CGM for the Early Detection and Management of Hyperglycemia in Pregnancy (IMAGINE)

RCT
Statistical Analysis Plan

Note: Table shells are in a separate document.

Version 1.0
Protocol Version 1.2
Author: Zoey Li
Date: April 1, 2025

Revision History

The following table outlines changes made to the Statistical Analysis Plan.

Version Number	Author	Senior Statistician	Effective Date	Study Stage
1.0	Zoey Li	Peter Calhoun	April 1, 2025	Planning.

Version Number	Revision Description
1.0	Original version

Approvals

Role	Digital Signature or Handwritten Signature/Date
Author: Zoey Li	
Senior Statistician: Peter Calhoun	
Jaeb Principal Investigator: Judy Sibayan	

Contents

1. Consistency of SAP with Protocol

This SAP was written with reference to protocol version 1.2. If the protocol is subsequently updated, then this SAP will be reviewed to ensure consistency with the new protocol. The SAP will not be revised unless the protocol changes require modification of the analyses. This SAP is consistent with the protocol stats chapter and details the analyses specifically for the RCT portion of the study.

2. Overview

The study will assess whether early detection (i.e. prior to gestational week 17) and treatment of elevated glycemia risk using continuous glucose monitoring (CGM) will reduce the prevalence of adverse pregnancy outcomes (APOs) when compared to standard care.

Participants will be screened for dysglycemia (defined as 5 to <25% of CGM values >140 mg/dL) prior to gestational week 17 through the data collected from a CGM sensor. The screening process will be done in two stages. Eligibility will first be assessed after 5 days of CGM wear. If the participant does not meet the eligibility requirement after 5 days, then the CGM sensor will be worn for an additional 5 days and all CGM data collected during the 10-day period will be reassessed for eligibility.

Approximately 6,000 participants will be screened, with an expected ~1,500 participants who will meet the criteria for dysglycemia. These 1,500 participants will then be randomized prior to gestational week 17 to either the gestational diabetes mellitus (GDM) Prevention Group (Glucose Lowering) or the Usual Care Group in a 1:1 ratio, stratifying by site. Participants in the Glucose Lowering group will undergo GDM treatment enhanced by the usage of an unblinded Dexcom G7 or Abbott's FreeStyle Libre 3 CGM sensor, while participants in the Usual Care group will undergo standard gestational care while wearing a blinded CGM sensor for 10-14 days approximately every 4 weeks throughout pregnancy. Participants in the Usual Care group will conduct a standard oral glucose tolerance test (OGTT) at 24-28 weeks' gestation; participants in the Glucose Lowering group may complete an OGTT if they are not receiving glucose-lowering medications. Outcome data will be collected from electronic medical records (EMR) and case report forms (CRFs) from sites during and after completion of the pregnancy.

The objectives of the RCT are as follows:

- To determine whether the identification of dysglycemia using continuous glucose monitoring (CGM) early in the second trimester of pregnancy with initiation of dietary and glucose management at this early timepoint (Glucose Lowering group) results in fewer APOs in the mother and infant compared with usual care (Usual Care group).

- To compare CGM-measured glycemic metrics during the second and third trimesters of pregnancy in the Glucose Lowering group versus the Usual Care group.
- To identify factors that modify the effect of Glucose Lowering group compared with the Usual Care group on a reduction in the risk of APOs.

3. Statistical Hypotheses

The primary endpoint is a test for superiority in the incidence of the primary composite endpoint (described in Section 5.1) comparing the two treatment groups (the Glucose Lowering group and the Usual Care group).

Let π_{GL} and π_{UC} represent the proportion of pregnancies with the composite endpoint for the Glucose Lowering and Usual Care groups, respectively. The statistical hypotheses for the primary endpoint are as follows:

Null Hypothesis (H_0): $\pi_{GL} - \pi_{UC} = 0$ (there is no treatment group difference in the incidence of the composite endpoint).
Alternative Hypothesis (H_a): $\pi_{GL} - \pi_{UC} \neq 0$ (there is a nonzero treatment group difference in the incidence of the composite endpoint).

4. Sample Size

Sample size has been computed for the primary composite endpoint. We expect ~25% of screened participants to meet the eligibility requirement of 5% to <25% time >140 mg/dL prior to 17 weeks' gestation. To achieve 90% power in detecting a difference in composite endpoint prevalence between the Glucose Lowering and Usual Care groups, assuming an estimated 41% of participants with the composite endpoint in the Usual Care group, an expected relative risk reduction of the composite endpoint of 25% in the Glucose Lowering group, a two-sided test, and a type I error rate of 5%, the calculated sample size is 900 participants randomized. The sample size has been increased to 1,500 randomized (~6,000 screened) to account for dropouts and to increase precision of the endpoint analyses. Statistical power for testing the primary composite endpoint is 98%, assuming an estimated 41% of participants with the composite endpoint in the Usual Care group, an expected relative risk reduction of the composite endpoint of 25% in the Glucose Lowering group, a two-sided test, a type I error rate of 5%, and 1,425 participants completing the study. A sample size of 1,425 completed participants also provides 88% power to detect a 20% reduction in the primary composite endpoint assuming a primary composite endpoint prevalence of 41% in the Usual Care group. A separate document details the justification of these power calculations.

5. Outcome Measures

5.1 Primary Efficacy Endpoint

The primary endpoint is:

- A composite that includes the following:
    - Large for gestational age (LGA, using INTERGROWTH definition)
    - Shoulder dystocia (including humeral or clavicular fracture)
    - Elevated bilirubin requiring phototherapy
    - Neonatal intensive care unit (NICU) admission
    - Maternal hypertensive disorder of pregnancy (HDP)

5.2 Secondary Efficacy Endpoints

- Composite consisting of neonatal complications of LGA (using INTERGROWTH definition), shoulder dystocia (including humoral or clavicular fracture), elevated bilirubin requiring phototherapy, NICU admission
- Maternal HDP
- LGA (using INTERGROWTH definition)
- Shoulder dystocia
- Elevated bilirubin requiring phototherapy
- Neonatal death
- NICU admission

5.3 Safety Endpoints

- Small for gestational age (SGA, $<10^{th}$ percentile using INTERGROWTH definition)
- CGM-measured time <54 mg/dL (during comparative periods where Usual Care group has blinded CGM)
- Severe hypoglycemia events
- Other serious adverse events

CGM-measured time <54 mg/dL will be calculated as described in section 12.

5.4 Exploratory Endpoints

Exploratory Maternal Endpoints

- Weight gain from randomization to end of pregnancy
- C-Section
- Pre-term birth

Exploratory Fetal/Neonatal Endpoints

- Large for gestational age (using GROW definition)
- Infant chromosomal abnormality, congenital anomaly, or malformation
- Evidence of other birth injury
- Length of neonatal hospital stay

6. Analysis Datasets

All efficacy and safety analyses comparing the treatment groups will follow the intention-to-treat approach, which means participants will be analyzed in the treatment arm assigned by randomization regardless of actual treatment received. All randomized participants will be included in the efficacy and safety analyses unless otherwise specified.

7. Per Protocol Analyses

A per-protocol analyses will be performed for the primary endpoint only if >5% of participants will be excluded based on the following requirements for inclusion in the per-protocol analysis:

- Glucose Lowering group: Wore CGM for ≥70% of their post-randomization gestational period.

8. Analysis of Outcomes

8.1 Analysis of the Primary Efficacy Endpoint

The prevalence of the primary composite endpoint in each treatment group will be tabulated. The primary composite endpoint at delivery will be compared between the Glucose Lowering and Usual Care groups using a logistic regression model adjusting for baseline % time >140 mg/dL with a compound symmetry covariance structure to handle the correlation within site. If the primary composite endpoint at delivery is missing, multiple imputation with treatment group and baseline % time >140 mg/dL in the model will be used to handle missing values.

Imbalances between groups in important covariates are not expected to be of sufficient magnitude to produce confounding. However, the presence of confounding will be evaluated in the sensitivity analyses by including factors potentially associated with the outcome for which there is an imbalance between groups (section 8.1.1).

A 95% confidence interval (CI) will be constructed on the treatment group risk difference (Glucose Lowering minus Usual Care).

8.1.1 Sensitivity Analyses

The following sensitivity analyses will be conducted for the primary outcome.

Confounding

A sensitivity analysis will be conducted if potential confounding factors collected at baseline are imbalanced between treatment groups. The imbalance will be assessed based on clinical judgment reviewing the distributions in the two treatment arms, not on a p-value. The person making this judgement will be unaware of whether there is an association between baseline variables and study outcome. All variables obtained on a continuous scale will be entered into the models as continuous variables, unless it is determined that a variable does not have a linear relationship with the log odds of the outcome. In such a case, categorization and/or transformation will be explored.

Missing Data

Missing data will be handled using multiple imputation using treatment group and baseline % time >140 mg/dL in the imputation model assuming the outcome is missing at random. It is worth noting that all statistical methods for handling missing data rely on untestable assumptions and there is no one correct way to handle missing data. The goal is to minimize the amount of missing data so that the results will not be sensitive to which statistical method is used.

To that end, sensitivity analyses will be performed to explore whether results are similar for the primary analysis when using different methods. Sensitivity analyses will only be performed if >5% of the primary composite outcome is missing. The following methods will be applied:

- Multiple imputation adjusting for baseline % time >140 mg/dL with a pattern mixture model assuming the dropout trajectory of the Glucose Lowering group is that of the Usual Care group
- Available cases only
- Two-way tipping point analysis

8.2 Analysis of Secondary Efficacy Endpoints

8.2.1 Binary Outcomes

The prevalence of each binary endpoint in each treatment group will be tabulated. Binary outcomes will be compared between the Glucose Lowering and Usual Care groups using models similar to that of the primary composite endpoint (section 8.1). A 95% confidence interval (CI) will be constructed on the treatment group risk difference (Glucose Lowering minus Usual Care).

Endpoints are only tested if there are enough events (at least 5 events combined between the two treatment groups).

8.3 Analysis of Exploratory Endpoints

All exploratory endpoints listed in section 5.4 will be compared between treatment groups. Mean ± SD or summary statistics appropriate to the distribution will be tabulated for each treatment group at baseline (if applicable) and follow-up. Continuous endpoints will be compared using a linear mixed effects regression model adjusting for baseline % time >140 mg/dL and a random site effect. A 95% CI will be constructed on the treatment group mean difference (Glucose Lowering minus Usual Care). If the residual values from the regression model have a skewed distribution, then an appropriate alternative transformation or nonparametric analysis will be performed instead.

Binary endpoints will be compared using a logistic regression model adjusting for baseline % time >140 mg/dL with a compound symmetry covariance structure to handle the correlation within site. A 95% CI will be constructed on the treatment group risk difference (Glucose Lowering minus Usual Care). Endpoints are only tested if there are enough events (at least 5 events combined between the two treatment groups).

8.3.1 Association between CGM Metrics and Adverse Pregnancy Outcomes

Within each group, the association of hyperglycemia and other CGM metrics listed in section 12 with vs. without the following maternal and fetal/neonatal outcomes will be explored if at least 20 events occur:

- Primary composite endpoint
- Composite neonatal endpoint
- Maternal HDP
- LGA (using INTERGROWTH definition)
- SGA
- Shoulder dystocia
- Elevated bilirubin requiring phototherapy
- Neonatal death
- NICU admission

The difference in CGM metrics between those with vs. without end maternal/fetal endpoint will be assessed using a linear mixed effects regression model adjusting for the baseline value of the metric, type of sensor, and site (random effect). Residual values will be examined for an approximate normal distribution. If residuals are highly skewed, then an appropriate alternative

transformation or nonparametric analysis will be performed instead. Missing data will be handled using direct likelihood.

9. Safety Analyses

All randomized participants will be included in the safety analyses. All reportable adverse events will be tabulated by treatment group. Safety analyses for the RCT will include events occurring on or after randomization until hospital discharge following the delivery.

For the following outcomes, the number of participants with ≥1 event and number of events will be tabulated by treatment group:

- SGA (<10th percentile)
- Severe hypoglycemia events
- Other serious adverse events

The prevalence of SGA will be summarized and a formal statistical comparison between treatment groups will be performed using a model similar to that used for the primary composite endpoint (section 8.1) if enough events occur (at least 5 events combined between the two treatment groups). Treatment group differences in rates of severe hypoglycemic events and other serious adverse events will not be tested due to reporting differences between the two groups.

CGM-measured time <54 mg/dL will be calculated and compared between treatment groups using a linear mixed effects regression model adjusting for the baseline time <54 mg/dL, sensor type, and site (random effect). A separate variance will be modelled for each treatment group. Residual values will be examined for an approximate normal distribution. If residuals are highly skewed, then an appropriate alternative transformation or nonparametric analysis will be performed instead. Missing data will be handled using direct likelihood.

10. Planned Interim Analyses

Re-estimation of the sample size will be undertaken as specified in the Sample Size Re-estimation Plan. The analysis will determine if the eligibility criterion based on the blinded CGM screening and the planned RCT sample size meets the desired power. The sample size may be modified or eligibility criteria revised based on the analysis.

11. Subgroup Analyses

In exploratory analyses, the primary composite endpoint, LGA (using INTERGROWTH definition), and HDP will be compared in baseline subgroups. Treatment group by subgroup factor interaction terms will be added to the models described in sections 8.1 and 8.2.1.

- Formal statistical testing for interaction will be done only if the overall result is statistically significant. Results will be tabulated by subgroups regardless of statistical significance. For continuous variables, results will be displayed in subgroups based on the distribution of the data although the analysis will utilize the variable as continuous. Cut point selection for display purposes will be made masked to the outcome data.
- Subgroups will be analyzed according to the following baseline factors:
    - Baseline HbA1c
    - Baseline % time >140 mg/dL
    - Baseline % time >120 mg/dL
    - Age
    - Race/Ethnicity
    - BMI
    - Education level

12.Glucose Exposure Assessment

Glucose exposure will be assessed using CGM which will be unblinded in the Glucose Lowering group and blinded in the Usual Care group. CGM data collected during screening will be used to calculate the baseline values for each CGM metric. CGM data collected after randomization up until the day before the delivery date will be used to calculate the CGM metrics for the follow-up period. During the follow-up period, wear periods of CGM data in the Glucose Lowering group will be matched to the blinded CGM wear periods in the Usual Care group by either determining which gestational weeks have CGM data in a substantial percentage of both groups or matching CGM wear periods on a participant-level.

CGM Metrics

- Mean glucose
- Glucose standard deviation
- Glucose coefficient of variation
- % Time 63-120 mg/dL
- % Time 63-140 mg/dL
- % Time <63 mg/dL
- % Time >120 mg/dL
- % Time >140 mg/dL
- Glucose level at 5 AM

For the baseline and follow-up periods, at least 1,152 CGM readings will be required to calculate the CGM metrics in each period. Summary statistics (mean ± SD or median [interquartile range])

will be reported for the CGM metrics at baseline and follow-up for each treatment group. CGM metrics will also be summarized by daytime and nighttime, by trimester, and by 4-week time periods. Daytime is defined as 6:00am to 11:59pm and nighttime from 12:00mn to 5:59am.

CGM metric differences between treatment groups will be compared using a linear mixed effects regression model adjusting for the baseline value of the metric, type of sensor, and site (random effect). A separate variance will be modelled for each treatment group. A 95% CI will be constructed on the treatment group mean difference (Glucose Lowering minus Usual Care). Residual values will be examined for an approximate normal distribution. If residuals are highly skewed, then an appropriate alternative transformation or nonparametric analysis will be performed instead. Missing data will be handled using direct likelihood.

Within the Glucose Lowering group, additional tabulations of CGM metrics will be done by gestational week.

Glycemic metrics will be also be tabulated separately for participants in the Usual Care group who started CGM due to a GDM diagnosis, overall and by trimester. All CGM data collected in the Usual Care group will be included for this analysis.

13.Analysis Windows

The following table defines the analysis windows for post-randomization CGM data inclusion:

Period	**Start date-time†**	**End date-time***
Overall gestational period	Midnight of estimated pregnancy start date	Midnight of delivery date (or estimated delivery date, if no delivery date is recorded)
1st trimester	Midnight of estimated pregnancy start date	23:59:59 on estimated pregnancy start date† + 90 days
2nd trimester	Midnight of estimated pregnancy start date + 91 days	23:59:59 on estimated pregnancy start date† + 181 days
3rd trimester	Midnight of estimated pregnancy start date + 182 days	Midnight of delivery date (or estimated delivery date, if no delivery date is recorded)
1st 4-week period	Midnight of estimated pregnancy start date	Start date-time + 28 days
2nd to 10th 4-week periods	End date-time of previous 4-week period	Start date-time + 28 days

1st gestational week	Midnight of estimated pregnancy start date	Start date-time + 7 days
2nd to 40th gestational week	End date-time of previous gestational week	Start date-time + 7 days

†The estimated pregnancy start date is defined as the estimated delivery date minus 279 days. All start date-times are truncated so that they start midnight of the randomization day + 1 day at the earliest.
*If pregnancy ends early due to miscarriage or termination, then the end date-time is truncated.

Each period requires at least 1,152 CGM readings.

14.Multiple Comparison/Multiplicity

The primary statistical analysis will be conducted using a p value threshold for significance of 0.05. For the secondary, subgroup, and exploratory efficacy analyses, the false discovery rate will be controlled using the adaptive Benjamini-Hochberg procedure. Analyses will be grouped into the following false discovery rate categories:

- Secondary efficacy endpoints
- Exploratory endpoints
- Primary composite endpoint subgroup analyses
- HDP subgroup analyses
- LGA subgroup analyses
- CGM metric by treatment group analyses
- CGM metric by APO analyses

P-values from the safety analyses, sensitivity analyses, and per-protocol analyses will not be adjusted for multiple comparisons.

15.Additional Tabulations and Analyses

The following tabulations will be performed according to treatment group:

- Baseline demographic and clinical characteristics:
    - Age at randomization
    - Ethnicity
    - Race
    - BMI
    - HbA1c
    - Screening CGM sensor type
    - Highest education level
    - Health insurance type

    - Gestational age at randomization
    - Polycystic Ovary Syndrome (PCOS) history
    - Gestational hypertension/pre-eclampsia history
    - Renal disease history
    - Chronic hypertension type
    - Low-dose aspirin use
    - Number of previous pregnancies
    - Number of previous births
    - Number of term births (≥37 weeks)
    - Number of early pre-term births (23-<34 weeks)
    - Number of late pre-term births (34-<37 weeks)
    - Number of indicated pre-term births
    - Number of abortions/miscarriages <14 weeks
    - Number of 2nd trimester (14 and <23 weeks) losses
    - Assisted reproduction type
- Flow chart accounting for all participants
- Protocol deviations
- Device issues
- Post-randomization CGM data availability (will include all CGM data collected for both the Glucose Lowering and Usual Care group)
- Type of labor onset
- Delivery mode
- Gestational age at delivery
- Infant birth weight
- Infant birth length
- Infant sex
- Miscarriage, stillbirth, or termination
    - Gestational age at miscarriage, stillbirth, or termination
- SGA (GROW definition)
- Length of maternal hospital admission
- Prevalence of antenatal corticosteroids used to enhance fetal lung maturity
- Neonatal hypoglycemia

The following will be tabulated in the Usual Care group only:

- Diagnosis of GDM and treatment
- Unblinded CGM use

The following will be tabulated in the Glucose Lowering group only:

- Percent CGM use
- Completion rate of weekly CGM reviews, overall and by site
- Reasons for not completing weekly CGM reviews
- Reasons for discontinuing CGM

8.3.2 Glucose Lowering Treatments in Glucose Lowering Group

The following glucose lowering treatment-related outcomes will be tabulated:

- Percentage of participants who started metformin at any point during follow-up
    - Gestational week of metformin start
    - Metformin dosage
    - Reason for metformin start or adjustment at any point during follow-up
    - Number of times metformin dosage changed
    - Percentage of participants on metformin who stopped metformin at any point during follow-up
- Percentage of participants who started long/intermediate acting insulin at any point during follow-up
    - Gestational week of long/intermediate-acting insulin start
    - Reason for long/intermediate-acting insulin start or adjustment at any point during follow-up
    - Number of times long/intermediate-acting insulin dosage changed
    - Percentage of participants on long/intermediate-acting insulin who stopped it at any point during follow-up
    - Average daily long/intermediate-acting insulin dosage
- Percentage of participants who started short/rapid-acting insulin at any point during follow-up
    - Gestational week of short/rapid-acting insulin start
    - Reason for short/rapid-acting insulin start or adjustment at any point during follow-up
    - Number of times short/rapid-acting insulin dosage changed
    - Percentage of participants on short/rapid-acting insulin who stopped it at any point during follow-up
    - Average daily short/rapid-acting insulin dosage