



**Design Proposal for Nested Case Control Study  
Evaluating Association Between Herbicide Exposures  
and Adverse Pregnancy and Birth Outcomes  
in the nuMoM2b Study Cohort**

Draft

December 23, 2020

**Melissa Perry and Marlaina Freisthler**  
Milken Institute School of Public Health  
George Washington University

The objectives of this study are to investigate associations between herbicide exposure and adverse pregnancy and birth outcomes among pregnant women living in the Heartland of the United States, a 13 state Midwestern region with intense agricultural activity.

**Research Question:** Is there an association between herbicide exposure and adverse pregnancy or birth outcomes among pregnant women in the Heartland?

**Specific Aim 1:** To examine the association between biomarkers of herbicide exposure and adverse pregnancy or birth outcomes among participants of a sample cohort of the Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-Be (nuMoM2b) study located in the Heartland Study region of interest (Ohio, Indiana, Illinois).

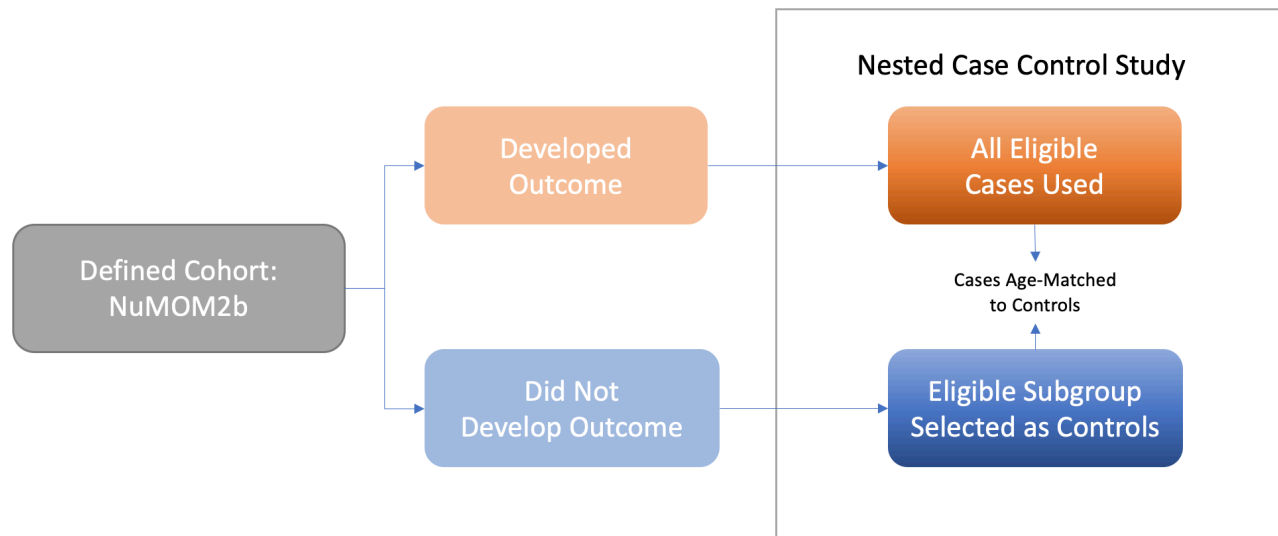
**Null Hypothesis:** There is no difference in biomarkers of herbicide exposure between women with adverse pregnancy or birth outcomes and women with normal pregnancy and birth outcomes in the nuMoM2b cohort.

**Alternative Hypothesis:** There is a difference in biomarkers of herbicide exposure between women with adverse pregnancy or birth outcomes and women with normal pregnancy and birth outcomes in the nuMoM2b cohort.

**Research Ethics:** The research proposal will be submitted to the Institutional Review Board at the George Washington University and Indiana University, as well as at any other institution where the study will be conducted. Informed consent from all women in the nuMoM2b cohort covered future unspecified use of biospecimens, and thus no new consenting is required. The consent forms also specifically state that research participants in nuMoM2b will not be notified of results individually.

**Study Design:** The study will be conducted as a nested case control study. All eligible cases of the pregnancy or birth outcome(s) of interest will be age-matched to controls who experienced no adverse pregnancy or birth outcomes in the nuMoM2b study. The strength of the nested case control (NCC) approach is its efficiency in that the cases of adverse pregnancy and birth outcomes have already been established in this cohort. This substantially reduces the cost and time involved in recruiting cases where events of interest are relatively rare. An important disadvantage is that only one outcome can be studied at a time. Each outcome requires a separate (but not necessarily unique) set of controls.

### Diagram of Proposed Nested Case Control Study Design:



**Data Source:** First trimester urine specimens will be tested for specific herbicide biomarkers at Chimiste Responsable division Environnement Centre de Toxicologie du Québec (CTQ) Unité Laboratoire de Toxicologie (Éric Gaudreau, M.Sc.) using protocols consistent with CDC multianalyte and glyphosate testing methods. First trimester urine was selected in consideration of the fact that embryonic development is the gestational period most sensitive to chemical exposure. Results of the biomarker analysis will be linked with data related to pregnancy and birth outcomes collected under the nuMoM2b study protocol.

**Study Population and Setting:** All cohort participants who developed the relevant case criteria will be considered a case for analytical purposes. Cases will be matched by age with a control from the group of the sample cohort that did not experience any adverse pregnancy or birth outcomes and also reported being a non-smoker. A separate (but not necessarily unique) set of controls will be selected for each outcome considered. To the extent possible, the number of controls matched to each case will be based on the ratio necessary to achieve a power of at least 80%. A ratio of 1:1 is preferred when feasible to limit unnecessary costs.

**Exposure:** Exposure to herbicides will be estimated using biomarker results. Quantified and imputed results are expected to be approximately normally distributed. Values below the limits of detection will be taken from machine read values. Initially, we will receive results quantifying glyphosate, AMPA, glufosinate, and its major metabolite. As soon as a multi-analyte method is available quantifying 2,4-D, dicamba, and other pesticides, the urine samples will also be tested using this method. Statistical analysis will then be performed to determine associations between adverse birth outcomes and pesticide biomarkers, individually and jointly, detected across all samples.

**Outcomes:** Outcomes of interest include spontaneous preterm birth, spontaneous miscarriage/stillbirth, gestational diabetes, and (see Table 1 below). Outcomes will be dichotomized for purposes of case identification.

Illinois, Indiana & Ohio Sites, n=2749	Total	Percent
Spontaneous Preterm Birth	35	1.27
Spontaneous Abortion/Stillbirth	47	1.71
Diagnosis of GDM	131	4.77

**Table 1.** Case outcomes in selected cohort.

The following case definitions apply for purposes of case identification:

Preterm birth is defined as delivery of a liveborn or stillborn infant for any cause between 20 and 0/7 and 36 and 6/7 weeks of gestation. *Spontaneous preterm birth* is defined as delivery occurring subsequent to spontaneous onset of preterm labor or premature rupture of the membranes (PROM) or fetal membrane prolapse.

*Stillbirth* is defined as birth of a dead fetus at 20 0/7 weeks of gestation or greater.

*Spontaneous abortion* is defined as the delivery of a liveborn or fetus experiencing fetal death for any cause before 20 and 0/7 weeks of gestation.

*Gestational diabetes* includes patients that were diagnosed with gestational diabetes during the course of the pregnancy, excluding women with pregestational diabetes. That is, the definition includes only Class A gestational diabetics as categorized under White's classification.

**Important Covariates:** Geographical location of residence is expected to be causally associated with the exposure variable, as is the time of the year in which the sample was collected (herbicide spray season vs. non-spray season). Socioeconomic status may also be causally associated with the exposure variable in that it may influence dietary behaviors (e.g., produce consumption, water source). Age and gravidity/parity are also associated with outcomes, but in this case will be controlled through study design (i.e., all participants are nulliparous; cases will be age matched to controls) and will therefore be unavailable for analysis in the regression model. Other potential covariates include smoking status, adiposity, medical comorbidities (e.g., Polycystic Ovarian Syndrome (PCOS), autoimmune disorders), history of *In Vitro* Fertilization, and race/racism.

**Sample Size and Power Calculations:** In the following tables, a given case to control ratio was only considered for each outcome if the required number of analyses could be performed within budget ( $n \leq 1700$ ). Odds ratio for each power calculation was estimated from the literature, and an alternate odds ratio was estimated for the full cohort using observed differences in case occurrence based by geographical location ("agricultural site") as a proxy. This method is

expected to be a less accurate estimate for odds ratio of case outcome. Power will likely be altered somewhat by choice of case definition.

Case	Case N	Total N	Ratio	Estimated OR	OR Source	Power
Spontaneous Preterm Birth	35	70	1:1	1.8	Literature	0.816
Spontaneous Miscarriage/ Stillbirth	47	235	1:4	1.4	Literature	0.485
Gestational Diabetes	131	262	1:1	2.2	Literature	0.999

**Table 2. Power Estimates.** Power calculation based on cases and controls from Heartland Study geographic region with minimum case-control ratio. Odds ratios based on estimates from literature.

### Statistical Analysis Plan:

Data on urine biomarkers will be procured from CTQ in its standard form and linked to existing REDCap databases containing nuMoM2b data. Two-sided tests with alpha=0.05 a priori level of statistical significance will be utilized throughout the statistical analysis, which will be performed using SAS version 9.4.

Statistical tests will be applied to compare baseline characteristics of case and control groups by describing the frequencies and patterns of characteristics of participants, exposures, and outcomes. Mean biomarker levels and standard deviations will be calculated for case and control groups. Differences in biomarker level will be evaluated for the primary outcome of interest using paired difference t-test, with statistically significant difference being demonstrated at a level of  $p < 0.05$ . Biomarker level will also be evaluated based on expected covariates, such as geographic location, socioeconomic status, adiposity, smoking status, medical comorbidities, and IVF history. Box plots will be generated for biomarker level by categorical variables. Differences in outcome will also be evaluated based on covariates using chi-square test, with differences considered statistically significant at a level of  $p < 0.05$ . Bar and pie charts will be produced for visualization of any differences in outcomes on categorical variables.

Inferential statistical analysis will be performed using conditional logistic regression to evaluate the effect of exposure on outcome in matched pairs, while controlling for any statistically relevant covariates. Odds ratios with 95% confidence intervals will be calculated. Linear regression will be explored based on the continuous nature of the outcome such as infant birthweight.