Early Gestational Diabetes Screening in the Gravid Obese Woman (EGGO)

Statistical Analysis Plan (v1.0)

Prepared by

Biostatistics and Data Management Core (BDMC)
Center for Women's Reproductive Health (CWRH)
UAB

Revised March 29, 2017

Table of Contents

EGGO STATISTICAL ANALYSIS PLAN

1. INTRODUCTION

Over 1/3 of reproductive age women are obese. Obese women have higher rates of poor pregnancy outcomes including stillbirth, fetal growth problems, diabetes, hypertension, and maternal death. Gestational diabetes (GDM) is a frequent pregnancy complication in obese women and is associated with increased risk of large infants, cesarean delivery, and high blood pressure. Although treatment of GDM improves these outcomes, obese women seem to benefit less from treatment than non-obese women. This is perhaps because obese women develop GDM earlier in pregnancy than normal weight women, leading to the baby being exposed to elevated blood sugars for a longer period prior to diagnosis and treatment.

The purpose of this study is to determine if screening obese women for gestational diabetes early in their pregnancy (at 14-20 weeks) improves pregnancy outcomes compared to screening obese women for gestational diabetes at the routine time (24-28 weeks). This study was created at UAB and will enroll 1,160 obese women at two study sites: UAB and Ochsner (New Orleans, LA). A confidential, computer-generated randomization scheme will be prepared by a designated study biostatistician. Randomization will be stratified by morbid obesity (BMI≥40.0 vs BMI <40). Randomization will also be stratified for both participating sites.

2. DESCRIPTIVE STATISTICS

Patient characteristics at randomization will be summarized by randomization group. The two study groups will be labeled as Early (for early gestational diabetes screening 14-18 weeks) and Routine (for routine gestational diabetes screening 20-24 weeks). For continuous variables, means/medians and standard deviations/interquartile ranges will be reported. To assess and/or identify covariates for which adjusted sensitivity analyses might be conducted, Student t-tests will be used to compare means between study groups. Where appropriate, medians and quartiles will be reported and the Wilcoxon rank sum test will be used as an alternative comparison procedure. Categorical measures will be presented as counts and percentages and will be compared using the χ^2 tests of association to identify potential group differences. For rare outcomes such that the χ^2 test of association is not appropriate, Fisher's exact test will be used. Balance overall is expected because of the large sample size and we expect approximately 5% to be different by chance, since we are not adjusting these baseline comparisons for multiple testing. Any group characteristics that are identified as statistically significantly different between the two groups at a 0.05 level of significance will be considered as covariates in multivariable models in subsequent analyses of the primary study outcome.

3. PRIMARY HYPOTHESIS

3.1. Primary Hypothesis #1: Impact of Early GDM Screening on the Primary Outcome (a Composite of Key Perinatal Outcomes)

H_0: There will be no difference in the rate of the primary outcome (quantified by the composite outcome of key perinatal outcomes) in EGGO participants randomized to the Early screening group compared to participants randomized to the Routine screening Group.

3.1.1. Primary Outcome: The primary outcome is a composite of the following items: macrosomia, primary cesarean, pregnancy induced hypertension (gestational hypertension or preeclampsia), shoulder dystocia, neonatal hyperbilirubinemia, and neonatal hypoglycemia. The occurrence of 1 or more of these items will be considered an occurrence of the primary study outcome. Individual definitions of these outcomes are listed below.

a. Macrosomia: Will be defined by measurement of infant birth weight. Macrosomia is a birth weight >4000 g.

b. Primary cesarean: A primary cesarean delivery for any indication will be considered as having the outcome of interest.

c. Gestational hypertension: Defined by systolic blood pressure ≥140 or diastolic blood pressure ≥90 without proteinuria (urine protein/creatinine ratio <0.19 or mg).

d. Preeclampsia: Defined by systolic blood pressure ≥140 or diastolic blood pressure ≥90 with proteinuria (urine protein/creatinine ratio ≥0.30). Preeclampsia diagnosis will also be refined by mild or severe. Severe preeclampsia will be defined as systolic blood pressure ≥160, diastolic blood pressure ≥110, platelets <100, AST >80 serum creatinine >1.2.

e. Shoulder dystocia: As documented by delivering physician; defined as requiring more than routine downward traction to achieve delivery of anterior shoulder. The length of the shoulder dystocia and maneuvers required to relieve the shoulder dystocia will be documented.

f. Neonatal hyperbilirubinemia: Neonatal bilirubin levels will be recorded. Hyperbilirubinemia will be defined >95th percentile gestational age and hour of life. The treatment of hyperbilirubinemia (i.e. phototherapy) will also be recorded.

g. Neonatal hypoglycemia. Blood sugar levels will be abstracted from neonatal charts. Neonatal hypoglycemia will be defined as blood sugar <35mg/dL.

3.1.2. Analysis Plan: This hypothesis effect of Early versus Routine GDM screening in obese pregnant women. The χ^2 test of association will be used to evaluate whether the rate of the primary outcome differs between the treatment groups. The primary analysis will be an intention-to-treat analysis of all individuals randomized to the two treatment groups, regardless of whether evaluation plans were followed. Risk ratios (RR) and 95% confidence intervals (CIs) will be computed and presented.

Standard comparison of the covariate characteristics at baseline between study groups will be undertaken (Section 2) to assess the randomization balance on measured covariates. Tests of significance will be two-sided and evaluated at the 0.05 level of significance. Since the sample size for EGGO is large, it is unlikely that "important" covariates (such as race, BMI or age) will be imbalanced through the

randomization process. Further, randomization is stratified by BMI (BMI <40 and BMI ≥40) and site (UAB and Oschsner).

In the unlikely event that imbalance occurs, adjustment for covariates will be made by modeling the log-odds of the outcome using multivariable logistic regression including terms indicating group membership and each of the imbalanced covariates. In other words, the log-odds of the primary outcome for any individual will be modeled using the following form:

$$\ln\left(\frac{P(outcome)}{1-P(outcome)}\right) = \beta_0 + \beta_1 x_1 + \sum \beta_j z_j.$$

Above, $x_1 = 1$ if the patient is in the Early group and equals 0 if the patient is in the Routine group. The set of z_j's consists of all covariates in the model including site. Hence, under this model, the term β_1 represents the difference in the log-odds between the two groups. To address the hypothesis of interest for primary hypothesis #1, we will perform a statistical test of H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$. Quantification of the treatment effect will be performed by exponentiation of the estimated coefficient, $\exp(\beta_1)$, to obtain an estimate of the odds ratio. The confidence interval will be determined similarly.

The stratification by BMI and by site allows for the investigation of differential effects across sites and across BMI thresholds. The presence of a homogeneous effect across center and across BMI thresholds will be examined using the Breslow-Day test.

3.1.3. Sample Size for Primary Outcome: Based on prior studies performed at UAB, the incidence of macrosomia, cesarean delivery, shoulder dystocia, and preeclampsia in obese women with gestational diabetes was 56%. Using a conservative estimate of these outcomes as 50%, to detect a 50% reduction in the incidence of adverse outcomes (α=0.05, β=0.80), 58 gestational diabetics per group are necessary.

The diagnosis of GDM will occur after patients in this study have been screened for GDM. As a result, a large number of patients must be randomized to each diagnostic procedure in order to achieve 58 gestational diabetics in each group. Prior reports of the incidence of GDM are between 6-20% depending on ethnicity and the screening and diagnostic criteria used. Using an incidence of 10% GDM in obese women, 580 subjects per screening group are necessary. Approximately 3,100 women per year attend the UAB clinic for obstetric visits, and approximately 40% of these women are obese. Assuming a 50% acceptance of enrollment, we anticipate enrolling 620 women per year. Therefore, we feel that we can enroll 1,160 subjects in 3 years.

3.2. Interim Monitoring: One interim analysis, evaluating both efficacy and futility, will be conducted after 600 women have completed the study (300 in each group have delivered). Under the assumption that 10% of obese women have GDM, we expect to have approximately 30 gestational diabetics in each group. Using an O'Brien-Fleming alpha spending function and one interim analysis, we will use the following critical stopping values:

At the interim look: If $\chi^2 > 7.82$ (i.e. $p < 0.005$) then reject H_0 and consider stopping the trial; Otherwise, continue.

At the final look: If $\chi^2 > 3.91$ (i.e. $p < 0.048$) then reject H_0.

We do not propose a formal stopping rule for safety or futility, but note that the DSMB will be presented with safety data and information on study progress. In addition, we note that the overall sample size of 58 gestational diabetics per group will provide 80% power for the hypothesized effect size using a modified alpha level of 0.048 at the final analysis. Conditional power estimates may be provided to assist the DSMB in their evaluation of efficacy and futility.

The interim analysis will focus only on outcomes specified in Section 3 of this Statistical Analysis Plan. Secondary outcomes may be presented at that time in aggregate form (only safety events will be presented by study group). Formal statistical hypothesis testing for those secondary outcomes will be conducted at the conclusion of the trial.

At the time of the interim analysis, the Ochsner study site will have contributed a small number of patients to the study. These patients will be excluded from the interim analysis evaluation.

4. Secondary Hypotheses and Sub-Analyses

For the primary outcome, separate analyses will be conducted to evaluate each component of the composite. In addition, the rates of preterm birth <37 weeks will be evaluated. Specific aims 2 and 3 will also be investigated.

4.1. Specific Aim 2: To test the hypothesis that a lower diagnostic threshold for GDM at 14-18 weeks will result in improved detection of GDM and reduce the need for third-trimester testing.

4.1.1. Secondary Hypothesis #1: Identification of a blood sugar value identified at early GDM screening that detects GDM outcomes and reduces the need for third-trimester screening

H_0: There will be no association between early screening blood glucose levels and the rates of (a) hypoglycemic medication use, (b) the primary outcome, (c) macrosomia, and (d) third trimester testing in EGGO participants randomized to the Early screening group

4.1.2. Statistical Analysis: Therefore, we will evaluate the ability of an early GDM screen to predict several outcomes. We will compare early GDM screening to GDM testing at 24-28 weeks with respect to the following outcomes: the need for hypoglycemic medications during pregnancy, the composite outcome from Specific Aim #1, macrosomia (a common complication of GDM), and third trimester testing. Logistic regression models will also be considered. Receiver operator characteristics (ROC) curves will be created to identify an ideal cut point. The area under the curve (the c statistic) will be used to assess the models.

4.2. Specific Aim 3: To test the hypothesis that 1,5-anhydroglucitol, a sensitive marker of recent hyperglycemic excursions, can be used as a simple and sensitive serum test for GDM in the obese population.

4.2.1. Secondary Hypothesis #2: Association between AG levels and GDM diagnosis

H_0: There will be no difference in the mean AG levels in EGGO participants with GDM and in EGGO participants without GDM.

4.2.2. Statistical Analysis: The AG levels of women with and without GDM will be compared using a student's t-test or Mann-Whitney U test, as appropriate. A receiver operator characteristics curve will be created for AG to predict GDM (as diagnosed by a positive test at 24-28 weeks), need for hypoglycemic medications, and macrosomia. The area under the ROC curve will be calculated and an ideal cut-point for AG will be determined. The sensitivity, specificity, positive predictive value, and negative predictive value of AG will be determined.

5. Additional Analyses

The EGGO study provides an opportunity for investigation beyond the research hypotheses explicitly stated in the original grant application and study protocol. The EGGO Steering Committee will assess the pre-specified secondary analysis concepts listed below for feasibility and potential for scientific impact. Approved concepts will be prioritized; the committee may not approve all concepts. Furthermore the evaluation of some concepts is dependent upon knowledge of the study's primary results and sufficient numbers of subjects enrolled with specific characteristics.

Secondary analysis proposals may include:

3. Any comparisons of non-GDM and GDM women
4. Any lab comparisons
5. Anything with obstetric complications as the outcome
6. Neonatal outcomes
7. Postpartum outcomes

6. References

Jennison C and Turnbull BW. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC: Boca Raton, 2000.