

Multilevel Quantile Function Modeling with Application to Birth Outcomes

Luke B. Smith^{1,*}, Montserrat Fuentes¹, Brian J. Reich¹, Amy H. Herring², and Peter H. Langlois³

¹Department of Statistics, North Carolina State University, Raleigh, North Carolina, 27695-8203, U.S.A.

²Department of Biostatistics, University of North Carolina at Chapel Hill,

Chapel Hill, North Carolina, 27599-7420, U.S.A.

³Texas Department of State Health Services, Austin, Texas 78714-9347, U.S.A.

**email*: luke.smith@ncsu.edu

SUMMARY: Infants born preterm or small for gestational age have elevated rates of morbidity and mortality. Using birth certificate records in Texas from 2002-2004 and Environmental Protection Agency air pollution estimates, we relate the quantile functions of birth weight and gestational age to ozone exposure and multiple predictors, including parental age, race, and education level. We introduce a semi-parametric Bayesian quantile approach that models the full quantile function rather than just a few quantile levels. Our multilevel quantile function model establishes relationships between birth weight and the predictors separately for each week of gestational age and between gestational age and the predictors separately across Texas Public Health Regions. We permit these relationships to vary nonlinearly across gestational age, spatial domain and quantile level and we unite them in a hierarchical model via a basis expansion on the regression coefficients that preserves interpretability. Very low birth weight is a primary concern, so we leverage extreme value theory to supplement our model in the tail of the distribution. Gestational ages are rounded into weekly values, so we present methodology for modeling quantile functions of discrete response data. In a simulation study we show that pooling information across gestational age and quantile level substantially reduces MSE of predictor effects relative to standard frequentist quantile regression. We find that ozone is negatively associated with the lower tail of gestational age in south Texas and across the distribution of birth weight for high gestational ages. Our methods are available in the R package **BSquare**.

KEY WORDS: Birth weight; Discrete; Extremes; Gestational Age; Graphics processing units; Ozone; Quantile.

This paper has been submitted for consideration for publication in *Biometrics*

Supplementary Materials

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

[Figure 16 about here.]

[Figure 17 about here.]

[Figure 18 about here.]

[Figure 19 about here.]

[Figure 20 about here.]

[Figure 21 about here.]

[Figure 22 about here.]

[Figure 23 about here.]

[Figure 24 about here.]

[Figure 25 about here.]

[Figure 26 about here.]

[Figure 27 about here.]

[Figure 28 about here.]

[Figure 29 about here.]

[Figure 30 about here.]

[Figure 31 about here.]

[Figure 32 about here.]

[Figure 33 about here.]

[Figure 34 about here.]

[Figure 35 about here.]

[Figure 36 about here.]

[Figure 37 about here.]

[Figure 38 about here.]

[Figure 39 about here.]

[Figure 40 about here.]

[Figure 41 about here.]

[Figure 42 about here.]

[Figure 43 about here.]

Received October 2013.

Appendix MCMC

Posterior distributions of θ_{mjg} were updated singly via random walk Metropolis using Gaussian candidate distributions with variances tuned to have acceptance probability between 30 and 40%. Posterior distributions of the autocorrelation parameters ρ_{mj} and range parameters ϕ_{mj} were updated using independent Metropolis updates. The basis function means μ_{mj} were given Gaussian priors with mean μ_0 and precision δ_0 . The basis function precisions γ_{mj} were given Gamma priors with shape parameter a and scale parameters b . These means and precisions have conjugate posteriors and were updated using Gibbs sampling.

$$\mu_{mj}|rest \sim N \left(\frac{\gamma_{mj} \mathbf{1}^T \Sigma_{\mathbf{mj}}^{-1} \theta_{\mathbf{mj}}}{\gamma_{mj} \mathbf{1}^T \Sigma_{\mathbf{mj}}^{-1} \mathbf{1} + \delta_0}, (\gamma_{mj} \mathbf{1}^T \Sigma_{\mathbf{mj}}^{-1} \mathbf{1})^{-1} \right)$$

$$\gamma_{mj}|rest \sim \text{Gamma}(G/2 + a, (0.5(\theta_{\mathbf{mj}} - \mathbf{1}\mu_{\mathbf{mj}}) \Sigma_{\mathbf{mj}}^{-1} (\theta_{\mathbf{mj}} - \mathbf{1}\mu_{\mathbf{mj}}) + \mathbf{1}/b)^{-1})$$

where $\theta_{\mathbf{mj}} = \theta_{\mathbf{mj}1}, \dots, \theta_{\mathbf{mj}G}$.

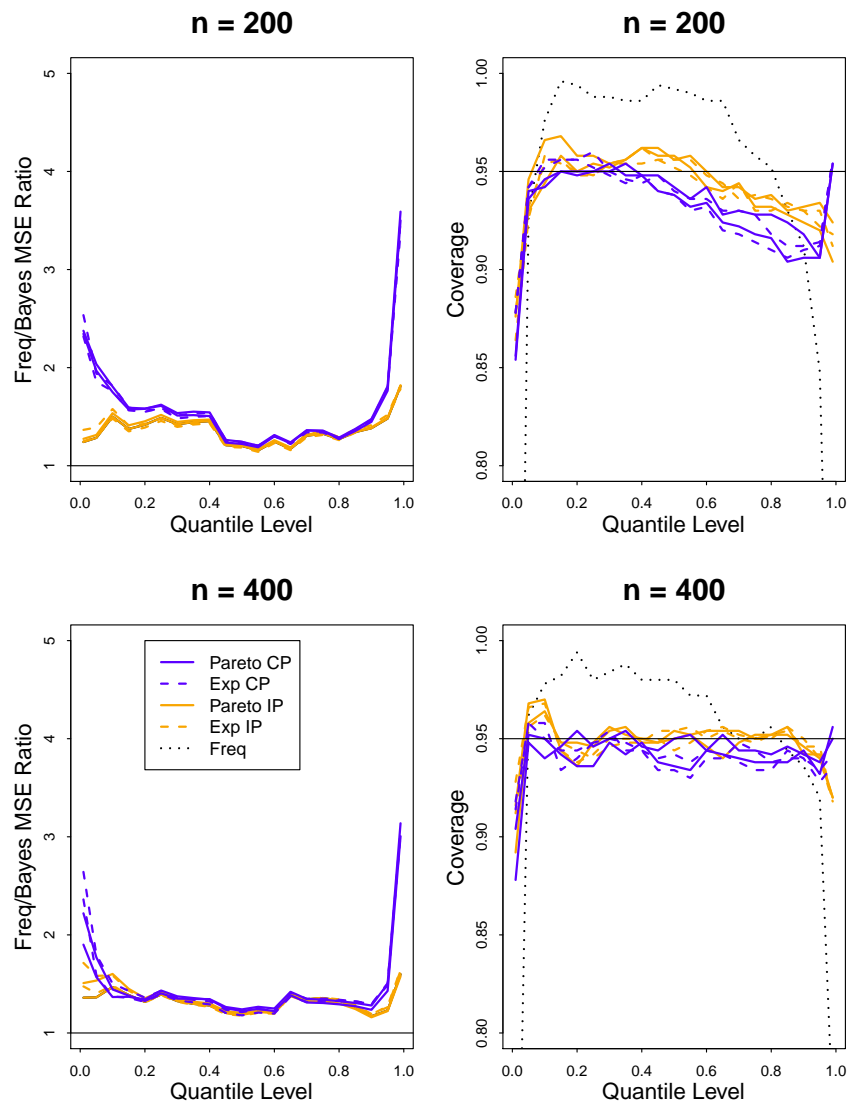


Figure 1. MSE ratios of quantile function estimator over frequentist quantile estimator and coverage probabilities for beta quantile functions. The maximum Monte Carlo standard error of MSE was 0.2 for Bayesian estimators and 0.3 for the frequentist estimator. The maximum Monte Carlo standard error for the MSE ratio was 0.09.

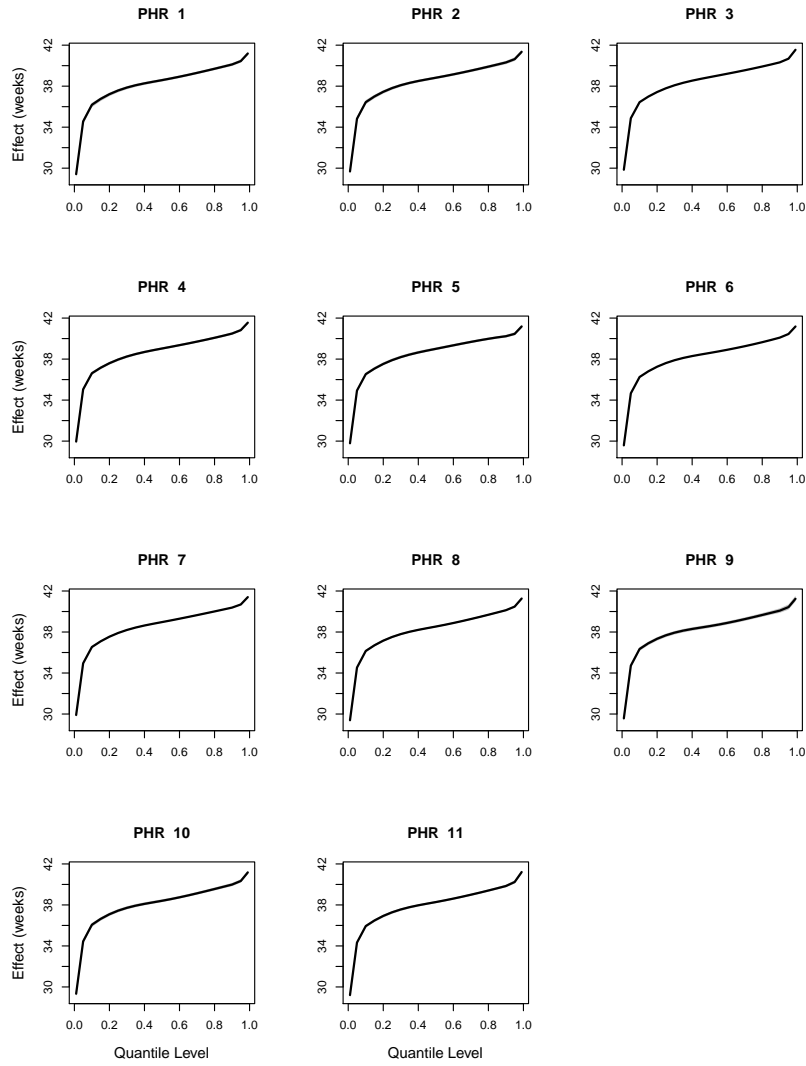


Figure 2. 95% credible limits for the posterior distribution of the intercept process for gestational age by Public Health Region (PHR). The intercept process represents the information regarding gestational age not explained by the predictors and was permitted to vary by PHR. The intercept process is not interpretable, as we had binary variables that were valued at either -1 or 1.

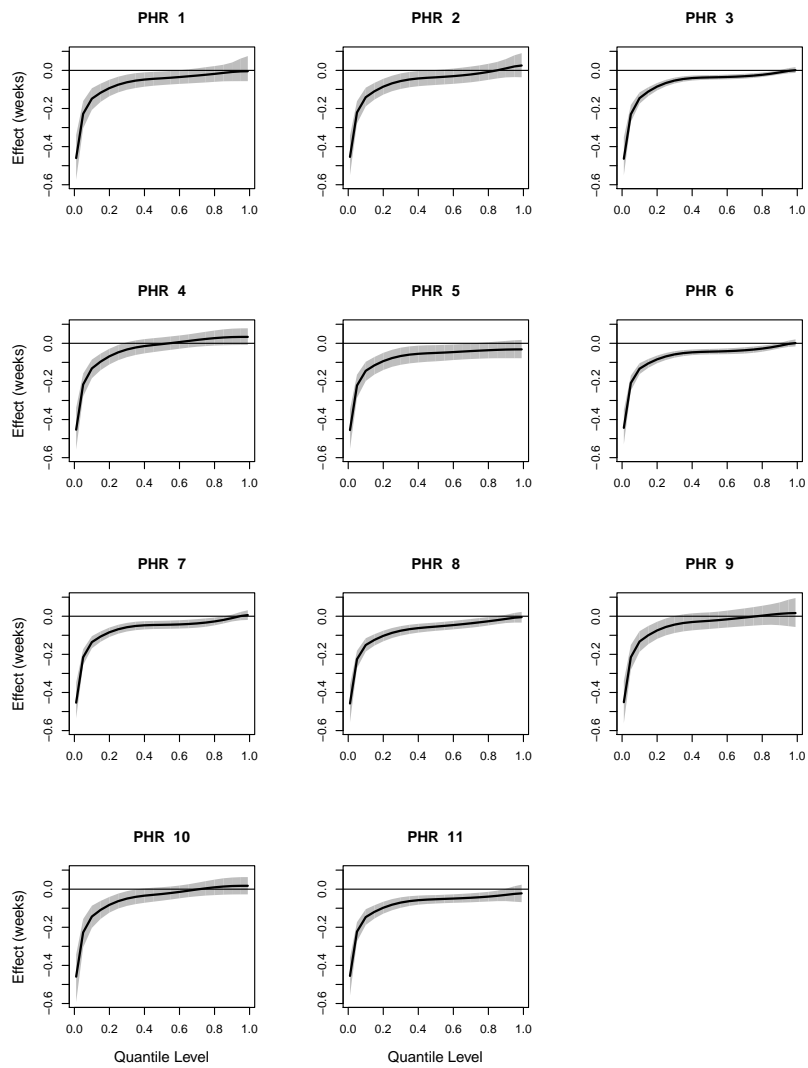


Figure 3. 95% credible limits for the posterior distribution of the effect of being male vs. female on gestational age by Public Health Region (PHR).

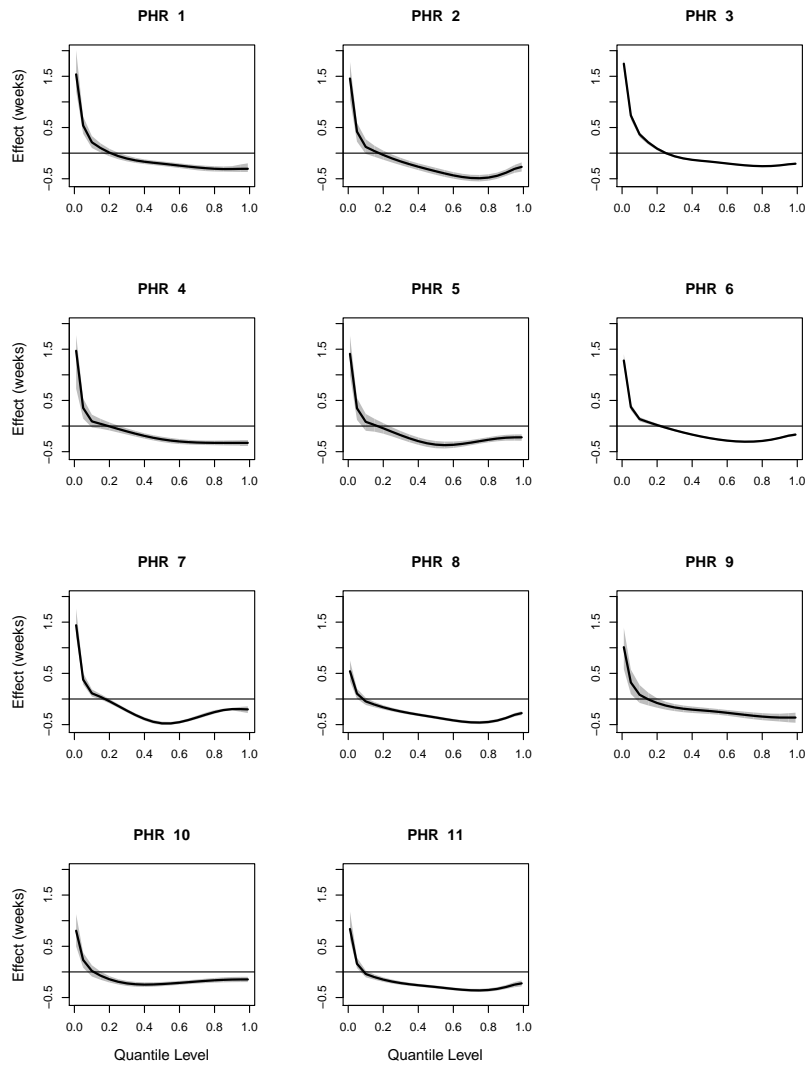


Figure 4. 95% credible limits for the posterior distribution of the effect of maternal parity on gestational age by Public Health Region (PHR).

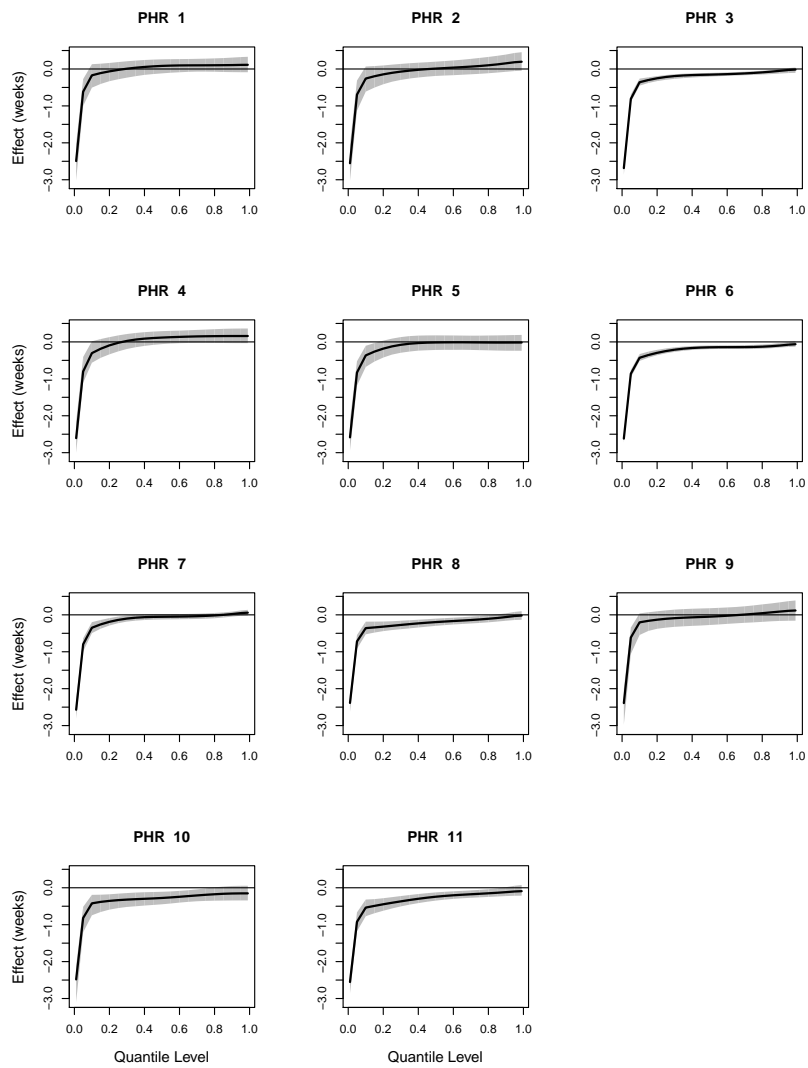


Figure 5. 95% credible limits for the posterior distribution of the effect of maternal age 40 and above on gestational age by Public Health Region (PHR).

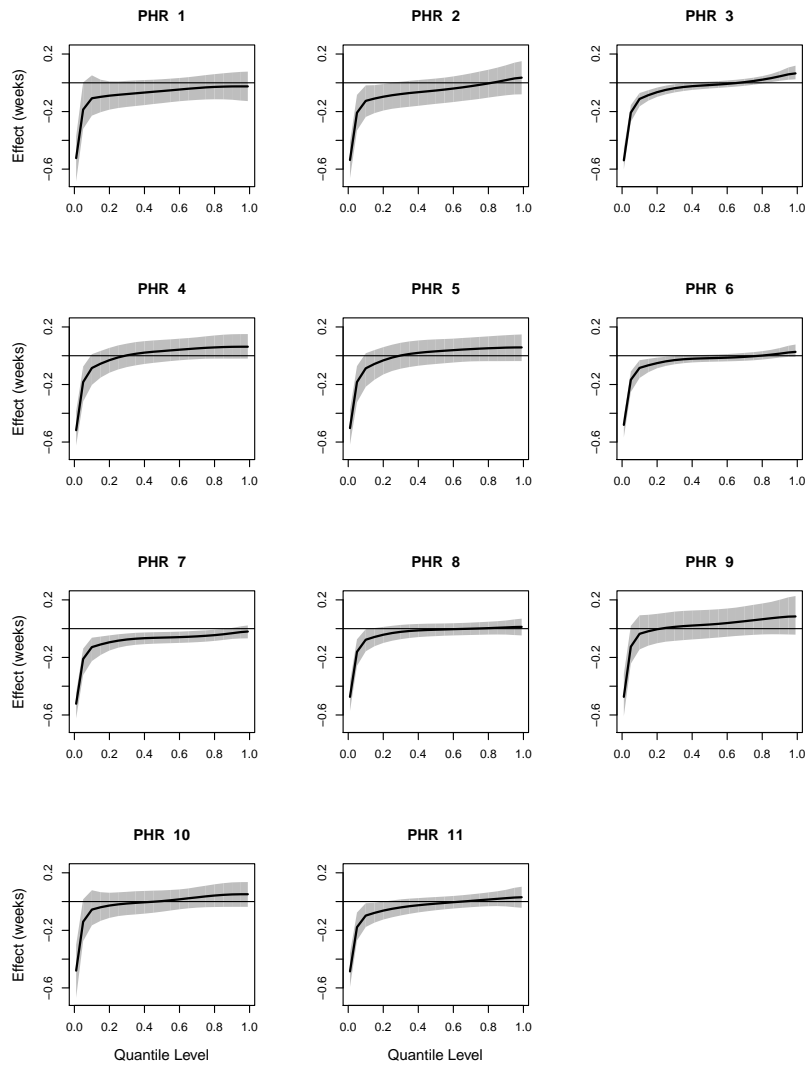


Figure 6. 95% credible limits for the posterior distribution of the effect of paternal age 40 and above on gestational age by Public Health Region (PHR).

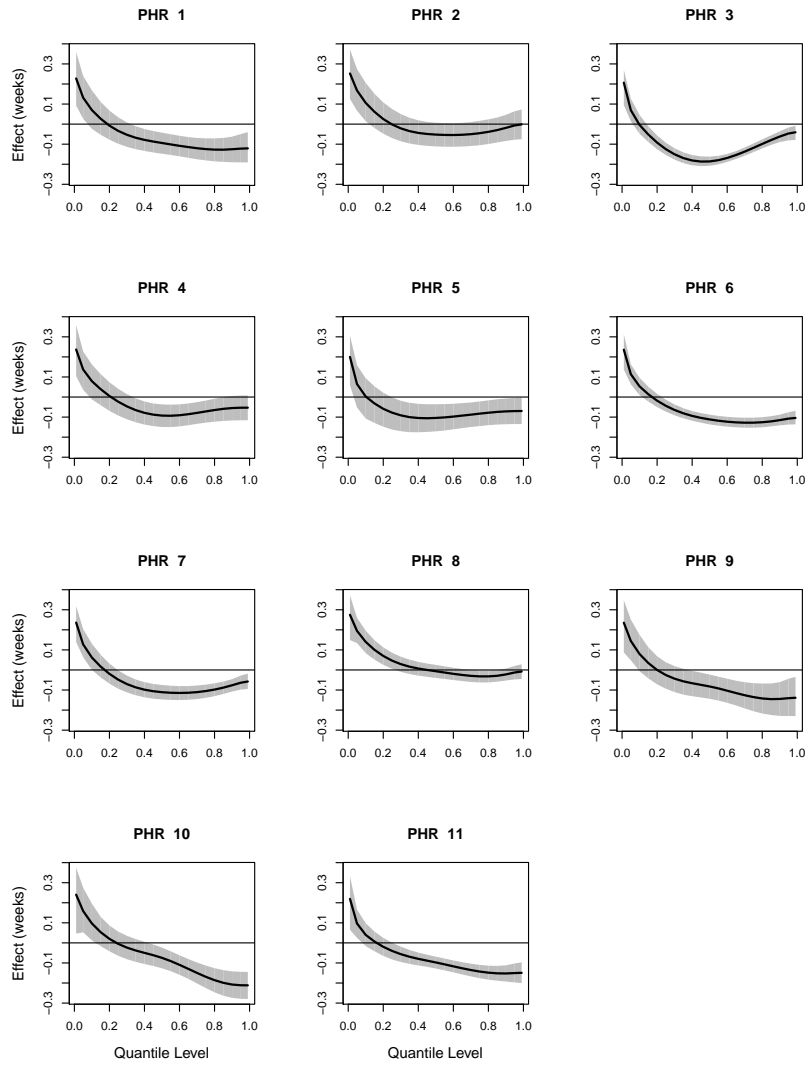


Figure 7. 95% credible limits for the posterior distribution of the effect of the mother finishing high school relative to not finishing high school on gestational age by Public Health Region (PHR).

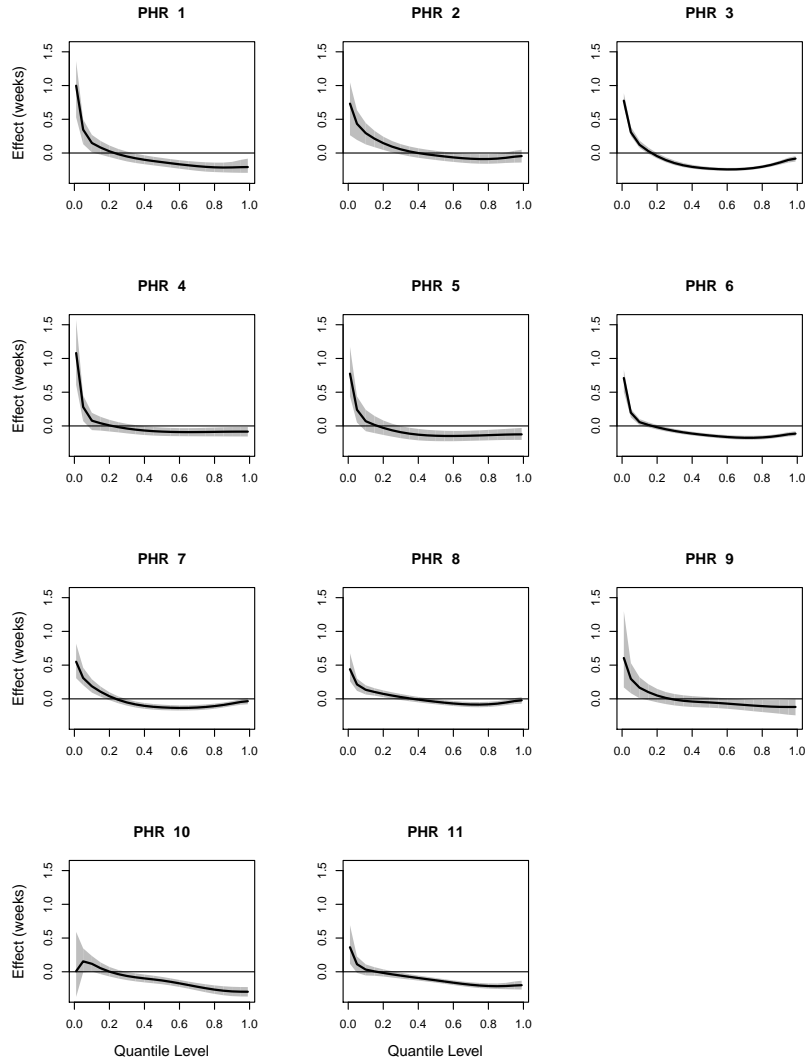


Figure 8. 95% credible limits for the posterior distribution of the effect of the mother finishing education above high school relative to not finishing high school on gestational age by Public Health Region (PHR).

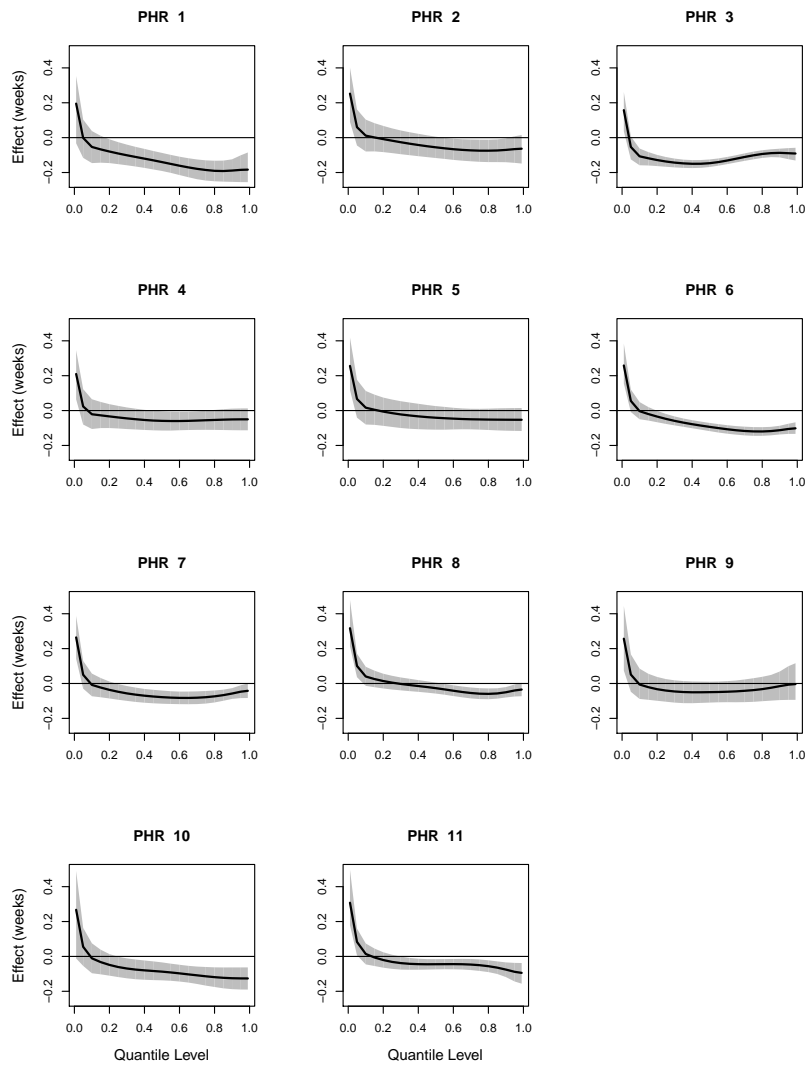


Figure 9. 95% credible limits for the posterior distribution of the effect of the father finishing high school relative to not finishing high school on gestational age by Public Health Region (PHR).

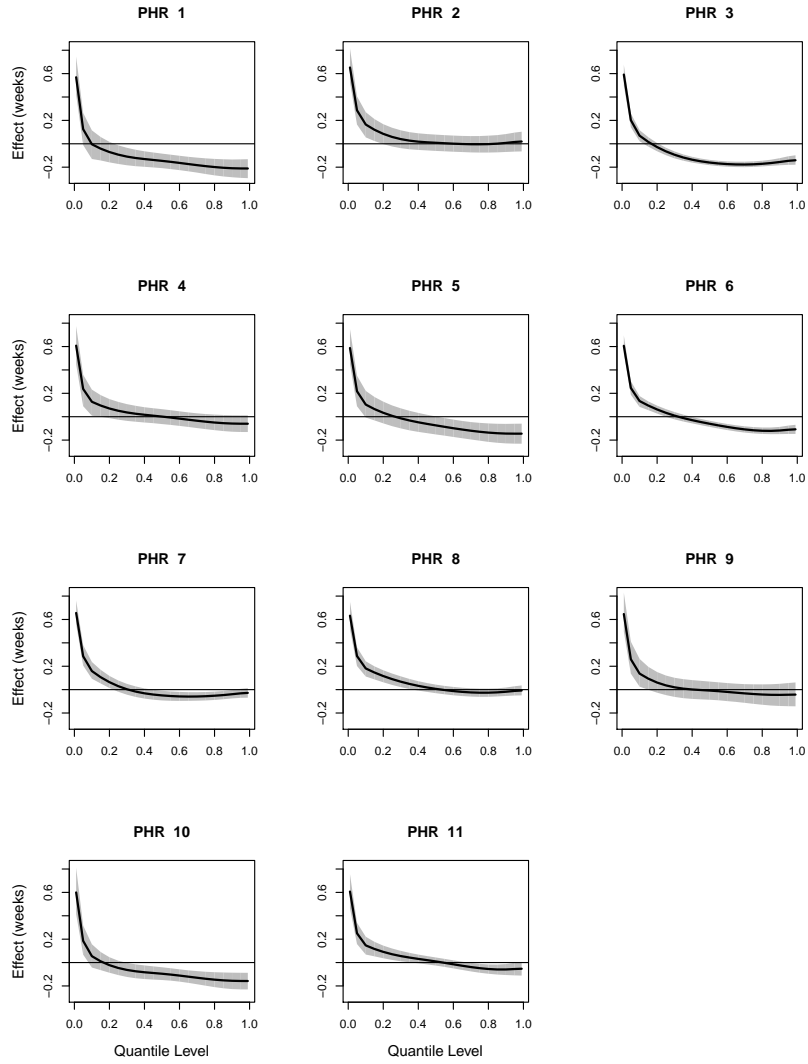


Figure 10. 95% credible limits for the posterior distribution of the effect of the father finishing education above high school relative to not finishing high school on gestational age by Public Health Region (PHR).

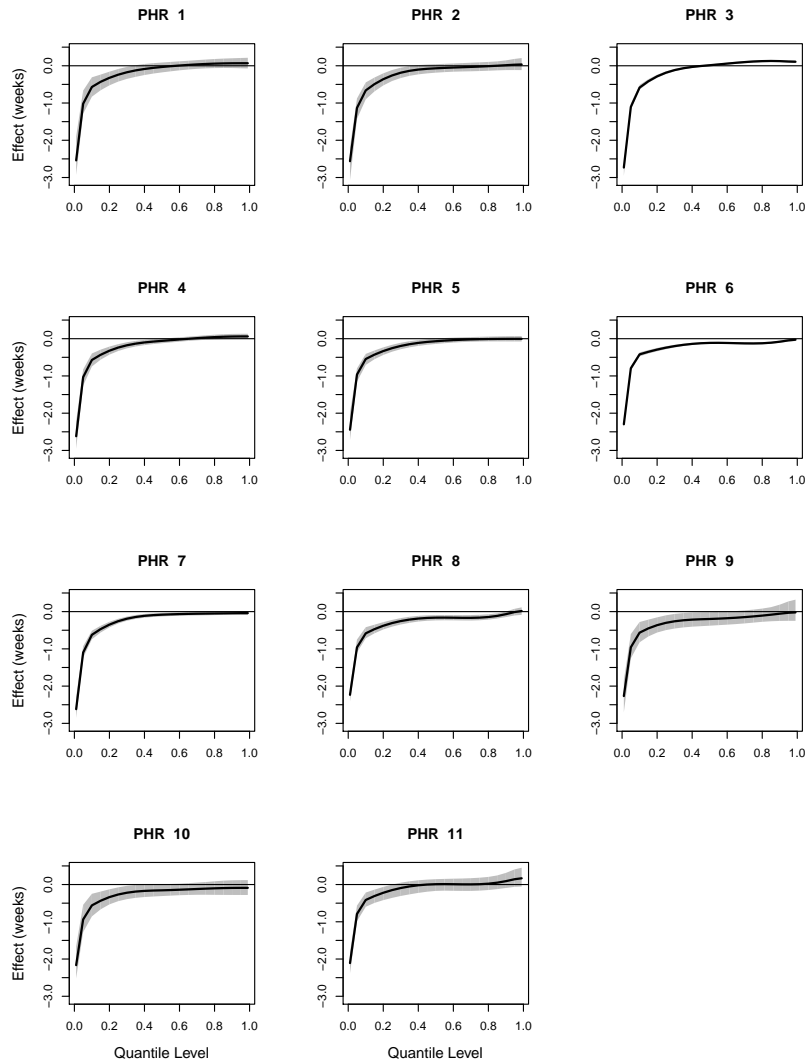


Figure 11. 95% credible limits for the posterior distribution of the effect of black non-Hispanic maternal ethnicity relative to white non-Hispanic maternal ethnicity on gestational age by Public Health Region (PHR).

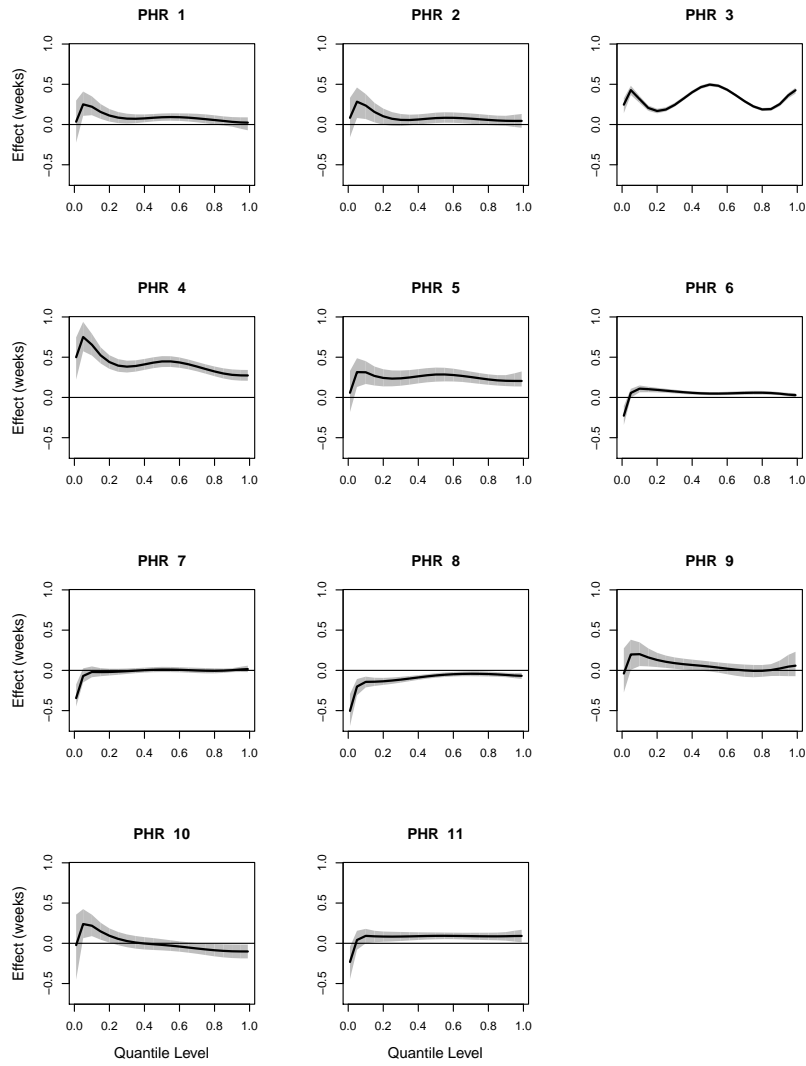


Figure 12. 95% credible limits for the posterior distribution of the effect of Hispanic maternal ethnicity relative to white non-Hispanic maternal ethnicity on gestational age by Public Health Region (PHR).

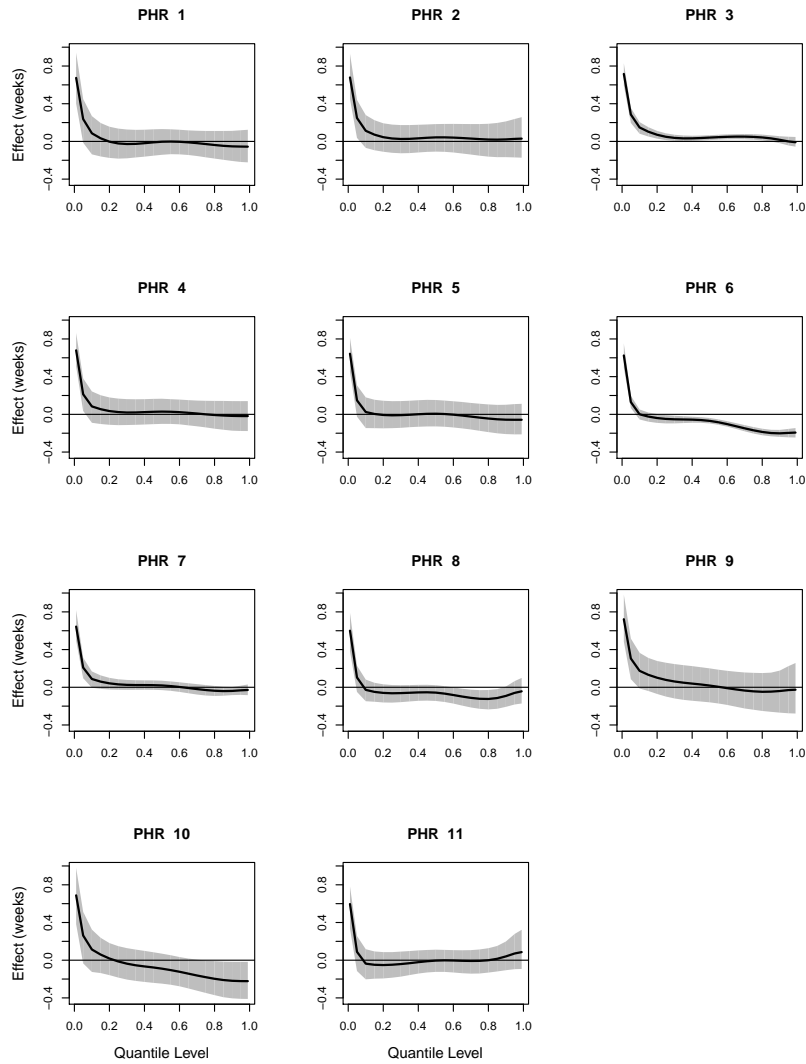


Figure 13. 95% credible limits for the posterior distribution of the effect of other maternal ethnicity relative to white non-Hispanic maternal ethnicity on gestational age by Public Health Region (PHR).

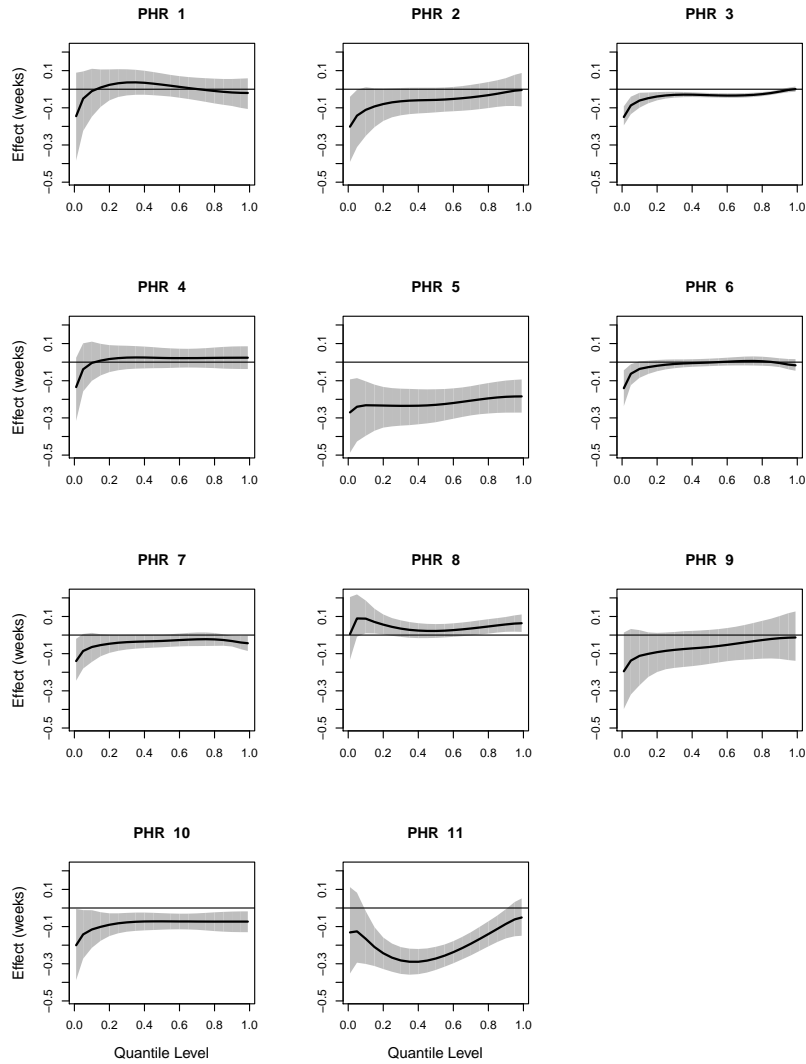


Figure 14. 95% credible limits for the posterior distribution of the effect of a one-unit increase in first trimester ozone exposure on gestational age by Public Health Region (PHR). All ozone values were linearly transformed into $[-1,1]$, so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.

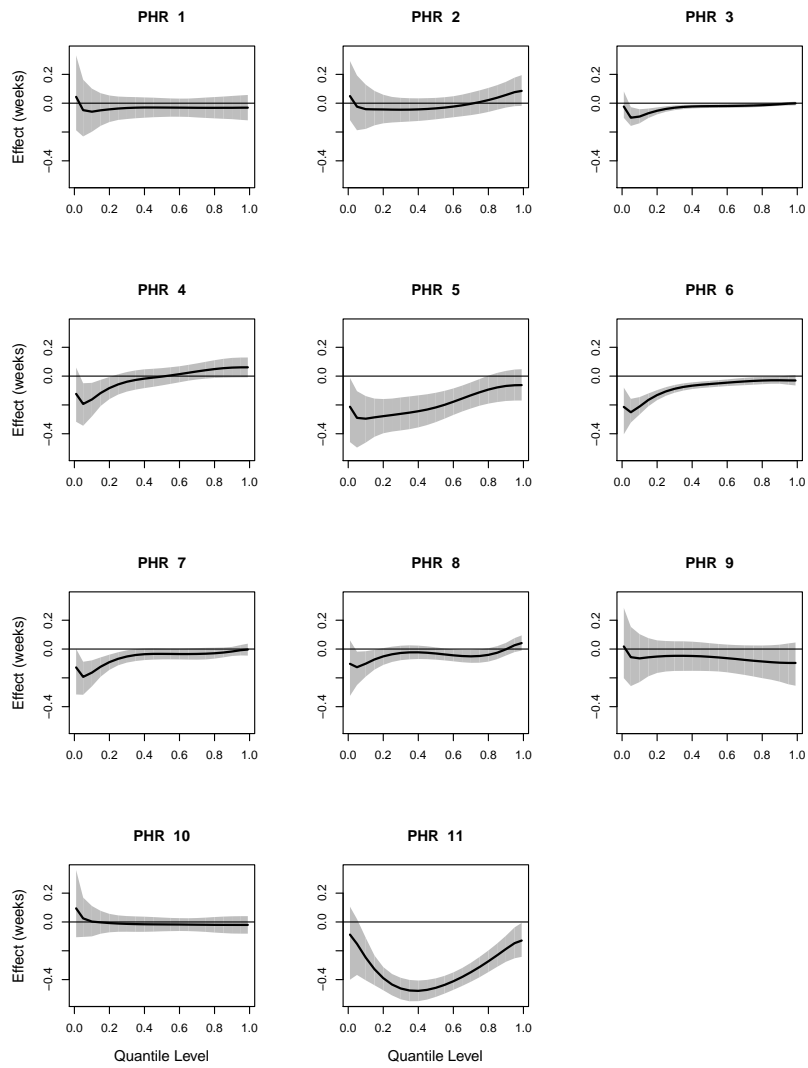


Figure 15. 95% credible limits for the posterior distribution of the effect of a one-unit increase in second trimester ozone exposure on gestational age by Public Health Region (PHR). All ozone values were linearly transformed into $[-1,1]$, so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.

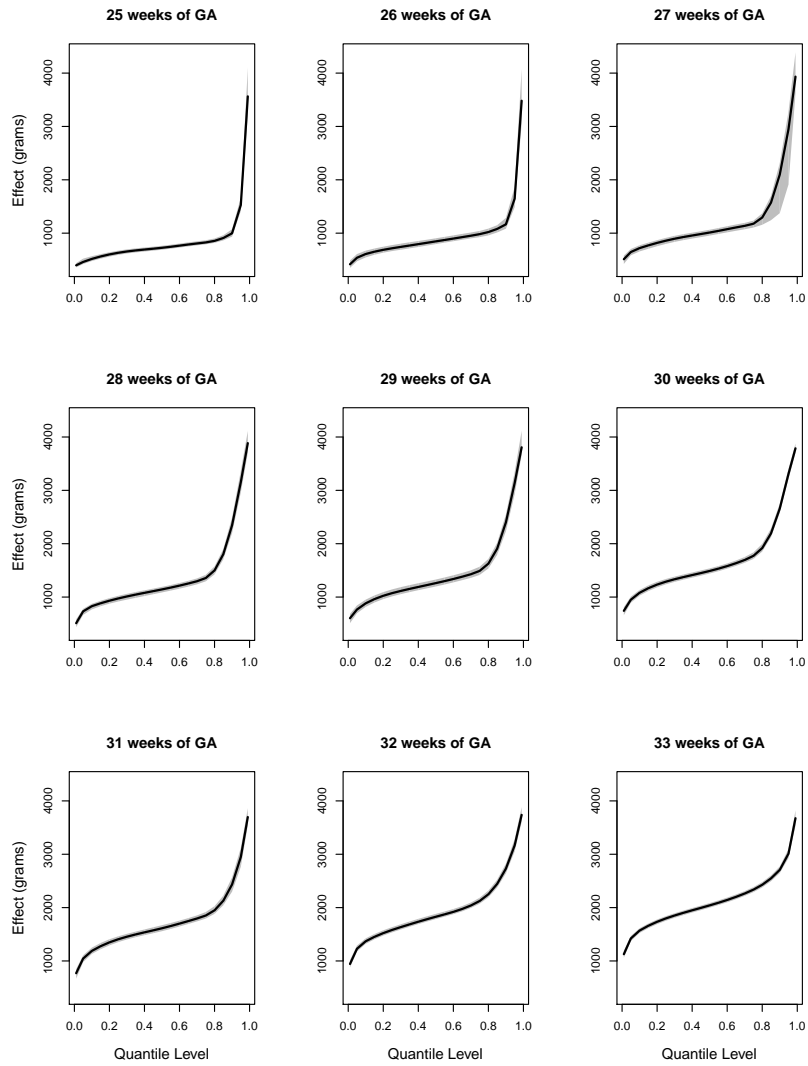


Figure 16. 95% credible limits for the posterior distribution of the intercept of birth weight for gestational ages 25-33 weeks.

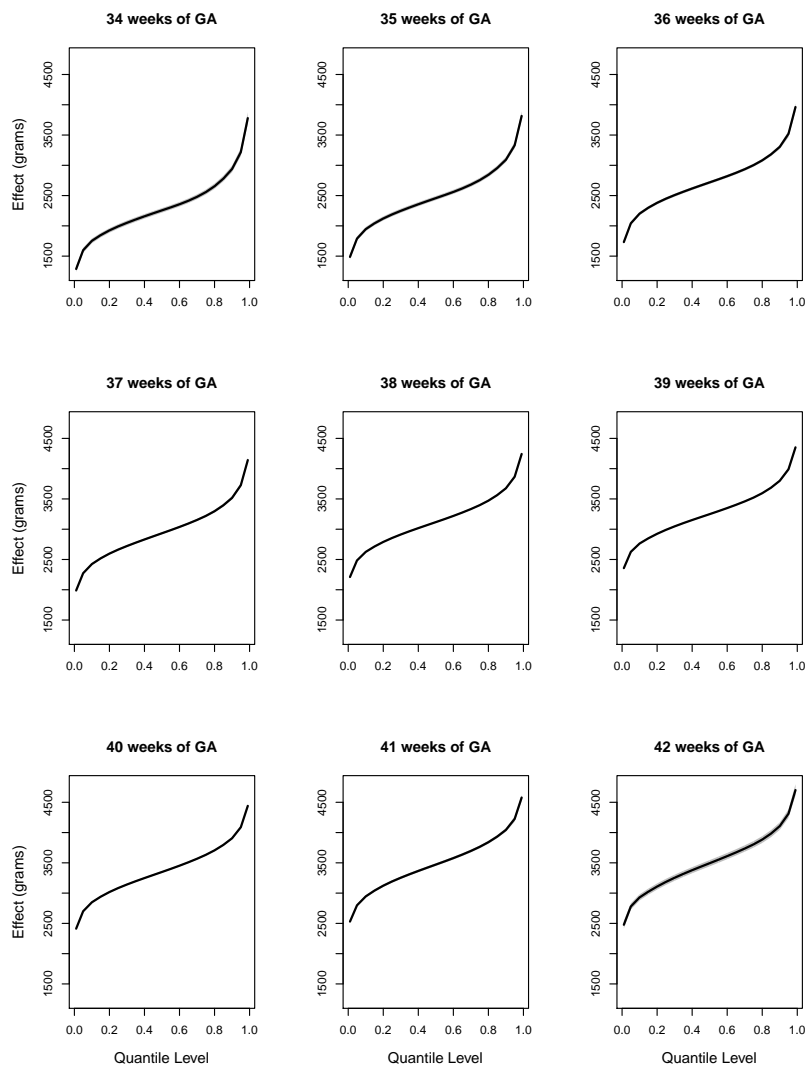


Figure 17. 95% credible limits for the posterior distribution of the intercept of birth weight for gestational ages 34-42 weeks.

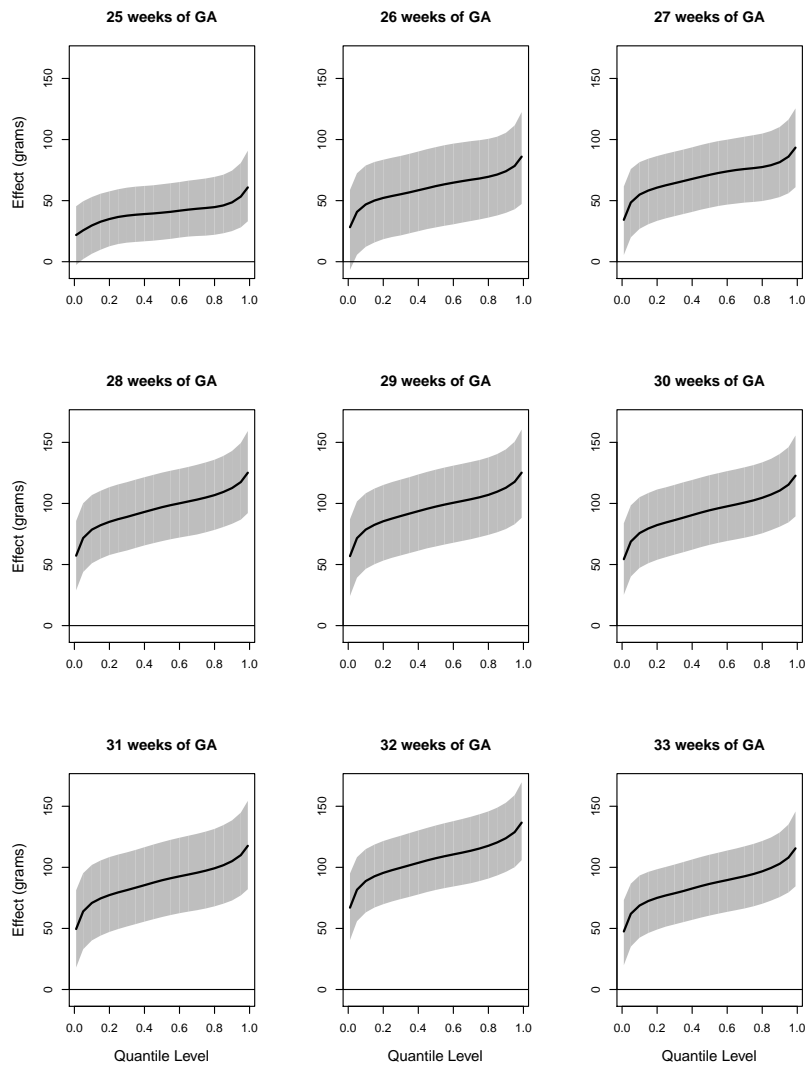


Figure 18. 95% credible limits for the posterior distribution of the effect of male sex relative to female on birth weight for gestational ages 25-33 weeks.

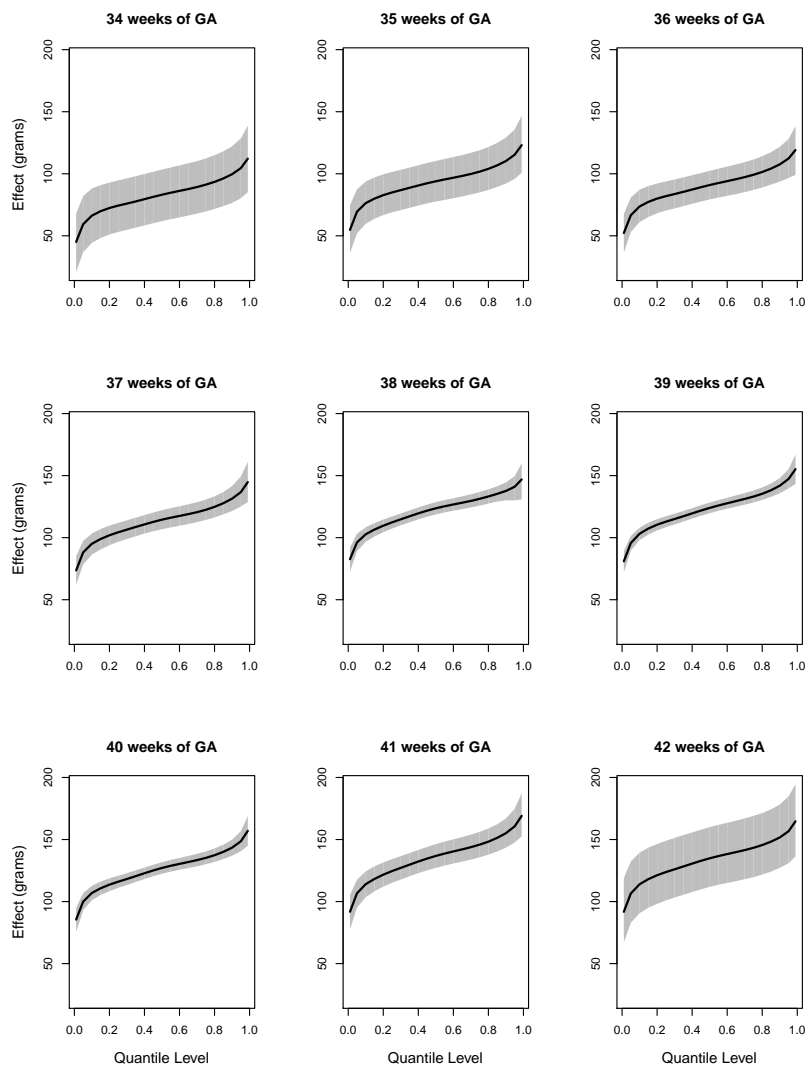


Figure 19. 95% credible limits for the posterior distribution of the effect of male sex relative to female on birth weight for gestational ages 34-42 weeks.

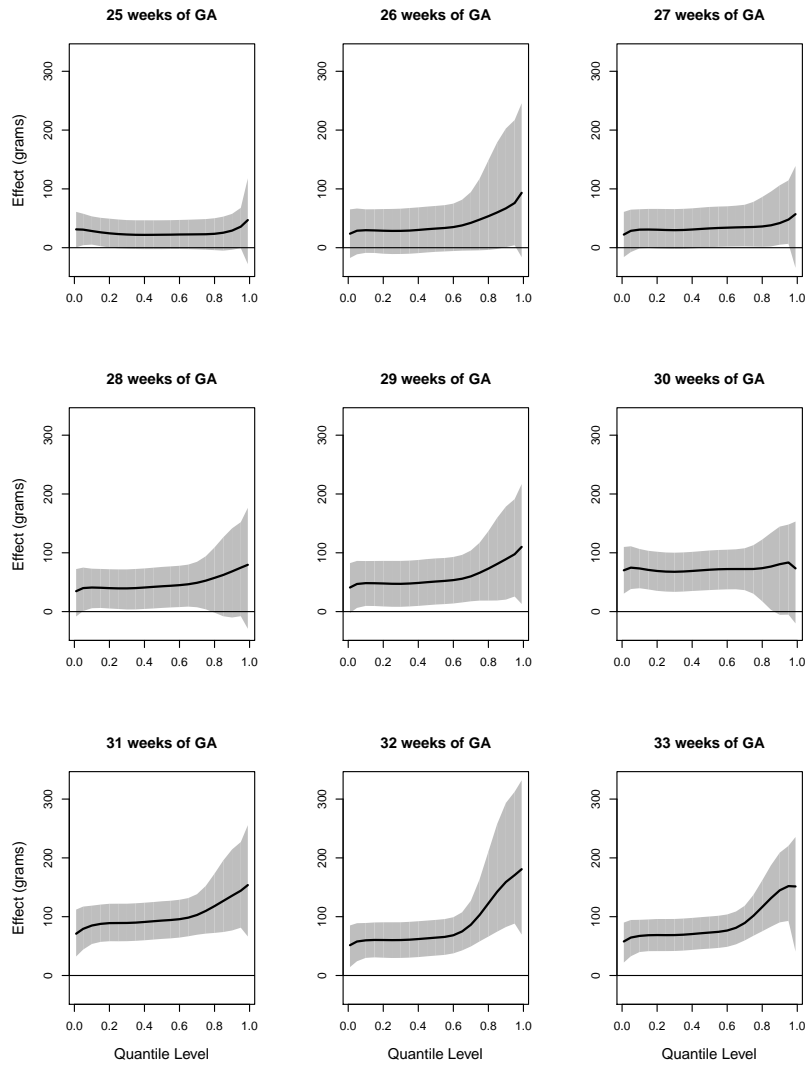


Figure 20. 95% credible limits for the posterior distribution of the effect of maternal parity on birth weight for gestational ages 25-33 weeks.

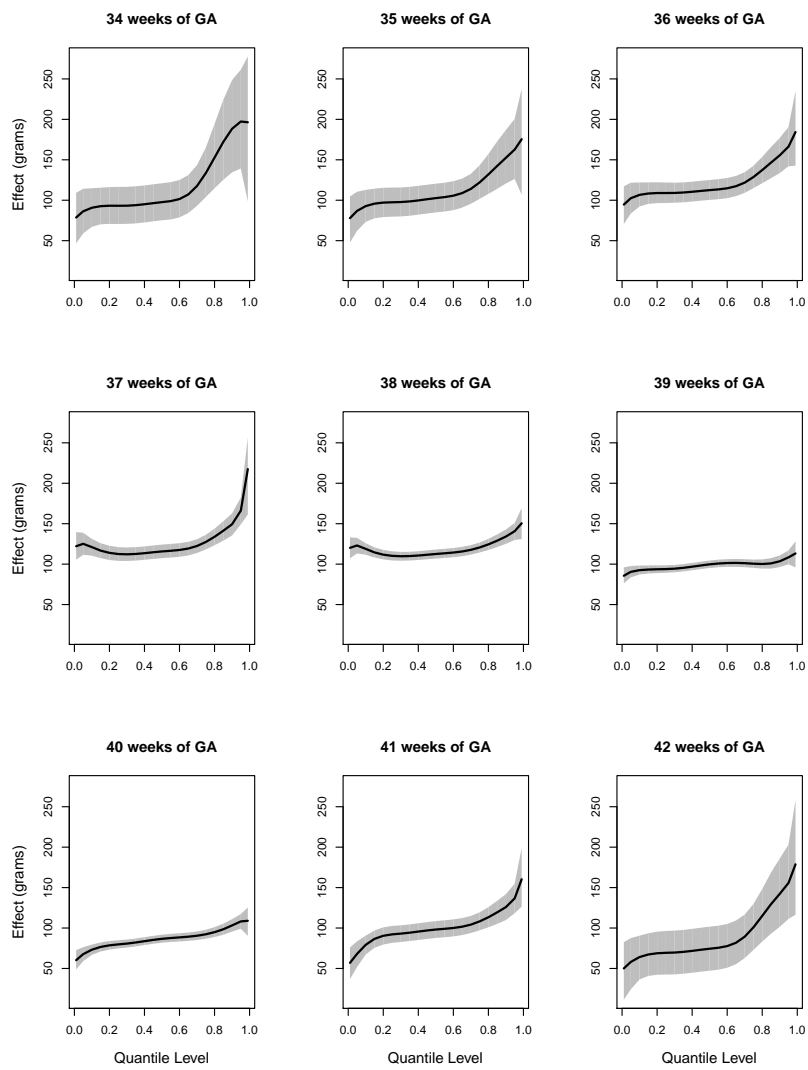


Figure 21. 95% credible limits for the posterior distribution of the effect of maternal parity on birth weight for gestational ages 34-42 weeks.

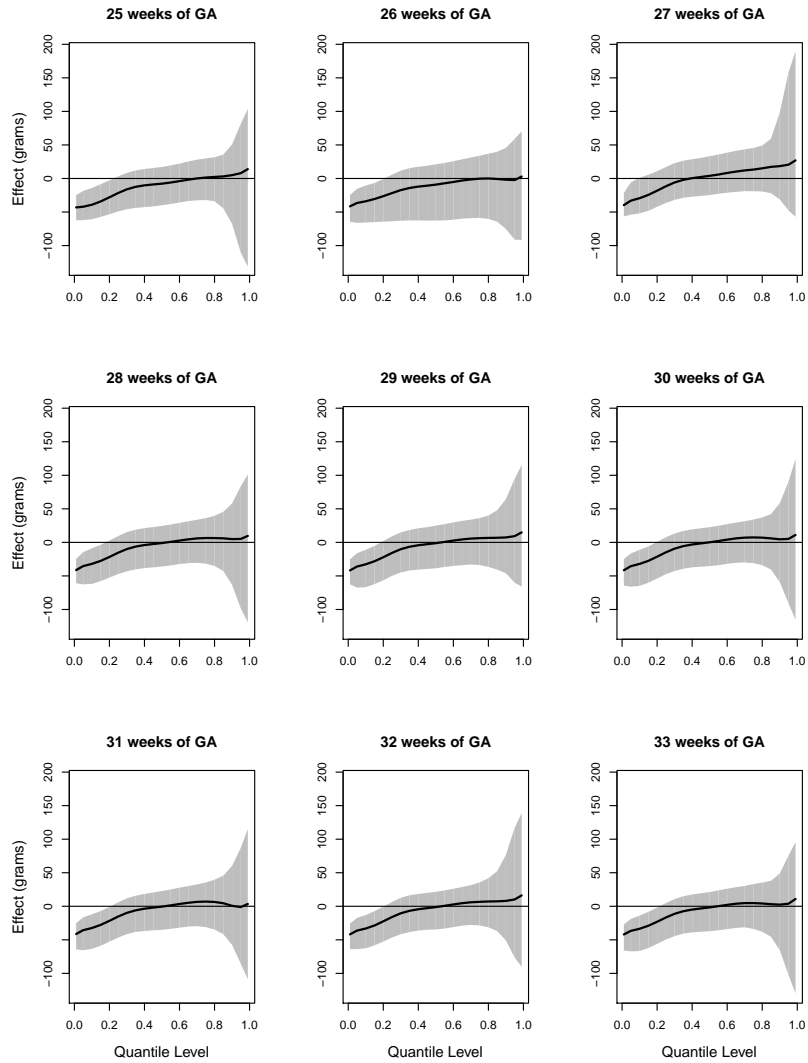


Figure 22. 95% credible limits for the posterior distribution of the effect of maternal age greater than 40 relative to maternal age less than 40 on birth weight for gestational ages 25-33 weeks.

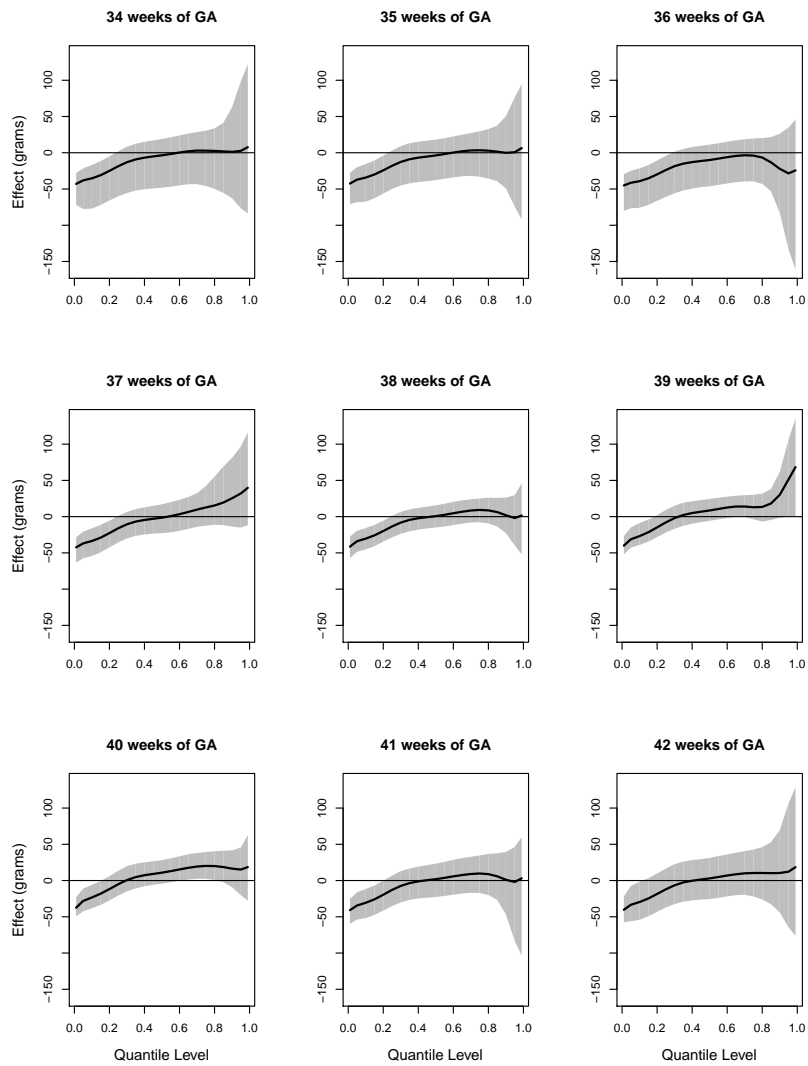


Figure 23. 95% credible limits for the posterior distribution of the effect of maternal age greater than 40 relative to maternal age less than 40 on birth weight for gestational ages 34-42 weeks.

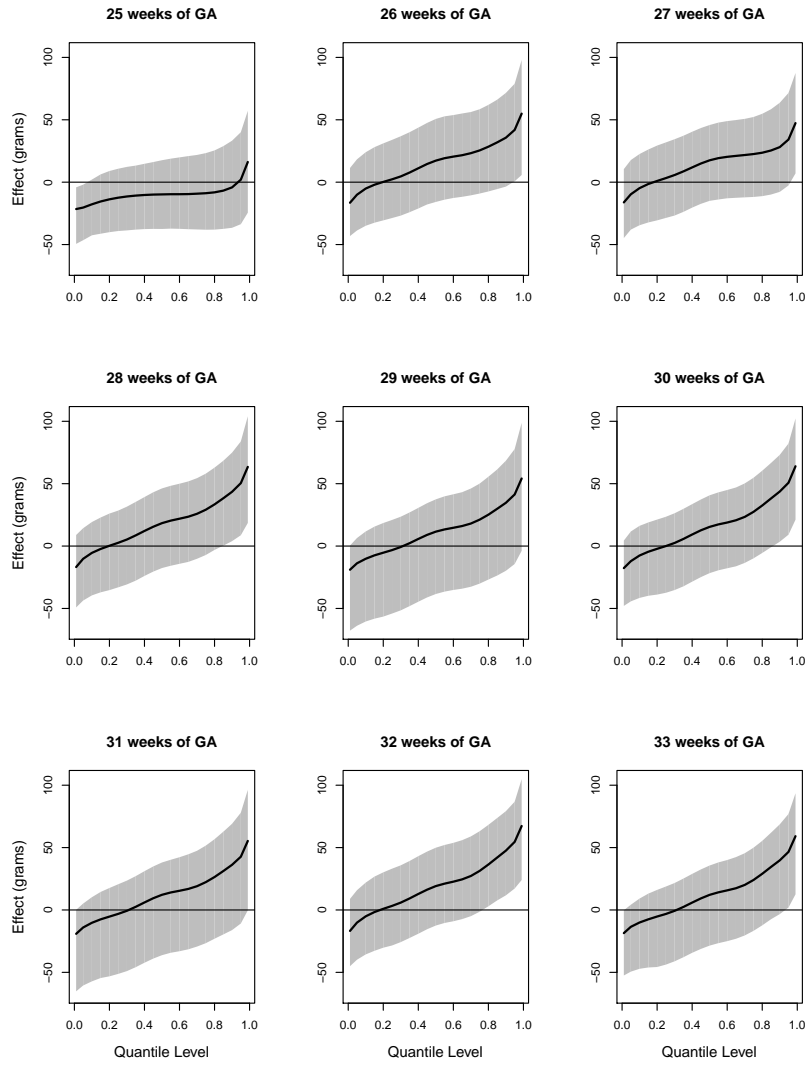


Figure 24. 95% credible limits for the posterior distribution of the effect of paternal age greater than 40 relative to paternal age less than 40 on birth weight for gestational ages 25-33 weeks.

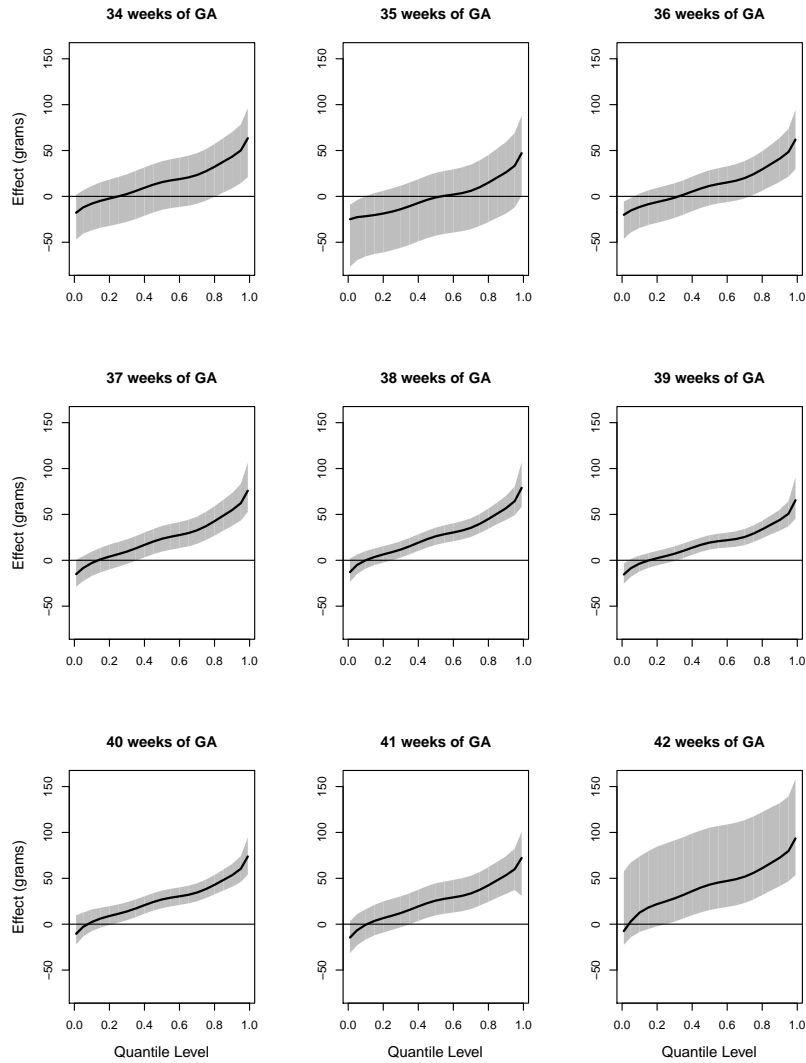


Figure 25. 95% credible limits for the posterior distribution of the effect of paternal age greater than 40 relative to paternal age less than 40 on birth weight for gestational ages 34-42 weeks.

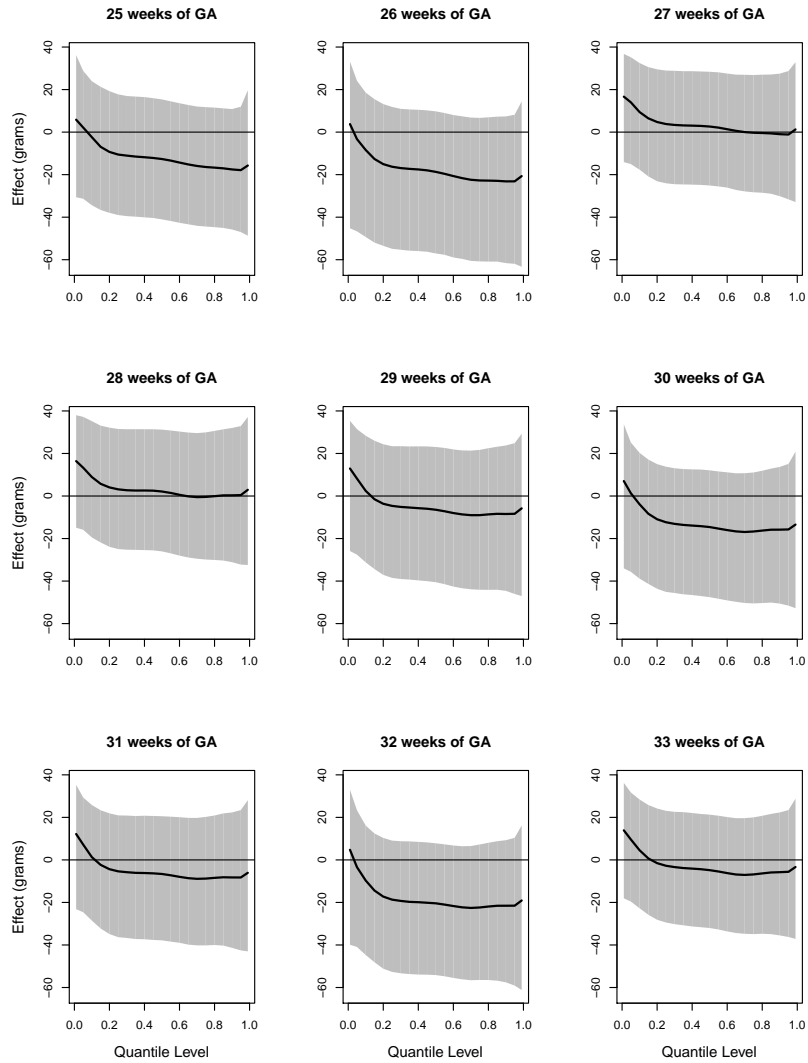


Figure 26. 95% credible limits for the posterior distribution of the effect of the mother finishing high school relative to not finishing high school on birth weight for gestational ages 25-33 weeks.

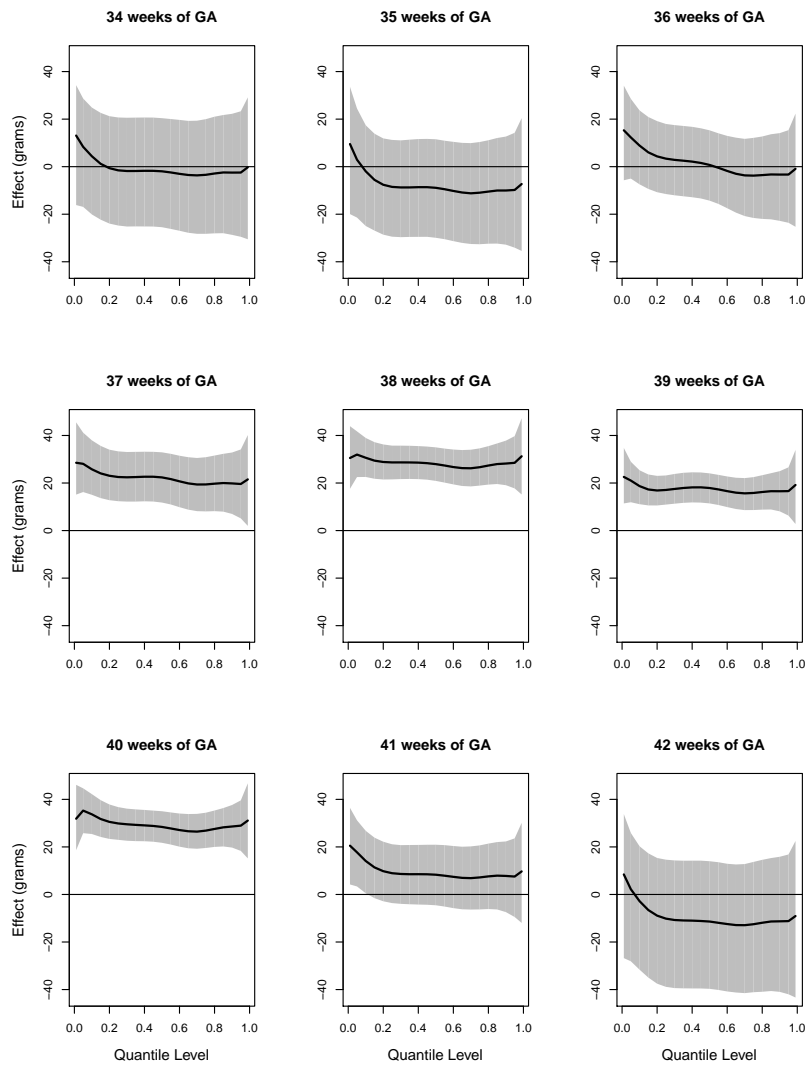


Figure 27. 95% credible limits for the posterior distribution of the effect of the mother finishing high school relative to not finishing high school on birth weight for gestational ages 34-42 weeks.

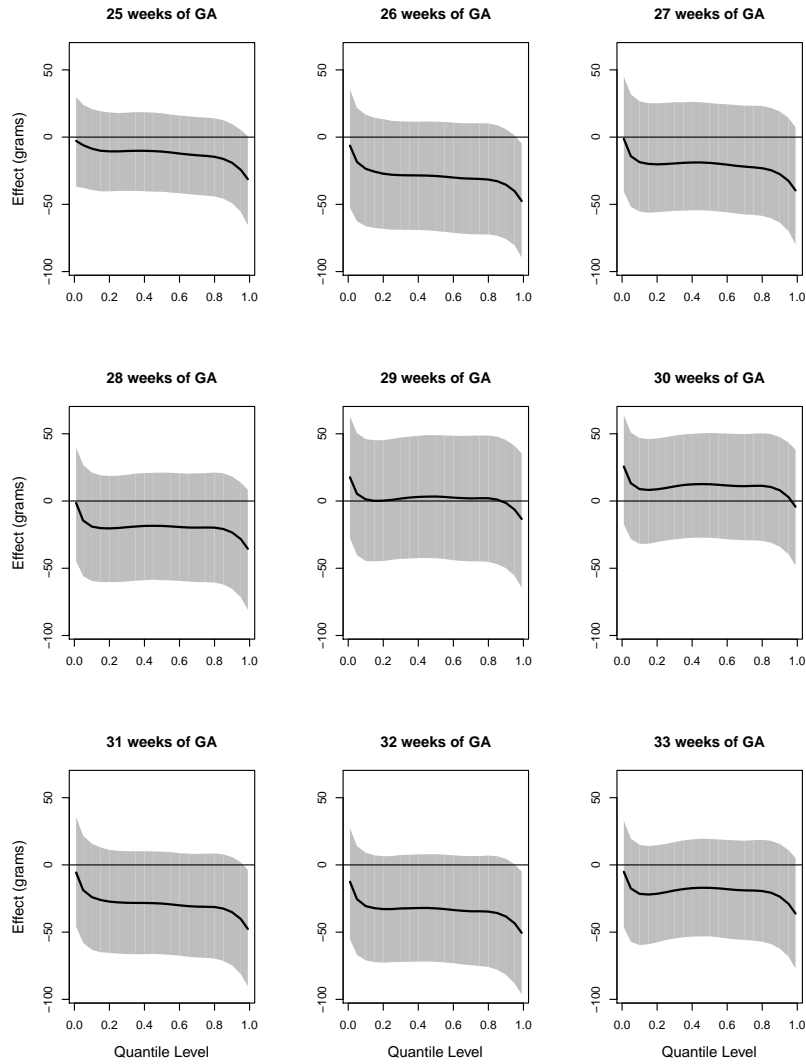


Figure 28. 95% credible limits for the posterior distribution of the effect of the mother finishing education above high school relative to not finishing high school on birth weight for gestational ages 25-33 weeks.

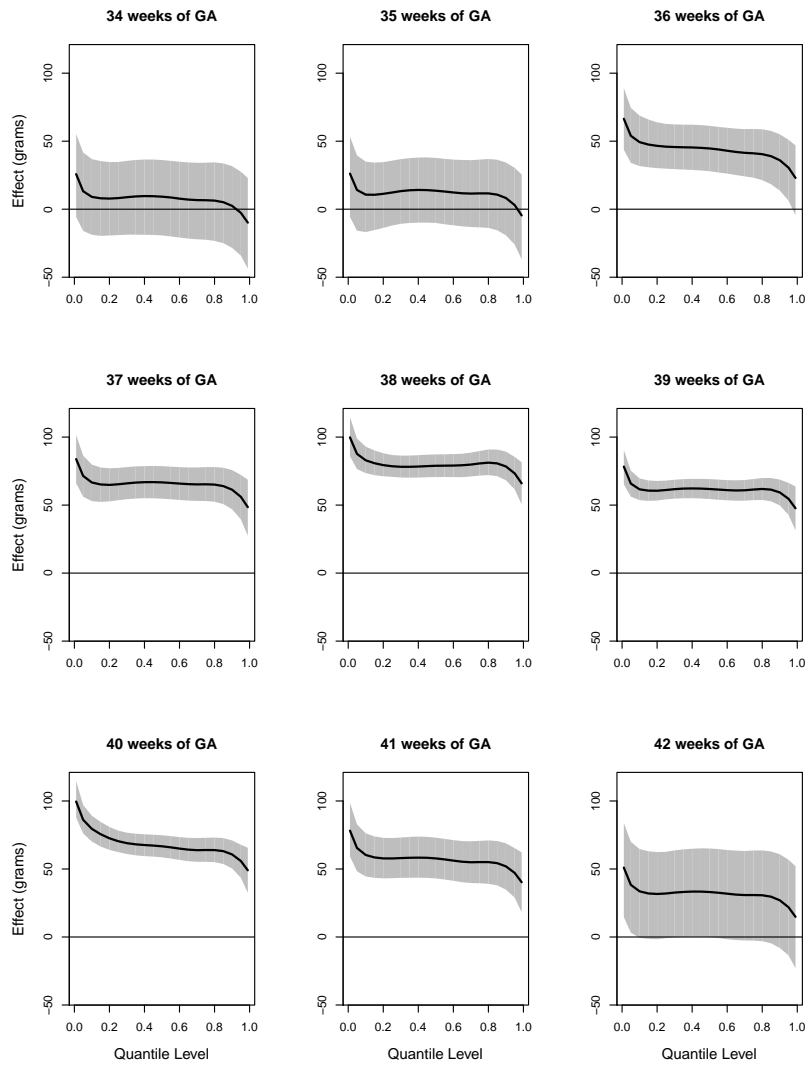


Figure 29. 95% credible limits for the posterior distribution of the effect of the mother finishing education above high school relative to not finishing high school on birth weight for gestational ages 34-42 weeks.

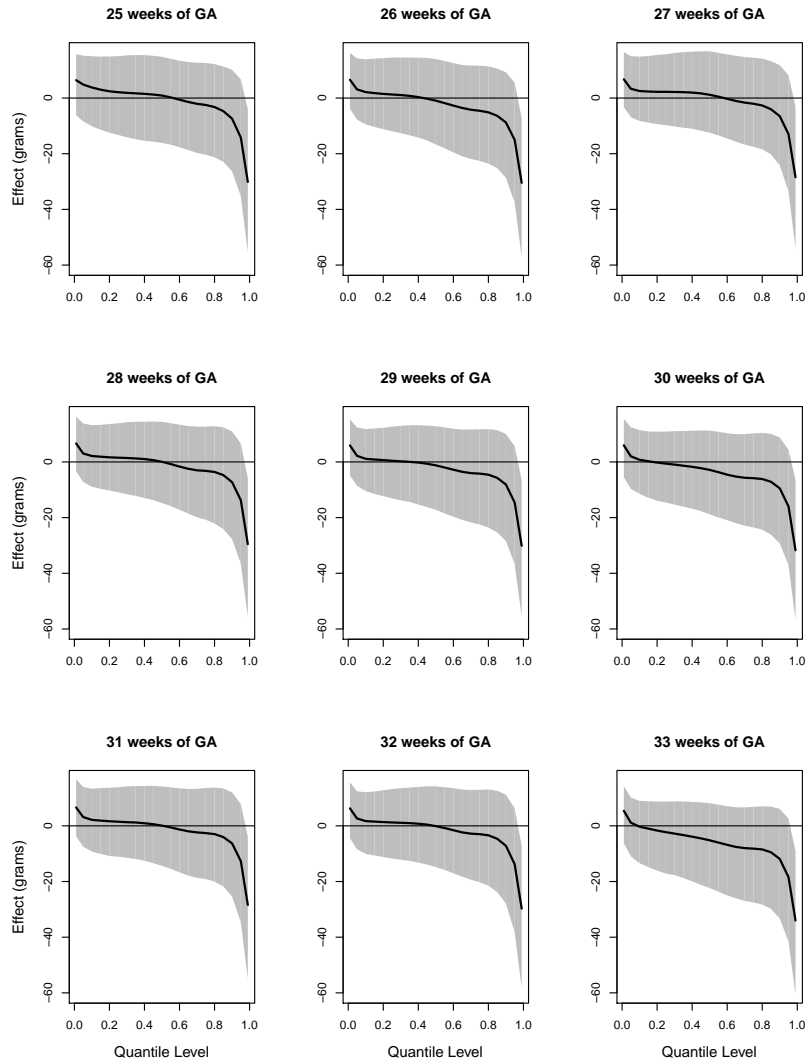


Figure 30. 95% credible limits for the posterior distribution of the effect of the father finishing high school relative to not finishing high school on birth weight for gestational ages 25-33 weeks.

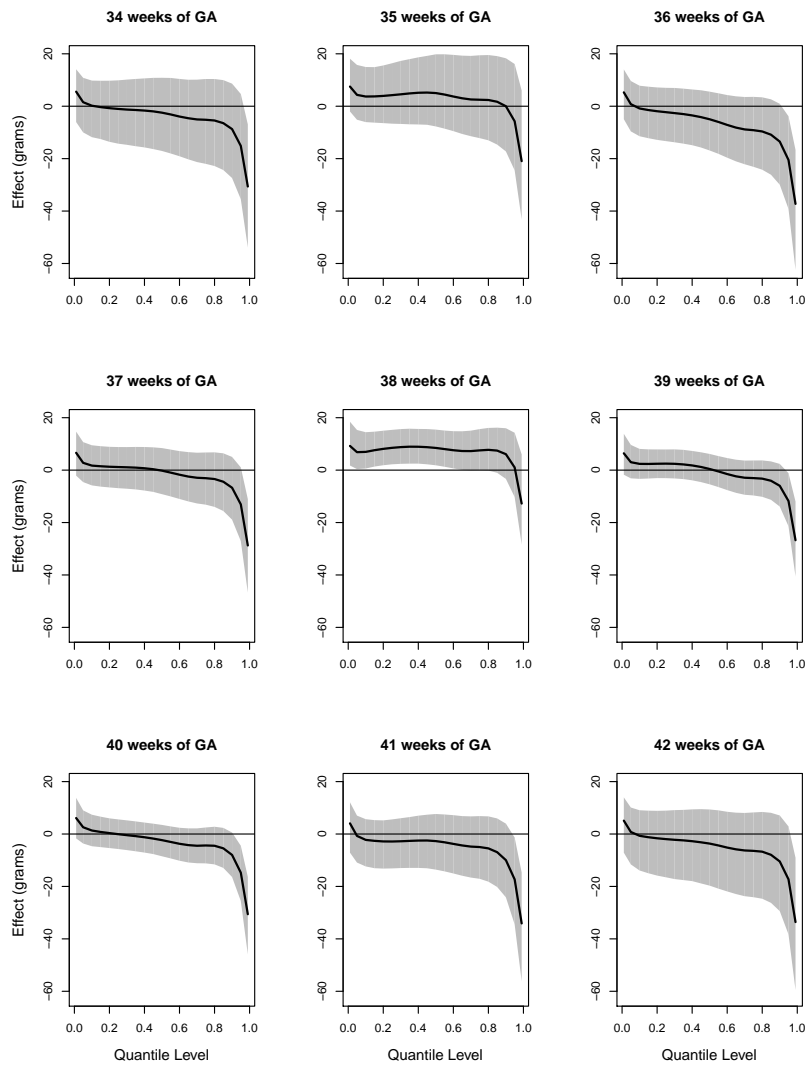


Figure 31. 95% credible limits for the posterior distribution of the effect of the father finishing high school relative to not finishing high school on birth weight for gestational ages 34-42 weeks.

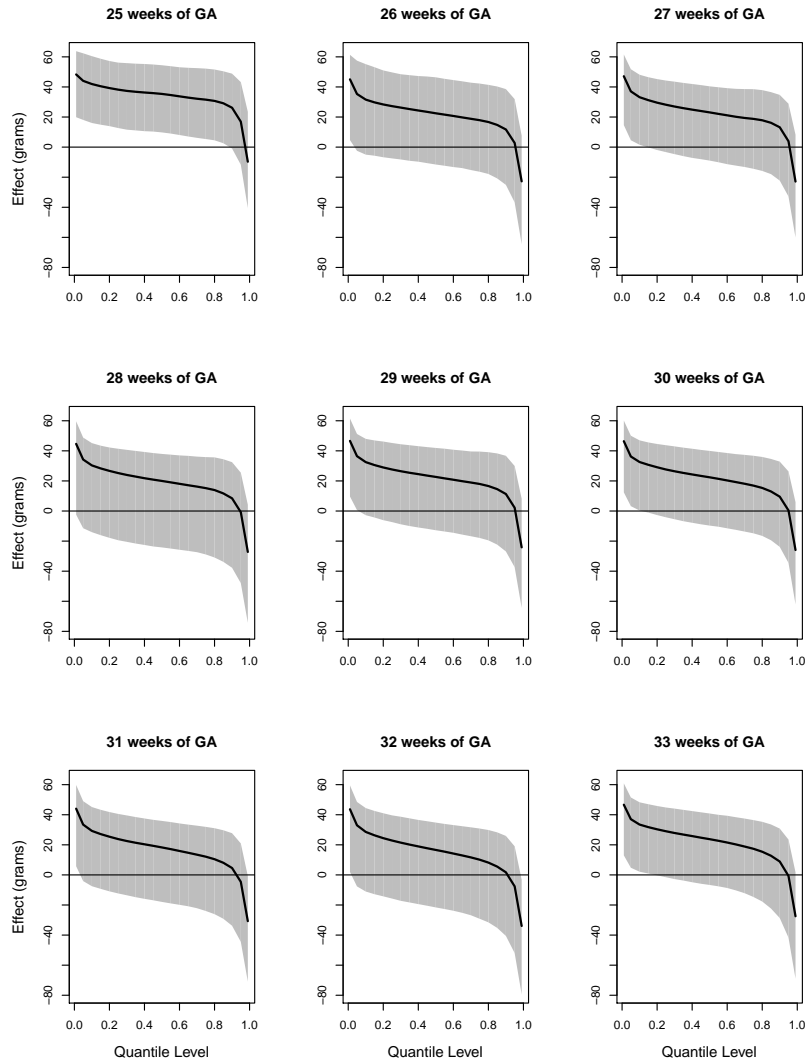


Figure 32. 95% credible limits for the posterior distribution of the effect of the father finishing education above high school relative to not finishing high school on birth weight for gestational ages 25-33 weeks.

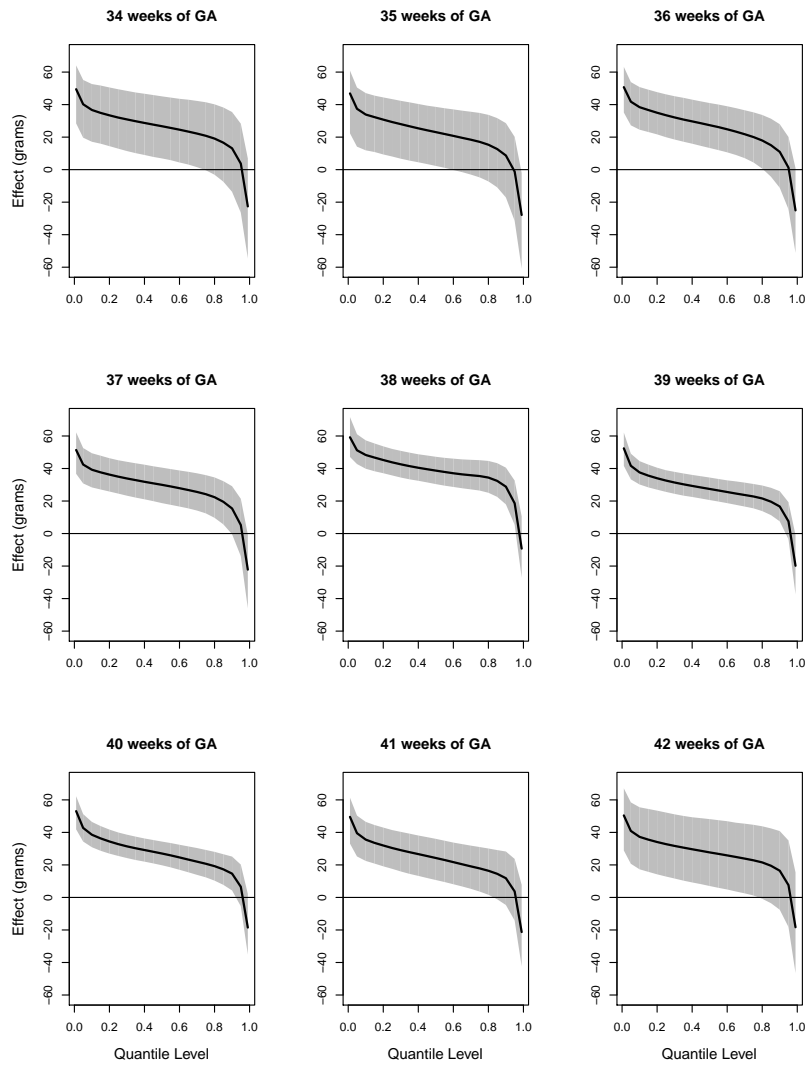


Figure 33. 95% credible limits for the posterior distribution of the effect of the father finishing education above high school relative to not finishing high school on birth weight for gestational ages 34-42 weeks.

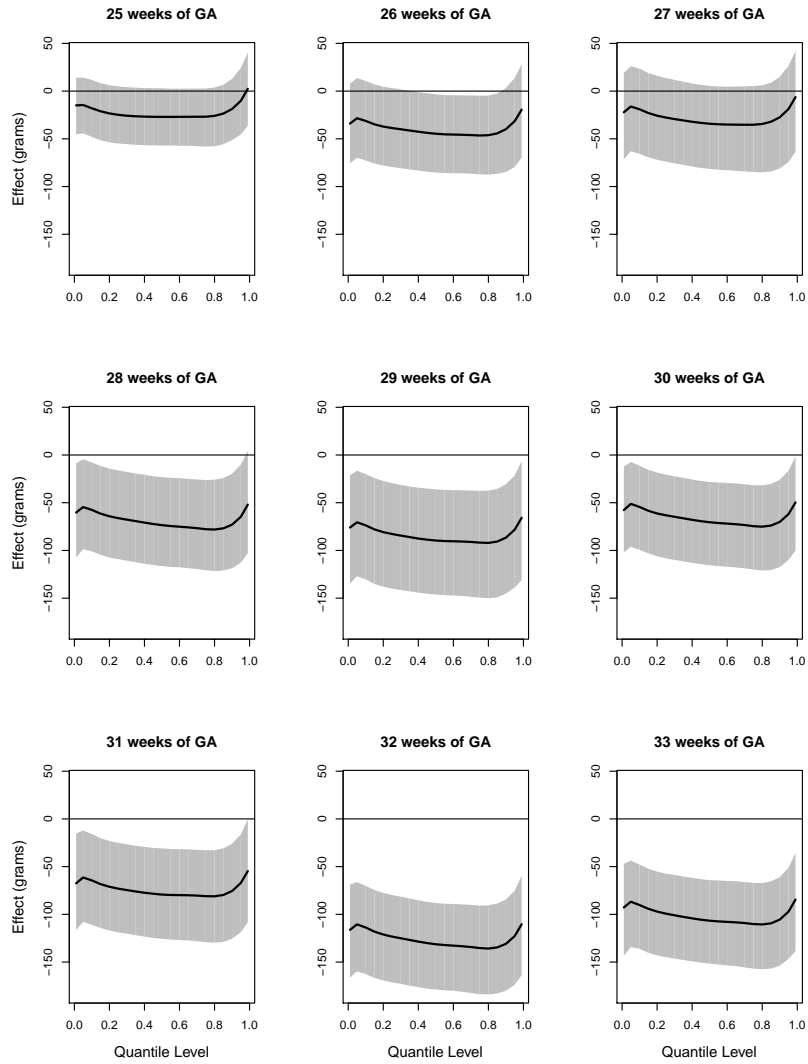


Figure 34. 95% credible limits for the posterior distribution of the effect of maternal black non-Hispanic ethnicity relative to white non-Hispanic ethnicity on birth weight for gestational ages 25-33 weeks.

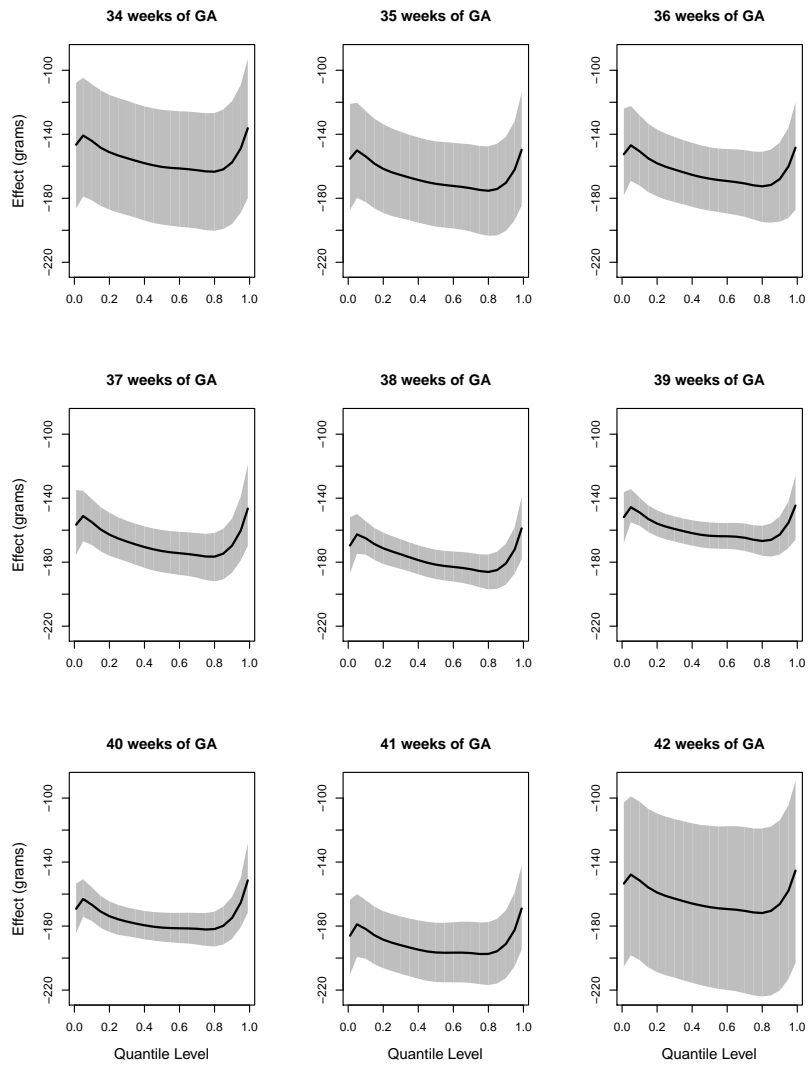


Figure 35. 95% credible limits for the posterior distribution of the effect of maternal black non-Hispanic ethnicity relative to white non-Hispanic ethnicity on birth weight for gestational ages 34-42 weeks.

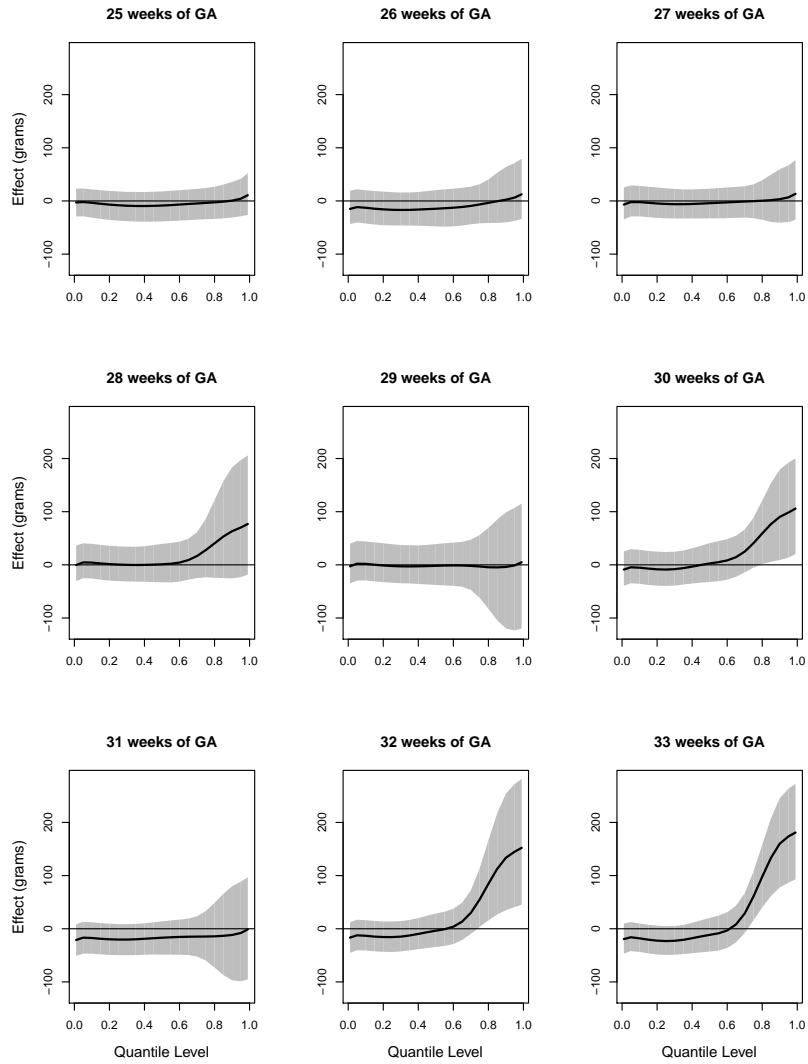


Figure 36. 95% credible limits for the posterior distribution of the effect of maternal Hispanic ethnicity relative to white non-Hispanic ethnicity on birth weight for gestational ages 25-33 weeks.

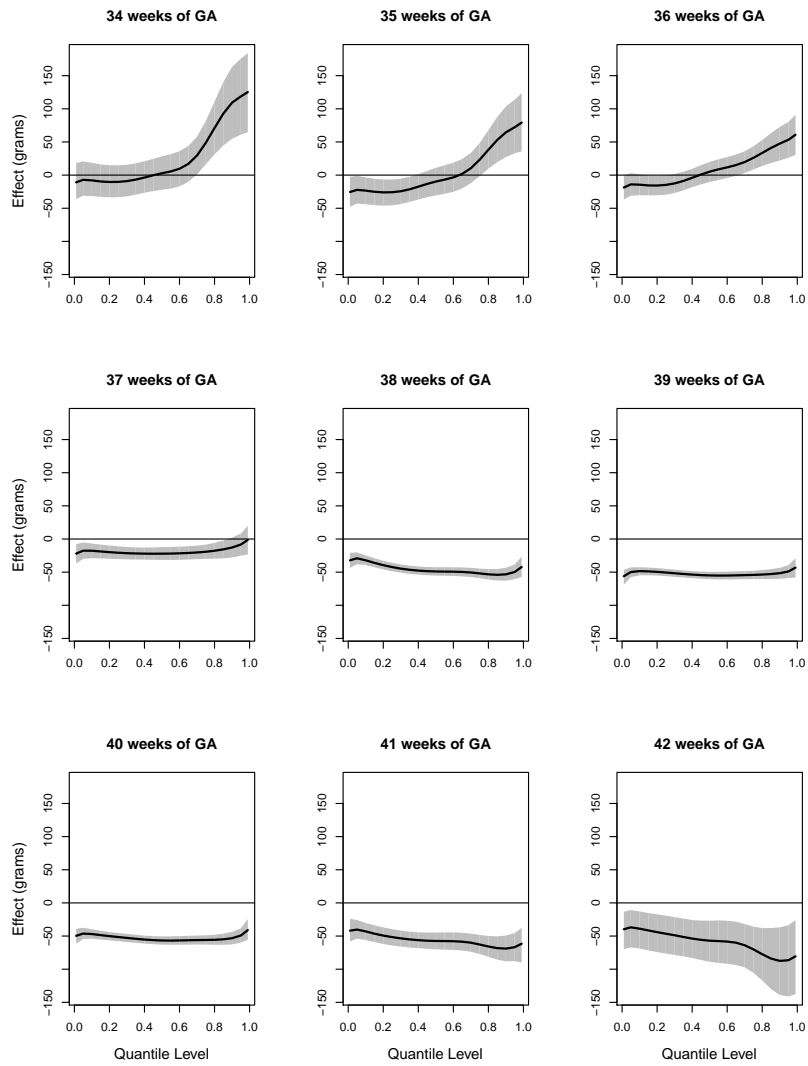


Figure 37. 95% credible limits for the posterior distribution of the effect of maternal Hispanic ethnicity relative to white non-Hispanic ethnicity on birth weight for gestational ages 34-42 weeks.

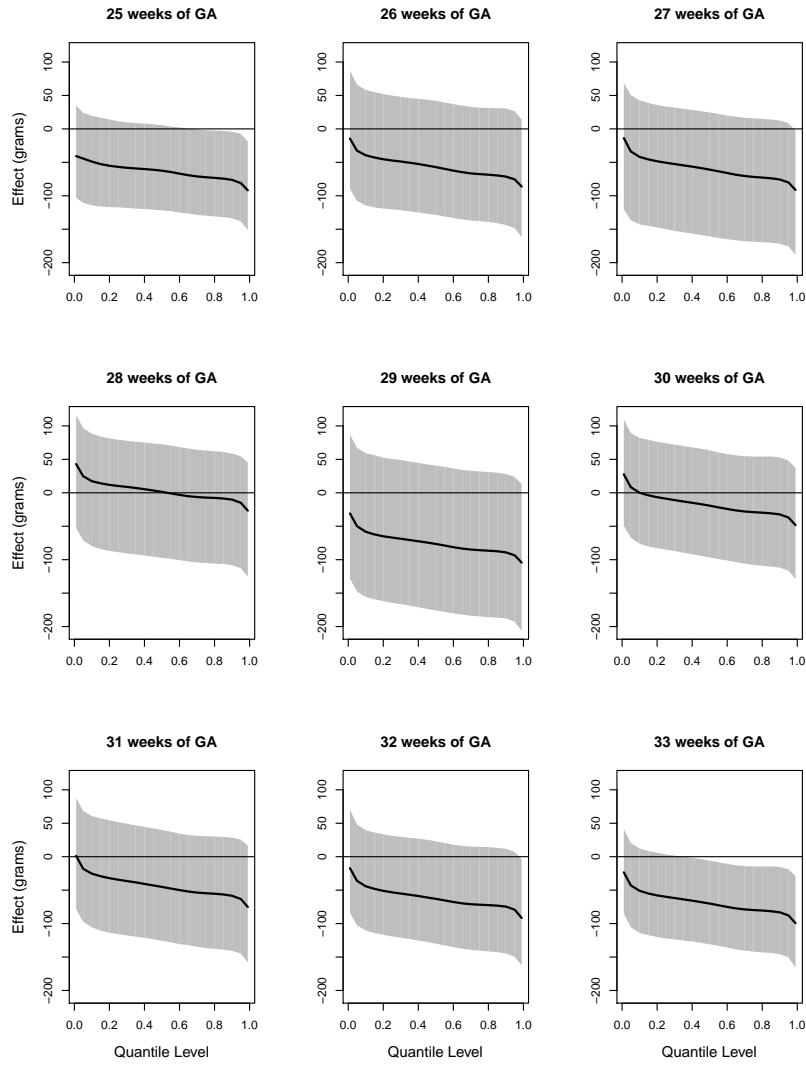


Figure 38. 95% credible limits for the posterior distribution of the effect of maternal other ethnicity relative to white non-Hispanic ethnicity on birth weight for gestational ages 25-33 weeks.

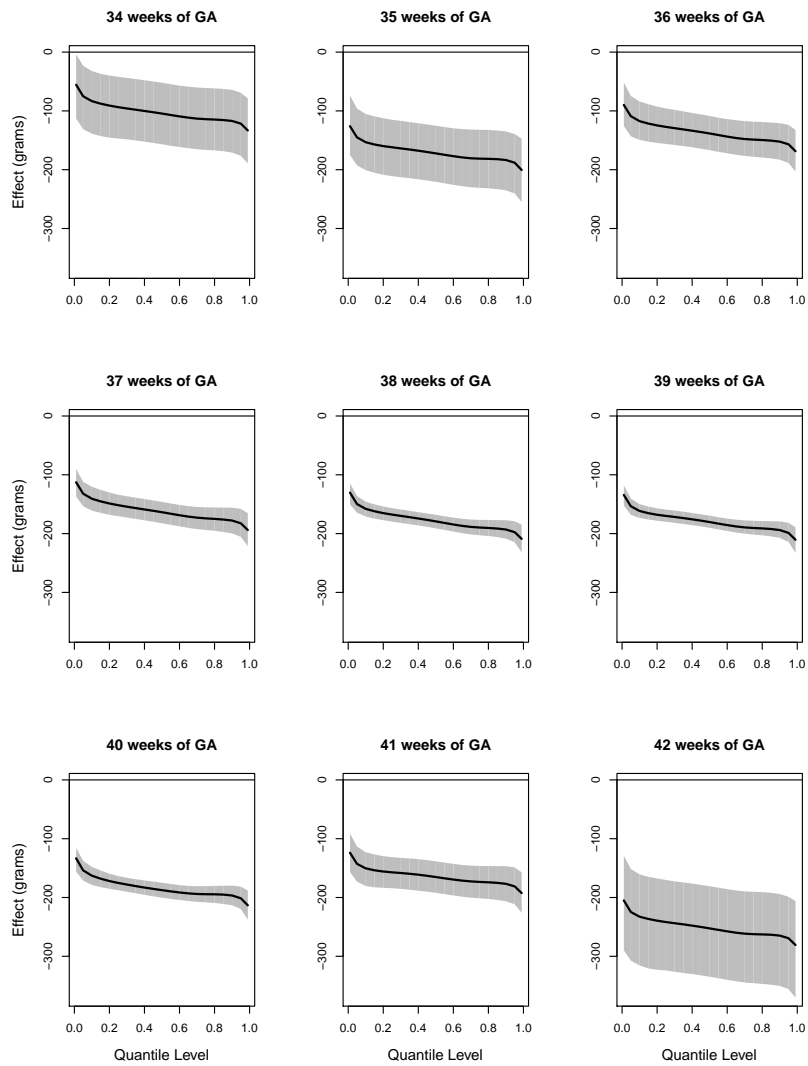


Figure 39. 95% credible limits for the posterior distribution of the effect of maternal other ethnicity relative to white non-Hispanic ethnicity on birth weight for gestational ages 34-42 weeks.

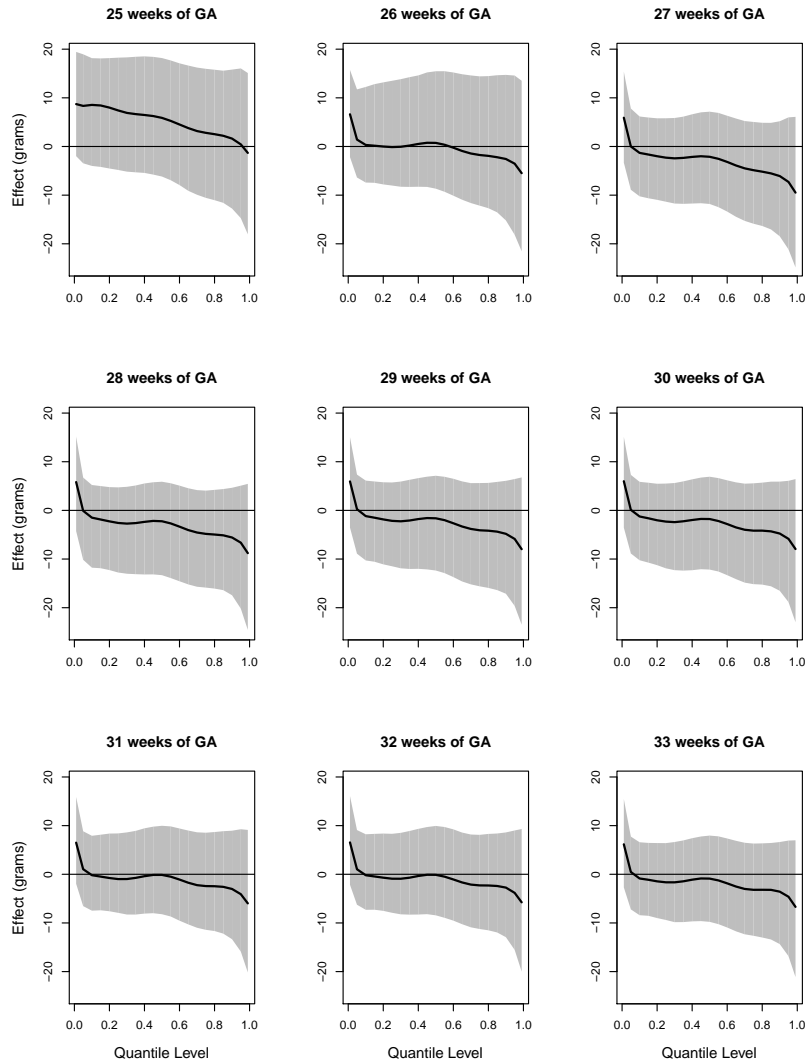


Figure 40. 95% credible limits for the posterior distribution of the effect of a one unit increase in first trimester ozone exposure on birth weight for gestational ages 25-33 weeks. All ozone values were linearly transformed into $[-1,1]$, so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.

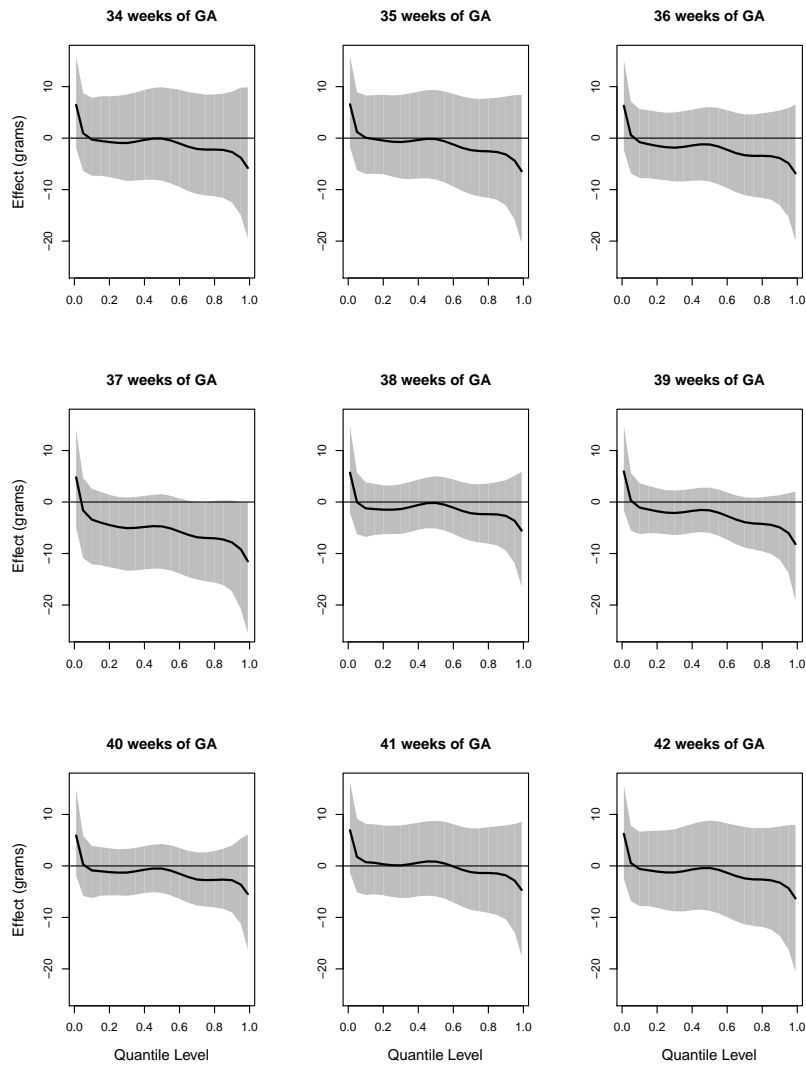


Figure 41. 95% credible limits for the posterior distribution of the effect of a one unit increase in first trimester ozone exposure on birth weight for gestational ages 34-42 weeks. All ozone values were linearly transformed into $[-1,1]$, so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.

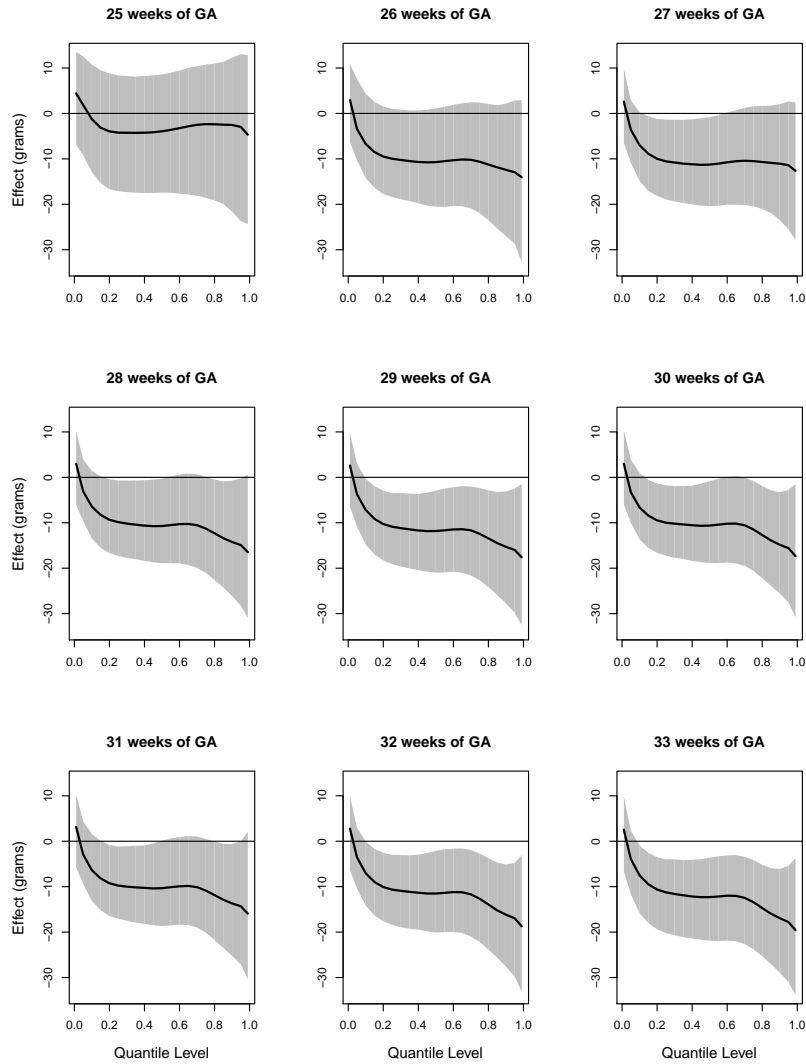


Figure 42. 95% credible limits for the posterior distribution of the effect of a one unit increase in second trimester ozone exposure on birth weight for gestational ages 25-33 weeks. All ozone values were linearly transformed into $[-1,1]$, so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.

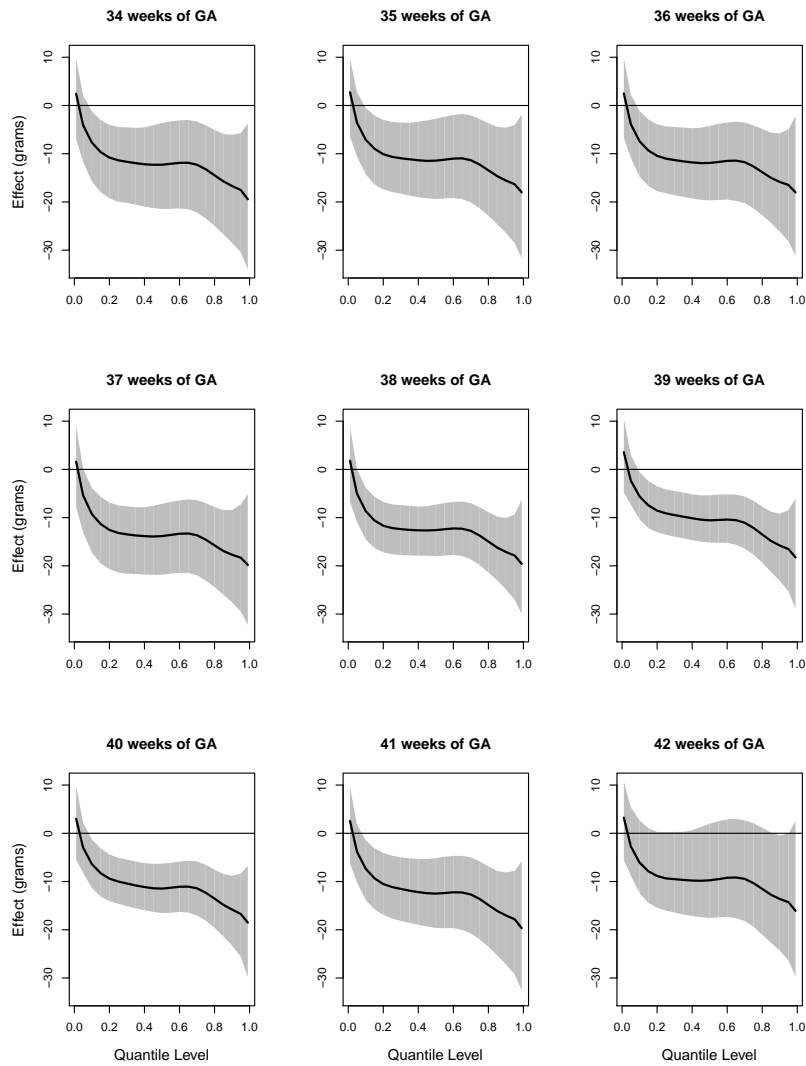


Figure 43. 95% credible limits for the posterior distribution of the effect of a one unit increase in second trimester ozone exposure on birth weight for gestational ages 34-42 weeks. All ozone values were linearly transformed into $[-1,1]$, so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.