☐	☒	☐	☒	☒	☒	☐
Perceived Support	Multidimensional Scale of Perceived Social Support (MSPSS)	☐	☒	☐	☒	☒	☒	☐
Trauma Symptoms	Post-Traumatic stress disorder checklist (PCL-6)	☐	☒	☐	☒	☒	☒	☐
Patient Satisfaction	Client Satisfaction Questionnaire (CSQ-8)	☐	☐	☐	☒	☐	☐	☐
Therapeutic Alliance	Working Alliance Inventory – Short Revise (WAI-SR)	☐	☐	☐	☒	☐	☐	☐
Disability Assessment	World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0)	☐	☒	☐	☒	☒	☒	☐
Dosage	Treatment Log: number of sessions attended	☐	☐	☒	☐	☐	☐	☐
Homework Adherence	Treatment Log: degree of activation	☐	☐	☒	☐	☐	☐	☐
COVID-19 exposure	COVID-19 Exposure (Y/N)	☐	☒	☒	☒	☒	☒	☐
Health Services Utilization	Health Service Utilization Questionnaire (HSU-Q)	☐	☒	☒	☒	☒	☒	☐
Health Benefits Access and Use	Health Benefits Questionnaire	☐	☒	☐	☒	☒	☒	☐
Medications	Currently taking medication (Y/N): • If yes, list:____	☐	☐	☒	☐	☐	☐	☐

Medication	Change in medication since last time the [medication] questionnaire was completed (Y/N)	☐	☐	☒	☐	☐	☐	☐
Screening: Alcohol use	Alcohol use disorder identification test (AUDIT)	☒	☐	☐	☐	☐	☐	☐
Screening: Treatment Preference	Treatment Preference (IP/TM)	☒	☐	☐	☐	☐	☐	☐
Maternal Characteristics: Age	Self-reported age (years)	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Immigrant status	Born in the country of current residence (Y/N)	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Immigrant status	Duration (years) lived in Canada/USA	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Ethnicity (**MC**)	Self-reported ethnicity	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Residence	Duration (years) of residence in current home	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Education Level (**MC**)	Highest level of education completed	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Marital status (**MC**)	Marital status	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Occupational status (**MC**)	Current employment/work status	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Income status (**MC**)	Household Income (before taxes)	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Income status (**MC**)	Ability to manage on current family income	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Mental health history	History of depression or anxiety (Y/N)	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Mental health history	History of depression related to birth of baby	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Mental health history	Age of first depression/anxiety	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Mental health history	History of seeing therapist	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Mental health history	Months of therapy since last session in the last year	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Psychosis symptoms	Psychosis screener	☒	☐	☐	☐	☐	☐	☐
Maternal Characteristics: Mania symptoms	Mania symptoms	☒	☐	☐	☐	☐	☐	☐

Maternal Characteristics: Drug use	Drug use screener	☒	☐	☐	☐	☐	☐	☐
Maternal Characteristics: Medical history (**MC**)	Medical conditions • High BP, PCOS, Diabetes, Kidney disease, Autoimmune disease, Thyroid disease, Obesity, HIV/Aids, None, Not wish to answer	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Medical history (**MC**)	Medication. If yes: • Over the counter. List: _____ • Prescription. List: _____ • Herbal. List: _____	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Medical history	Alcohol use (Y/N/Y but not while pregnant)	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Medical history	Alcohol use frequency	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Medical history	Cannabis use (Y/N/Y but not while pregnant)	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Medical history	Cannabis use frequency	☐	☒	☐	☐	☐	☐	☐
Maternal Characteristics: Sexual orientation and gender identity (**MC**)	Gender identity	☐	☒	☐	☐	☐	☒	☐
Maternal Characteristics: Sexual orientation and gender identity (**MC**)	Sexual orientation	☐	☒	☐	☐	☐	☒	☐
Maternal Characteristics: Sexual orientation and gender identity (**MC**)	Gender identity of partner	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Pregnancies	Pregnant (yes/no)	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Pregnancies	Weeks pregnant or post-partum	☒	☒	☐	☐	☐	☐	☐
Perinatal history: Pregnancies	Intentional pregnancy (yes/intentions kept changing/no)	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Pregnancies	History of pregnancies	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Miscarriages	History of miscarriages (< 20 weeks)	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Still birth	History of still births (>= 20 weeks)	☒	☒	☐	☐	☐	☐	☐
Perinatal history: Infant death	Infant death (if postpartum)	☒	☐	☐	☐	☐	☐	☐
Perinatal history: Preterm labor	History of preterm live births (< 37 weeks)	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Abortion	History of abortions	☐	☒	☐	☐	☐	☐	☐

Perinatal history: Medical history	History of ectopic pregnancy	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Medical history (**MC**)	Pregnancy conditions • Pre-eclampsia, High blood pressure, Gestational diabetes, Preterm labor	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Medical history (**MC**)	Medications during pregnancy • Insulin, Glyburide, Metformin, Betamethasone, Progesterone injections or suppositories to prevent preterm birth	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Pregnancies	Most recent delivery (singleton, twins, triplets, quadruplets)	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Child	Child date of birth	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Child	Child sex	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Child	Child due date	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Child	Child birth weight	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Child	Child birth length	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Child (**MC**)	Pregnancy outcome of indexed child	☐	☐	☐	☐	☐	☒	☐
Perinatal history: Child	Fetal anomaly	☒	☐	☐	☐	☐	☐	☐
Perinatal history: Births	Previous births (count)	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Number of children	Number of children	☐	☒	☐	☐	☐	☐	☐
Perinatal history: Labor (**MC**)	How labor began (on own, induced, Caesarean)	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Labor (**MC**)	Planned method of delivery (Vaginal, Caesarean section)	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Labor (**MC**)	Medication for pain during recent labor and delivery	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Labor	Pitocin (Oxytocin) at any point before your baby was born	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Labor	Bleeding after delivery	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Labor	Blood transfusion (Y/N)	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Labor	Baby admitted to NICU after delivery	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Breastfeeding	Plan for feeding infant	☐	☒	☐	☐	☐	☒	☐
Perinatal history: Breastfeeding	Pain, if any, breast feeding during the following time periods	☐	☒	☐	☐	☐	☒	☐
Notes: **MC**=Multiple Choice.								

TREATMENT PROVIDER VARIABLES			
VARIABLE	**MEASURE**	**TIME OF ASSESSMENT**	
		ENROLLMENT	**SUPERVISION SESSION-WISE**
Provider characteristics: Age	Self-reported age (years)	☒	☐
Provider characteristics: Sexual orientation and gender identity (**MC**)	Gender	☒	☐
Provider characteristics: Sexual orientation and gender identity (**MC**)	Gender Identity	☒	☐
Provider characteristics: Sexual orientation and gender identity (**MC**)	Sexual orientation	☒	☐
Provider characteristics: Ethnicity (**MC**)	Ethnicity	☒	☐
Provider characteristics:	Designation (NSP/SP)	☒	☐
Provider experience	Experience implementing evidence-based psychological treatments prior to SUMMIT (Y/N)	☒	☐
Provider experience	Experience seeing patients using telemedicine prior to SUMMIT (Y/N)	☒	☐
Weekly Supervision	Provider Attendance	☐	☒
Measurement-Based Supervision	Q-SUMMIT (Adapted from Quality of Healthy Activity Program: Q-HAP)	☐	☒
Therapy Quality	Q-SUMMIT (Adapted from Quality of Healthy Activity Program: Q-HAP)	☐	☒
Notes: **MC**=Multiple Choice.			

Other forms collected, when applicable:

- *Columbia Suicide Severity Rating Scale – Cox and Holden completed at any time point if a participant:
  - Has a positive verbal response to self-harm
  - Scores >0 on question 10 of the EPDS
  - Scores >0 on question 9 of the PHQ-9
- Infant Harm By A Child
- Infant Harm By Other Adult
- Infant Harm By Study Participant
- S/AEs

Appendix 2. Cut-off Values, Severity Classification and Psychometric Properties for Key Outcomes

Outcome	Instrument	Values	Cronbach's Alpha**
Depressive Symptoms	Edinburgh Postnatal Depression Scale (EPDS)	-Cut-off: EPDS total score ≥10[8, 67] -Severity Classification[9]*: • 0-9: None • 10-11: Mild • 12-19: Moderate • 20-30: Severe	0.86
Anxiety Symptoms	Generalized Anxiety Disorder Scale (GAD-7)	-Cut-off: GAD-7 total score ≥10[11] -Severity Classification[10]: • 0-4: None • 5-9: Mild • 10-14: Moderate • 15-21: Severe	0.90
Response & Remission	Patient Health Questionnaire-9 (PHQ-9)	-Cut-offs[68]: •Response: PHQ-9 total score <10 •Remission: PHQ-9 total score <5 -Severity Classification[69]: • 0-4: None • 5-9: Mild • 10-14: Moderate • 15+: Moderately severe & severe	0.87
Trauma Symptoms	Abbreviated PTSD Checklist (PCL-6)	-Cut-off: PCL-6 total score 14[16, 18, 70] -Severity Classification[18]: • 6-12: Low risk • 13-16: Medium risk • 17-25: High risk • 26-30: Very high risk	0.86
Disability Assessment	World Health Organization Disability Assessment Schedule (WHODAS)	-Cut-off: WHODAS total score 3.1[14, 71]	0.91
Child Mental Development	Bayley Child Mental Development Scales IV	-Mean score: 100	• Cognitive: 0.90 • Receptive Language: 0.93 • Expressive Language: 0.90
Perceived Support	Multidimensional Scale of Perceived Social Support (MSPSS)	Not Applicable	0.95
Patient-Reported Activation	Premium Abbreviated Activation Scale (PAAS)	Not Applicable	0.81
Therapy Quality	Q-SUMMIT	Not applicable	• Treatment-specific Skills: 0.86 • General Skills: 0.94

**In addition to previously-established cut-offs, an exploratory analysis using ROC curve will be conducted to compare symptom severity categories between the EPDS and PHQ-9.*

***Cronbach's alpha for these measures were conducted between August 2022 and March 2023 and based on a minimum sample size of n=279.*