

Supplement 1:

I. Trial Protocol

The 4th Trimester Statistical Analysis Plan version: v1.0

May 2025

Developed during trial conduct and finalized prior to primary outcome analysis

Study Outcomes

Validated instruments were used to measure dimensions of well-being, meaning in work, and distress. Likert-scored responses were scaled from 0 (not at all) to 4 (completely true/extremely so).

Primary Outcome: Burnout

The primary outcome was the change in burnout from enrollment to 24 weeks postpartum, measured by the Stanford Professional Fulfillment Index (SPFI). Burnout consists of two Likert-scored subscales:

- Emotional Exhaustion: 4 items
- Interpersonal Disengagement: 6 items

These were averaged and normalized to a 0–10 scale. Results are presented as both continuous and binary outcomes. Scores ≥ 3.325 on any burnout subscale or overall indicate high burnout. The SPFI was selected over the Maslach Burnout Inventory (MBI) due to its strong correlation ($r \sim 0.8$) and greater sensitivity to short-term changes, making it better suited for postpartum and intervention studies.

Secondary Outcomes

Secondary outcomes assessed change from enrollment to 24 weeks postpartum in the following well-being measures:

1. Professional Fulfillment:

SPFI subscale with three Likert-scored items, averaged and normalized to a 0–10 scale; scores ≥ 7.5 indicate high fulfillment.

2. Organization and Personal Values Alignment:

Three Likert-scored items summed to a 0–12 scale. This scale is inversely correlated with burnout and positively correlated with fulfillment.

3. Work Impact on Personal Relationships:

Four Likert-scored items, averaged and normalized to 0–10 scale. Higher scores indicate social isolation, associated with burnout and patient complaints.

4. PROMIS Sleep Impairment:

Scale of 8–40; scores ≥ 16 indicate fatigue. Sleep impairment correlates with burnout, reduced fulfillment, and increased risk of errors.

5. Edinburgh Postnatal Depression Scale (EPDS):

37 Scale 0–30; scores >13 denote a positive screen for depression. The EPDS demonstrates 81%
38 sensitivity and 88% specificity compared to gold-standard interviews.

39 6. Career Dissatisfaction:

40 Measured via agreement (strongly agree/agree) with statements on career regret,
41 thoughts of leaving training, and perceived conflict between career and family responsibilities,
42 compared to disagreement (strongly disagree/disagree/neutral).

43 **Study Inclusion and Exclusion Criteria**

44 Inclusion criteria:

- 45 1. Childbearing resident or fellow level MD or DO trainee from any specialty.
- 46 2. Pregnant (≥ 12 weeks gestation) with planned delivery during clinical rotations.
- 47 3. Currently in a clinical rotations/training program (not on research elective).

48
49 Exclusion criteria:

- 50 1. Trainees who are new mothers with a baby greater than 1 month of age.
- 51 2. Trainees who are not on clinical rotations/ clinical training program (on a research
52 electives).
- 53 3. Trainees who expect taking more than 16 weeks of maternity leave after delivery or
54 completing training within 6 months postpartum.
- 55 4. Trainees who have known medical or obstetric co-morbidities or (i.e. cardio-pulmonary,
56 gastrointestinal) which may interfere or preclude use of the smart bassinet and/or the
57 wearable breast pump.

58 59 **Local recruitment procedures:**

- 60 1. Identification: Trainees were recruited through a GME administered email flyer. Ongoing
61 identification of potential trainees suitable for study was made by the program director of
62 the training program or self. Subjects could either self-identify or could be identified by
63 training program directors by expressing interest via email to the study PI or study
64 personnel.
- 65 2. Responsible: The study PI or study personnel reviewed requests and recruit individuals
66 into the study.
- 67 3. Participants were recruited over the phone, email or via an in-person meeting to review
68 the requirements of the study, ensure they meet inclusion/exclusion criteria, and discuss
69 the associated risks/benefits/alternatives of participation in the trial. Consent can be
70 signed either in person or via secure electronic signature. After reviewing inclusion
71 criteria an additional in-person or virtual meeting was set up for consent signature,
72 providing study materials and random group allocation.

73

74
75
76

Measures

Variable	Source	Measure type, values
DEMOGRAPHICS		
Age	Self-reported	Numerical
Race	Self-reported	Categorical: Asian, Black or African American, White, More than one race, Other (Mediterranean, MENA, Hispanic/Latino alone)
Hispanic/Latino	Self-reported	
Additional children	Self-reported	Numerical
Pregnancy Trimester	Self-reported	Numerical, discrete: 1,2,3
Relationship status	Self-reported	Categorical: Married, Committed and/or long-term relationship, Single, Divorced/separated and single
PARTNER/HOUSEHOLD CHARACTERISTICS		
Partner work hours	Self-reported	Categorical: Stay at Home or <40, 40-59, 60+, No partner
Partner occupation	Self-reported	Categorical: Attending Physician or Surgeon, Other paid occupation, Resident/Fellow, Student/Stay at home parent
Physician partner	Self-reported	Categorical, binary: Yes, No
Household income	Self-reported	Categorical: <\$125,000, \$125,000 - \$249,999, >\$250,000

Expected childcare help	Derived from self-reported data	Categorical, binary: Yes to any= I have local friends/family (unpaid) to help with childcare, I have friends/family (unpaid) traveling in to help with childcare, I have friends/family to help with childcare who I am financially supporting, I am hiring additional help (e.g., daycare, nanny) , No
Childcare help (24 weeks)	Derived from self-reported data	Categorical, binary: Yes to any = Yes, I had local family/friends who were able to help out, Yes, I had family or friends who relocated to help out, I ended up hiring more help than I anticipated needing when I was pregnant , No
PROGRAM CHARACTERISTICS		
Procedural	Derived from self-reported data	Categorical, binary: Yes, No
Surgical	Derived from self-reported data	Categorical, binary: Yes, No
Hospital	Self-reported	Categorical: Site 1, Site 2, Site 3, Site 4, Site 5, Site 6, Site 7
Program size	Self-reported	Categorical: 1-3 chief residents or fellows, 4-5 chief residents or fellows, 6-10 chief residents or fellows, 11-15 chief residents or fellows, 15+ chief residents or fellows
Specialty (group)	Derived from self-reported data	Categorical:Anesthesiology (Adult and Child), Clinical + Interventional Specialties, Clinical Subspecialties, Critical Care, Diagnostic Specialties, Emergency Medicine (Adult and Child), General Surgery, Medicine, OBGYN and MFM, Pediatric Subspecialties, Pediatrics, Psychiatry (Adult and Child), Surgical Subspecialties

	Group	Specialties
	<i>Anesthesiology (Adult and Child)</i>	<i>Anesthesia, Peds Anesthesia</i>
	<i>Clinical + Interventional Specialties</i>	<i>Interventional Cardiology, Cardiovascular/Cardio imaging, Radiation Oncology</i>
	<i>Clinical Subspecialties</i>	<i>Cardiology, Hematology Oncology, Infectious Diseases, Neurology, Rheumatology, Allergy, Dermatology, Medicine-Dermatology, PMR, Endocrinology, GI, Nephrology</i>
	<i>Critical Care</i>	<i>Peds NICU, Peds Critical care, Pulmonary Critical Care</i>
	<i>Diagnostic Specialties</i>	<i>Radiology, Pathology</i>
	<i>Emergency Medicine (Adult and Child)</i>	<i>Emergency medicine, Pediatrics Emergency medicine</i>
	<i>General Surgery</i>	<i>General surgery</i>
	<i>Medicine</i>	<i>Internal Medicine</i>
	<i>OBGYN and MFM</i>	<i>Obstetrics and Gynecology, Complex Family Planning Fellowship, Obstetrics and Gynecology / Surgery, Maternal-Fetal Medicine</i>
	<i>Pediatric Subspecialties</i>	<i>Peds Pulmonology, Peds Neurology, Peds GI, Peds Nephrology, Peds Cardiology, Peds Pain Medicine, Neonatology</i>
	<i>Pediatrics</i>	<i>Pediatrics, Med-Peds, Pediatrics (Hospital Medicine)</i>
	<i>Surgical Subspecialties</i>	<i>Plastic Surgery, ENT, Ophthalmology, Peds Plastic Surgery (Craniofacial), Orthopedics</i>

	<i>Psychiatry (Adult and Child)</i>	<i>Psychiatry, Peds Psychiatry, Peds Triple Board, Triple Board</i>
Level of training	Derived from self-reported data	Categorical: Junior resident, Senior resident, Fellow
PGY	Self-reported	Categorical: 1, 2, 3, 4, 5, 6, 7, Fellow
Expected leave duration (weeks)	Self-reported	Numerical
Expected leave duration category (weeks)	Derived from self-reported data	Categorical: Short (< 8), Standard (8-12), Extended (>12)
Leave duration (weeks)	Self-reported	Numerical
Leave duration category (weeks)	Derived from self-reported data	Categorical: Short (< 8), Standard (8-12), Extended (>12)
Program support at enrollment	Self-reported	Categorical, binary: Yes = Leadership cares about my wellbeing as pregnant trainee, No
Program support during pregnancy and postpartum period (24 weeks)	Derived from self-reported data	Categorical, binary: Yes to any = Given lighter rotations (but still clinical work) during my third trimester, Given a reduced call schedule during my third trimester, Given rotations in which I had no clinical duties, and that still counted as time working that I did not have to make up, Given other duty adjustments to lighten my clinical workload that are not listed above, Put on a less demanding rotation, Had alterations to my call schedule No, checked both = No, I worked a normal schedule until delivery, No post-partum work adjustments

Written policies	Self-reported	Categorical: Duration of parental leave, Adjustment of clinical duties during pregnancy or after returning from parental leave, Clinical accommodations to allow time for breastfeeding/pumping, Childcare resources, A mentorship program supporting new parents
Written policies (binary)	Derived from self-reported data	Categorical, binary: Yes, No
Pregnancy complications	Self-reported	Categorical: Pre-term delivery (less than 37 weeks gestation), Gestational diabetes, High blood pressure (gestational hypertension or preeclampsia), Other
Readmission	Self-reported	Categorical: Baby required NICU stay, baby was readmitted, mother was readmitted
OUTCOMES		
Burnout score	Derived from self-reported data	Numerical (0-10)
High burnout	Derived from self-reported data	Categorical, binary: Yes (≥ 3.325), No (< 3.325)
Emotional Exhaustion subscale <i>(To what degree have you experienced the following during the past two weeks?)</i>		
“I have felt a sense of dread when I think about work I have to do”	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
“I have felt physically exhausted at work”	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
“I have felt lacking in enthusiasm at work”	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely

“I have felt emotionally exhausted at work”	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
Interpersonal Disengagement subscale (<i>“During the past two weeks, my job has contributed to me feeling:...”</i>)		
<i>“Less empathetic with my patients”</i>	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
<i>“Less empathetic with my colleagues”</i>	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
<i>“Less sensitive to others' feelings/emotions”</i>	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
<i>“Less interested in talking with my patients”</i>	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
<i>“Less connected with my patients”</i>	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
<i>“Less connected with my colleagues”</i>	Self-reported	Categorical: 0, Not at all 1, Very little 2, Moderately 3, A lot 4, Extremely
Professional fulfillment (<i>How true do you feel the following statements are about you at your work during the past two weeks?</i>)		

<i>"I feel happy at work"</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>"I feel worthwhile at work"</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>"My work is satisfying to me"</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>"I feel in control in dealing with difficult problems at work"</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>"My work is meaningful to me"</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>"I am contributing professionally (e.g., patient care, teaching, research, leadership) in the ways I value most"</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
Organization and Personal Values Alignment <i>(How true are the following statements about conditions in your principal training site?)</i>		
My input is valued in important administrative/training program decisions	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true

Our organizational goals and values fit well with my goals and values	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
Administration/the training program values my clinical work	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
Impact of work on personal relationships <i>(How has your job affected your personal relationships during the past month?)</i>		
<i>“Made it harder for me to enjoy my pregnancy”</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>“Made it harder for me to nurture existing personal relationships”</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>“Made it harder for me to develop new meaningful personal relationships”</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>“Contributed to conflict in my personal relationship(s)”</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
<i>“Contributed to me feeling more isolated or detached from the people who are most important to me”</i>	Self-reported	Categorical: 0, Not at all true 1, Somewhat true 2, Moderately true 3, Very true 4, Completely true
Career dissatisfaction or regret		
<i>“I have considered leaving my program in the last year (due to my experience of pregnancy/parenthood in my current training specialty)”</i>	Self-reported	Categorical: 1, Strongly agree 2, Agree 3, Neutral 4, Disagree 5, Strongly disagree

<i>"If I could choose again, I would choose a non-medical career more accommodating of parenthood"</i>	Self-reported	Categorical: 1, Strongly agree 2, Agree 3, Neutral 4, Disagree 5, Strongly disagree
<i>"If I could choose again, I would stay in medicine but choose a specialty more accommodating of parenthood than my current specialty"</i>	Self-reported	Categorical: 1, Strongly agree 2, Agree 3, Neutral 4, Disagree 5, Strongly disagree
<i>"I am changing my subspecialty/fellowship plans due to difficulty balancing parenthood with the demands of my original choice of subspecialty/fellowship"</i>	Self-reported	Categorical: 1, Strongly agree 2, Agree 3, Neutral 4, Disagree 5, Strongly disagree

78 **Summary of Protocol Amendments***
79 Original Protocol: IRB v11, approved December 12, 2022
80 First participant enrollment: May 25, 2023
81

Date / Timing Relative to Enrollment	Protocol Version	Description of Change	Rationale
Pre-enrollment	IRB clarification	Exclusion criteria amended to exclude trainees who had already given birth	To ensure baseline well-being measures reflected experiences while actively working during pregnancy (versus obtaining baseline measures during parental leave)
Pre-enrollment	IRB clarification	Randomization approach changed from use of online research randomization software (https://www.randomizer.org/) to permuted block randomization stratified by procedural vs non-procedural specialty and study site, implemented using SAS-generated randomization lists	To ensure balance across sites and specialty strata given anticipated small numbers within strata
Pre-enrollment	IRB clarification	Procedural specialty definition updated to include critical care	Alignment with procedural workload and training characteristics
Within two weeks of first enrollment	IRB amendment	Baseline survey timing changed from “within two weeks of enrollment” to “at the time of enrollment, prior to randomization”	To improve compliance and allow participants to complete surveys privately before allocation

Date / Timing Relative to Enrollment	Protocol Version	Description of Change	Rationale
Within two weeks of first enrollment	IRB amendment	Control-group compensation schedule adjusted (increased 16-week survey compensation from \$50 to \$75; reduced 24-week survey compensation from \$75 to \$50)	To improve survey completion; changes occurred before any compensation was issued
Within two weeks of first enrollment	IRB amendment	Inclusion criteria amended to require participants to be ≥ 12 weeks pregnant at enrollment	To minimize participant distress related to early pregnancy loss, which is more common during the first trimester.
Within two weeks of first enrollment	IRB amendment	Exclusion criteria amended to exclude trainees completing training before 6 months postpartum	To ensure completion of prespecified 24-week postpartum outcome assessments during active clinical training.
Within two weeks of first enrollment	IRB amendment	Exclusion criteria amended to exclude trainees planning ≥ 16 weeks of parental leave (which exceeds most boards' limits)	To ensure a comparable workload and define the high-risk cohort, trainees needed to be back at work when completing postpartum well-being metrics.
Within two weeks of first enrollment	IRB amendment	Recruitment procedures amended to allow self-referral prior to disclosure of pregnancy to program leadership	To preserve confidentiality and reduce barriers to participation
Date / Timing Relative to Enrollment	Protocol Version	Description of Change	Rationale

6/15/2023 (After enrollment began)	IRB amendment	Added Boston Children's Hospital as a study site	Planned site included after completion of local IRB administrative requirements
8/3/2023	IRB amendment	Added Beth Israel Deaconess Medical Center as a study site	Planned site included after completion of local IRB administrative requirements
9/13/2023	IRB amendment	Added Tufts Medical Center as a study site	To increase diversity in training program size and structure, thereby improving generalizability of findings across graduate medical education settings.
10/10/2023	IRB amendment	Added Hartford HealthCare–affiliated programs as study sites	To increase diversity in training program size and structure, improve generalizability, and include an independent (non-academic medical center) training program.
11/9/2023	IRB amendment	Added Brown University–affiliated programs as study sites	To increase diversity in training program size and structure, thereby improving generalizability of findings across graduate medical education settings.
Date / Timing Relative to Enrollment	Protocol Version	Description of Change	Rationale

After enrollment began (mid-study)	IRB amendment	Intervention procedures amended to allow additional smart bassinet leases on a case-by-case basis (e.g., multiple births or need for use at a secondary residence), with identical accessories and return requirements	Applied to two participants only: one participant with unexpected twins requiring two bassinets, and one participant with a secondary residence; accommodation did not alter intervention content or confer differential intervention intensity
------------------------------------	---------------	--	---

82 *No changes were made to the study's primary or secondary outcomes, measurement
83 instruments, or assessment timepoints after enrollment began. All protocol amendments
84 were implemented to improve feasibility, participant safety, confidentiality, or operational
85 execution.

86 **II. Statistical Analysis Plan**
87 *May 2025*
88 *Developed during trial conduct and finalized prior to primary outcome analysis*

89 **Study Sample**

90 We anticipated needing 78 subjects in each arm, 156 participants total, to detect a statistically
91 significant change in burnout score (by at least 1.025 using the SPFI), with 80% power with a two
92 tailed alpha of 10%, and allowing for 20% attrition of study subjects.

93 **Data Structure and Unit of Analysis**

94 Although some of our covariates included program-level characteristics (e.g., program type and
95 size), all outcomes were measured at the resident/fellow level. Thus, the unit of analysis for all
96 analyses was the resident/fellow.

97 **Data Transformations**

98 For all survey items with 5-point Likert scales, responses were dichotomized as strongly
99 disagree/disagree/neutral vs. agree/strongly agree. For validated numerical scales with a cutoff,
100 responses were dichotomized as above or below that value. Incomplete responses were not
101 permitted; missing data occurred only if a respondent had not returned to work by week 16.

102 **Unadjusted Analyses**

103 Baseline characteristics were compared between the intervention and control groups using
104 Pearson's chi-squared tests for categorical variables and independent samples t-tests for
105 continuous variables. These characteristics included resident and fellow demographics such as
106 age, race, ethnicity, number of children, trimester of pregnancy at enrollment, pregnancy
107 complications, and relationship status, as well as partner and household variables like partner work
108 hours, partner occupation, and household income along with family or friends expected childcare
109 help. Program-related factors including specialty and site, program size, level of training, expected
110 and actual leave duration, perceptions of program support both at enrollment and throughout
111 pregnancy, and the existence of written parental policies were also evaluated. For outcome
112 analyses, the Wilcoxon rank-sum test was used for two-group comparisons of continuous or ordinal
113 data, while the Friedman test assessed repeated measures of these outcomes, including burnout
114 scores, emotional exhaustion, interpersonal disengagement, professional fulfillment, alignment of
115 personal and organizational values, the impact of work on personal relationships, and sleep-related
116 impairment. Pearson's chi-squared test evaluated associations between categorical variables, and
117 Cochran's Q test was used for repeated measures of binary outcomes such as positive burnout,
118 high professional fulfillment, career regret, and high fatigue. Outcome measures were collected at
119 baseline, 16 weeks, and 24 weeks, with analyses including both between-group and within-group
120 comparisons to assess changes in unadjusted means at the end of the study.

121
122 **Multivariable Analyses**

123 We selected *a priori* program and resident-level demographic variables to include as fixed effects in
124 the mixed-effects models based on prior evidence of association with burnout from our past work
125 and literature and data collected at enrollment. Random effects were specified with individual ID
126 nested within hospital, using the structure (1 | hospital/record_id) to account for clustering at both
127 levels.

128 Robust linear mixed-effects models were used to analyze continuous outcomes, including burnout
129 score, professional fulfillment related to organization and personal values alignment, impact of
130 work on personal relationships, and sleep-related impairment, with adjustment for covariates and
131 random intercepts to account for clustering by hospital and individual. For binary outcomes, severe
132 burnout was defined as a burnout score ≥ 3.325 , professional fulfillment as SPFI > 7.5 , regret as
133 any affirmative response to regret questions, and fatigue as PROMIS score ≥ 16 . Mixed-effects
134 logistic regression models were fitted for these binary outcomes using the same covariate
135 adjustments and random effects. Odds ratios were derived by exponentiating the logistic model
136 coefficients, with confidence intervals calculated by the Wald method.

137 **Procedure for Handling Missing Data**

138 All analyses relied on a pairwise deletion approach for handling missing data, meaning that for
139 each robust mixed effects model, only observations with non-missing values for all variables in the
140 model formula were included in the analysis. The use of `na.action = na.exclude` in R mixed models
141 ensured that excluded rows were retained as missing in fitted values and residuals, preserving the
142 original data structure for further analysis and reporting. No imputation or additional modeling of
143 missing values was performed; cases with missing data in modeled variables were simply omitted
144 from model estimation. For our study, analyses focused on participants who had 16 weeks or less
145 of leave to avoid mixing data in which burnout scores were missing but sleep-related impairment
146 scores were available. Such responses would reflect experiences at home rather than at work,
147 pose no patient care risk, and would not represent typical scenarios for most specialties with
148 restrictions on time spent out of training. This approach was appropriate given that the burnout
149 evaluation was work-related and questions were directed at participants' attitudes towards patient
150 care. Additionally, spending at least one third of the six-month postpartum period on clinical
151 rotations was required as part of the inclusion criteria.

152 **Statistical Significance**

153 Statistical significance for all analyses was set at a two-tailed $p < 0.05$.

154 **Statistical Software**

155 R statistical software

156 **Summary of Statistical Analysis Plan Amendments***
 157 Original Statistical Analysis Plan: IRB v11, approved December 12, 2022. First
 158 participant enrollment: May 25, 2023

Date / Timing Relative to Enrollment	SAP Version	Description of Change	Rationale
12/12/2022 (Pre- enrollment)	Original SAP (IRB v11)	Original statistical analysis plan specified descriptive statistics and between-group comparisons using chi-squared tests for categorical variables and t-tests for continuous variables, with a priori power calculation based on SPFI burnout scores	Established primary and secondary outcomes, analytic intent, and target effect size prior to enrollment
7/2024	Interim analytic refinement	Certain covariates were reclassified into categorical or binary variables (e.g., parental leave duration categorized as short/standard/extended; program support, childcare help, and presence of written parental support policies modeled as binary variables)	To improve interpretability, reduce model complexity, and preserve statistical power given sample size, without altering outcomes or hypotheses
7/2024	Interim analytic refinement	Calculation and reporting of standardized effect sizes (Cohen's d) derived from model-based estimates	To facilitate interpretation of effect magnitude for clinically oriented readers; effect size reporting did not influence hypothesis testing or statistical inference
7/17/2024	Interim analytic refinement	Longitudinal mixed-effects modeling implemented to evaluate intervention effects while accounting for repeated measures and individual-level variability	Recommended during biostatistical consultation to appropriately model longitudinal data and separate intervention effects from individual variability

Date / Timing Relative to Enrollment	SAP Version	Description of Change	Rationale
9/10/2024	Final SAP refinement	Mixed-effects models were specified as the primary analytic approach. Alternative approaches (e.g., GEE or difference-in-differences) were considered but not implemented.	Analytic approach was prespecified based on consideration of data structure and model assumptions; outcomes, instruments, and timepoints unchanged
5/2025	Final SAP (May 2025)	Formalization of the analytic population (modified intention-to-treat) and approach to handling missing data (complete-case analysis)	To explicitly document analytic handling, consistent with CONSORT guidance and prespecified eligibility criteria
5/2025	Final SAP (May 2025)	Formalized full statistical analysis plan to fully specify modeling assumptions, covariates, diagnostics, and reporting conventions consistent with journal standards	To document and consolidate the final analytic approach prior to primary outcome analysis, without modification of prespecified outcomes, hypotheses, or assessment timepoints

*No changes were made to the study's prespecified primary or secondary outcomes, measurement instruments, or assessment timepoints after enrollment began. All statistical analysis plan amendments reflect analytic refinement rather than changes in study objectives or hypotheses and were finalized prior to primary outcome analysis.