

## Statistical Issues in Modeling Pregnancy Outcome Data

Germaine M. Buck,<sup>1</sup> Vanja M. Dukic,<sup>2</sup> Patrick J. Heagerty,<sup>3</sup> Thomas A. Louis,<sup>4</sup> Courtney J. Lynch,<sup>1</sup> Louise M. Ryan,<sup>5</sup> Enrique F. Schisterman,<sup>1</sup> Ann Trumble,<sup>1</sup> and the Pregnancy Modeling Working Group  
<sup>1</sup>National Institute of Child Health & Human Development, Department of Health and Human Services;  
<sup>2</sup>University of Chicago; <sup>3</sup>University of Washington; <sup>4</sup>Johns Hopkins Bloomberg School of Public Health;  
<sup>5</sup>Harvard University.

Risk of repeating an adverse pregnancy outcome is approximately two-fold, underscoring the need for statistical designs and analyses of studies that accommodate the dependent (correlated) nature of pregnancy outcomes. Failure to adequately address this dependency may lead to biased and inefficient exposure effect estimates, possibly masking important relations. Past analytic approaches have included ignoring the dependency, side-stepping it by analyzing only one pregnancy per women or by treating past history as a covariate. Newer approaches such as the use of robust standard errors or explicit modeling of the correlation structure enable researchers to account for the dependency when estimating risk or making clinical predictions.

To address how best to model a sequence of pregnancy outcomes we utilized data from the U.S. Collaborative Perinatal Project (CPP) to identify determinants of infant birth weight (in grams). The CPP Study enrolled approximately 48,197 pregnant women at one of 12 clinical centers in the United States between 1959-1964 and followed their children to approximately 7-8 years of age. For study purposes, we restricted our sample to the 2,717 primigravida women whom had two or more consecutive singleton pregnancies (n=4,915 total pregnancies) prospectively followed by the CPP Study and for whom complete information was available on study covariates (i.e., clinical site, maternal age, race, pre-pregnancy weight, cigarette smoking, family income, infant sex). Modeling strategies included generalized estimating equations (GEE) and mixed models with a variety of correlation structures (e.g., AR1 with random intercept or Toeplitz).

Differences in estimated effect size (slopes) and robustly estimated standard errors (subscripts) were observed across the various modeling approaches for known biologic determinants of fetal growth.

Maternal Characteristic	Ignoring Prior History (GEE) Slope <sub>(SE)</sub>	Modeling Prior History (GEE) Slope <sub>(SE)</sub>	Random Pregnancy (GEE) Slope <sub>(SE)</sub>	Random Intercept (Mixed Model) Slope <sub>(SE)</sub>
Smoking 1+ppd	-259 <sub>(28)</sub>	-199 <sub>(22)</sub>	-220 <sub>(34)</sub>	-217 <sub>(25)</sub>
Black race	-234 <sub>(34)</sub>	-180 <sub>(26)</sub>	-199 <sub>(35)</sub>	-224 <sub>(31)</sub>
Age ≥30 years	-141 <sub>(49)</sub>	-118 <sub>(39)</sub>	-163 <sub>(63)</sub>	-146 <sub>(55)</sub>

Our evaluations support use of mixed models or GEEs with robust standard error estimates for assessing determinants of birth weight in the context of past reproductive history. Our results underscore the importance of *a priori* model specification, the inefficiency associated with restricting analyses to one pregnancy per woman, and the attenuating effects of including past reproductive history as a covariate in the model. In the absence of a strong rationale for ignoring past reproductive performance, every effort should be made to make full and proper use of reproductive history information when assessing risk factors for adverse reproductive events.

Start and end date: March 2002 – July 2004.

Presentation(s): None to date.

Peer-reviewed publication(s): None to date.

Buck Louis GM, Dukic VM, Heagerty PJ, Louis TA, Lynch CD, Ryan LM, Schisterman EF, Trumble A, and the Pregnancy Modeling Working Group. Statistical Issues in Modeling Pregnancy Outcome Data. *Statistical Methods in Medical Research* 15(2), (in press).

Other publication(s): None to date.

Sponsors in addition to the LRI: None.

Abstract revision date: January 2006.