COMPARISON OF TWO SCREENING STRATEGIES FOR GESTATIONAL DIABETES-GDM2 TRIAL

Statistical Analysis Plan

NCT 02308138

Document Date: April 29, 2020

GDM2 Trial Final Statistical Analysis Plan

Study Design & Objectives[1]. Gestational diabetes mellitus (GDM) affects approximately 5 to 7% of pregnancies and is associated with increased risk for fetal overgrowth, cesarean delivery, birth trauma, and pre-eclampsia. GDM is commonly diagnosed in the US using a two-step screening and confirmatory approach, whereas a one-step approach is commonly used outside the US. Recent guidelines have suggested that the one-step approach should be used to diagnose GDM, although concern that this may increase the prevalence of GDM to approximately 18%—as well as the lack of clinical trials-based evidence regarding the difference in perinatal outcomes–has led to skepticism about adopting the one-step approach. The Comparison of Two Screening Strategies for Gestational Diabetes (GDM2) Trial is a single-center, parallel-arm, comparative effectiveness trial of one-step versus two-step GDM screening strategies on the rate of adverse perinatal outcomes including: large-for-gestational age (LGA) deliveries (primary outcome), small-for-gestational age (SGA), macrosomia, cesarean delivery, fetal growth and body composition, and maternal and neonatal composite outcomes. This paper describes the design and analysis plan of the GDM2 Trial as well as overall challenges in assessing the impact of screening and diagnosis strategy on adverse pregnancy outcomes. We will also assess whether the additional women diagnosed with the one-step approach benefit from treatment as assessed by metabolic profiles at one year postpartum. Ultimately, this study will provide the necessary evidence for establishing universal guidelines for GDM diagnosis and implementation into clinical care. The impetus for the GDM2 Trial is that women with mild glucose intolerance who are otherwise considered normal by the two-step approach would be identified by the one-step approach and subsequently treated with anticipated improved perinatal outcomes. The primary hypothesis is twofold: 1) Using an intention-to-treat approach, we will assess whether women diagnosed with the GDM one-step diagnostic approach have lower rates of LGA compared to those randomized to the two-step approach; and 2) We hypothesize that women classified as "No GDM" by the one-step approach will have lower rates of LGA compared to their counterparts in the two-step approach. We will also follow up a sub-cohort of women at one year postpartum where we hypothesize that women diagnosed with the one-step approach will have better metabolic profiles, compared to the two-step approach.

The GDM2 Trial is a single-center, parallel-arm, comparative effectiveness trial comparing two GDM screening strategies (one-step versus two-step approach) on the rate of adverse perinatal outcomes (e.g. large for gestational age [LGA], macrosomia) and differences at one year postpartum in the mothers' metabolic profiles and the infants' growth. We recruit women from 24 to 28 weeks GA to receive a one-hour, 50 g screening test; these women are then randomized in a 1:1 ratio to either the two-hour, 75 g or three-hour, 100 g diagnostic test. Women and their infants are followed for up to one year postpartum.

Sample Size Calculation. The first co-primary hypothesis compares LGA incidence between screening methods. We expect to see the rate of LGA in the two-step approach (regardless of GDM diagnoses) to be approximately 15%, similar to that observed in a previous cohort study. To detect an absolute risk reduction of 7% (to 8% in the one-step approach) for LGA, a sample size of 460 per arm is required accounting for 15% participant attrition. The absolute risk reduction and corresponding odds ratio is considered small to moderate.

For the second co-primary hypothesis, we anticipate 84% (n = 329) of women in the one-step arm and 95% (n = 372) of those in the two-step arm will be classified as "No GDM". Based on an estimated rate of LGA of approximately 13% among the "No GDM" women in the two-step arm, our anticipated sample size yields at least 80% power to detect a 7.1% absolute risk difference between screening arms. Since the analysis of the second hypothesis will be conducted on a subset of women after randomization, there could be imbalances between the "No GDM" groups

in both arms on baseline characteristics. The magnitude of this imbalance will be quantified by a variance inflation factor (VIF), which is the coefficient of determination when the study arms (one-step versus two-step) are regressed on the covariates of interest. Assuming a large VIF, the detectable risk difference of LGA and macrosomia increases to 8.3%. To preserve an overall type I error of 5%, each of the co-primary hypotheses will be conducted at the 2.5% significance level.

Interim & Final Analyses. This study will not have planned interim looks for the primary outcome. The final analyses will be conducted once study follow-up is complete, after all data is cleaned, and once the study database is locked.

Hypotheses. The primary hypothesis is twofold: 1) Using an intention-to-treat approach, we will assess whether women diagnosed with the GDM one-step diagnostic approach have lower rates of LGA compared to those randomized to the two-step approach; and 2) We hypothesize that women classified as "No GDM" by the one-step approach will have lower rates of LGA compared to their counterparts in the two-step approach. We will also follow up a sub-cohort of women at one year postpartum where we hypothesize that women diagnosed with the one-step approach will have better metabolic profiles, compared to the two-step approach.

Analysis Sets. The Full Analysis Set (FAS) will be based on an intention-to-treat (ITT) analysis, where all women are analyzed according to the study arm they were randomized to. The Per-Protocol (PP) analysis set will be defined as women who received the screening method they were randomized to. The As-Treated (AT) analysis set will be the same as the FAS, but the women will be analyzed according to the screening method they actually received. The Safety Analysis Set (SAS) will be a subset of the FAS and include all women who competed the intervention.

Study Outcomes. Primary and secondary outcomes will be abstracted from the electronic health record with the assistance of the Center for Assistance in Research using eRecord (CARe) system. Our primary outcome is large for gestational age (LGA), which is defined as a birthweight ≥ 90th percentile for gestational age after adjusting for sex differences. Secondary outcomes include: 1) small for gestational age (SGA), defined as ≤ 10th percentile for gestational age after adjusting for sex differences; 2) macrosomia, defined as birthweight ≥ 4000 g; 3) primary or repeat cesarean

Outcomes	Definition	
Primary		
Large for Gestational Age (LGA)	Birth weight ≥ 90th percentile for GA assessed using sex-specific US birth weight curves corrected for implausible ultrasound estimates.[11]	
Secondary		
Small for Gestational Age (SGA)	Birth weight ≤ 10th percentile for GA assessed using sex-specific US birth weight curves corrected for implausible ultrasound estimates.[11]	
Macrosomia	Defined as birth weight ≥ 4000 g.	
Cesarean Delivery[#]	Primary or repeat cesarean delivery.	
Fetal Growth & Body Composition	Measured using length, weight, head circumference, and a peapod assessment.	
Maternal Composite[##]	Any of the following: (a) third- or fourth-degree vaginal lacerations; (b) postpartum hemorrhage (defined as an estimated blood loss of > 1000 mL after vaginal delivery or 1500 mL following cesarean delivery with or without the need for blood transfusion); or (c) preeclampsia/hypertensive disorders of pregnancy (defined according to American College of Obstetrics and Gynecologists diagnostic criteria[12]).	
Neonatal Composite[###]	Any of the following: (a) hypoglycemia with a blood glucose value of < 40mg/dL in the first 24 hours of life; (b) hyperbilirubinemia requiring treatment; (c) still birth or absence of fetal heart tones before delivery; or (d) birth trauma, such as shoulder dystocia or brachial plexus injuries.	

[#] We will also note the primary indication for the cesarean delivery, including prior cesarean delivery with no trial of labor after cesarean (TOLAC) delivery, failure to progress, fetal intolerance to labor, breech presentation, cesarean delivery on maternal request, or abnormal placentation such as placenta previa.
[##] Obtained from delivery records and ICD 9 codes.
[###] Neonatal blood glucose is measured by heel stick within one hour after delivery, and bilirubin is measured transcutaneously starting at 35 hours post-delivery and as needed; these values will be obtained from the laboratory section of the neonatal electronic medical record.

delivery; 4) fetal growth and body composition outcomes (length, weight, and head circumference of the baby); 5) a maternal composite outcome; and 6) a neonatal composite outcome.

Handling of Missing Values. Reasons for attrition will be investigated at the time participants undergo the oral glucose tolerance test (OGTT), as well as at the end of the study. Missing outcome data will be investigated to see if they are ignorable, but we do not anticipate this to be an issue due to the integrated EHR system to ascertain clinical outcomes. If appropriate, we will consider various approaches to account for missing data, such as multiple imputations. Much of the missing data involving the outcomes of interest will be constrained by the natural design of the study, which includes a short time frame between study visits and the use of electronic health records to obtain pregnancy outcomes and healthcare utilization.

Statistical Analyses. Each of the primary and secondary outcomes will be described within each study arm using sample means or sample proportions along with 95% confidence intervals. Continuous outcomes will be assessed for departures from normality, and suitable transformations will be used if needed. Demographic and clinical characteristics will be compared between arms at baseline using two-sample t-tests and chi-square tests, as appropriate. Any variables that are significantly associated with either study arm will be included as covariates in primary and secondary analyses. Secondary and exploratory outcomes will be analyzed at the 5% significance level.

Primary co-hypothesis: The primary analyses will use a logistic regression model to quantify the probability of LGA as a function of the study arm (one-step versus two-step) and clinic type (resident versus non-resident). For the first co-primary hypothesis, all of the women randomized will be considered in the analysis, while only the women classified as "No GDM" will be considered for the analysis of the second co-primary hypothesis. To preserve an overall type I error of 5%, each of the co-primary hypotheses will be conducted at the 2.5% significance level. Analyses of the secondary perinatal outcomes of cesarean delivery, fetal growth, and the maternal and neonatal composite outcome variables will use a logistic regression model to quantify the probability of the outcome as a function of the study arm and clinic type. We will use a linear regression model for the fetal growth and body composition outcomes. Finally, we will compare the proportion of women with serious adverse events (SAEs) between study arms using logistic regression.

Secondary hypothesis: A secondary hypothesis is that women who undergo testing with the one-step approach will have greater improvements in their metabolic profile (cytokines, weight change, insulin resistance, and beta cell function) at one year postpartum compared to women in the two-step approach, because more women will be diagnosed with and treated for GDM. Metabolic profiles will be analyzed using linear regression with the change in outcome at one year postpartum modeled as a function of the study arm, clinic type, and any baseline covariates that significantly differ between arms. Analysis of Covariance (ANCOVA) will be used to model the one-year metabolic profile as a function of baseline metabolic profile, study arm, clinic type, and any baseline covariates related to study arm.

Secondary and Exploratory Analyses: The rate of healthcare utilization variables will be compared between screening arms using Poisson regression due to the count-nature of these outcomes. We will use logistic regression to model the outcomes of labor induction and admission to the neonatal intensive care unit, as well as to compare the rate of 6- to 12-week postpartum diabetes mellitus testing between screening arms. Finally, we will assess the sensitivity and specificity of the one-hour, 50 g GCT against each of the OGTT study arms.

References.

1. Abebe KZ, Scifres C, Simhan HN, et al. Comparison of Two Screening Strategies for Gestational Diabetes (GDM 2) Trial: Design and rationale. *Contemp Clin Trials*. 2017;62(January):43-49. doi:10.1016/j.cct.2017.08.012