Case Study 1 Written Report

Evan Knox, Zhuoqun Wang, Justin Weltz

January 2020

1 Abstract

Observational data on 2380 pregnant women was used to assess possible relationships between pesticide exposure (via DDE and PCB levels) and premature birth. This relationship was analyzed via quantile regression at the 10th and 20th percentiles, as well as ordinal logistic regression based on NIH preterm birth guidelines. Unit increases of total PCB's are related to decreases in gestation age of 0.1-0.2 weeks at these quantiles. Ordinal logistic regression demonstrates significant increases in odds of preterm birth for unit increases in total PCBs or in the first principal component. Overall, there is modest evidence to suggest that PCB exposure is associated with a risk of preterm births.

2 Introduction

The effects of pesticides on the human body are an area of continual interest to the general public and researchers alike, given the growing prevalence of pesticide-aided farming and the corresponding paucity of foodstuffs lacking trace amounts thereof. While many of these concerns focus on the concern of adults or children eating said foods, it is well-known that teratogenic effects (those harmfully altering the growth of humans pre-birth) can be caused by even very limited exposure to certain chemicals; we need look no further than the thalidomide crisis of the mid-20th century.

Not all such chemicals, even those with teratogenic effects, are likely to present such striking effects as thalidomide. Rather, one of the less direct ways we may observe teratogens acting is by causing preterm births, which cause babies to be born before they are fully developed, risking myriad health risks if born sufficiently preterm. Two common pesticides, DDT and PCB, are thought to have this sort of teratogenic effect, and thus must be subjected to careful scrutiny.

3 Materials and Methods

Data on pregnant women was collected through the National Collaborative Perinatal Project. A subset of this data, comprising 2380 women, was provided to us. This data includes rough length of gestation, measures of exposures to pesticides via their stable breakdown products, DDE and various PCB's (breakdown products of DDT and PCB), as well as numerous mostly demographic variables: age of the mother, race, smoking status, center of study entry, 3 normalized scores of SES (income, education, and occupation), and basic measures of cholesterol, triglycerides, and albumin.

The intended purpose of this subset of women was to establish a relationship, or the lack thereof, between exposure to pesticides and preterm births. While the observational and proxy-heavy state of this data made causal interpretations beyond our scope, it was nevertheless ideal to remove the effect of covariates as much as possible, so as to rule out other "causes".

No variables from the dataset were removed due to problems with face validity. Albumin was excluded due to extreme missingness. The three SES scores had nontrivial amounts of missingness as well. However, due to the possibility that these scores might be proxies for prenatal care and education, they were multiply-imputed using the 'mice' package in R. Centers were relevant both due to the possibility of differential missingness, but also as a proxy variable for location or for region. Race has a well-documented relationship with prenatal care, typically demonstrating both increased risk for preterm birth and worse care overall for African-American mothers. Cholesterol and triglycerides are both proxy measures of fat, which stores bioactive chemicals and maintains their levels in the body.

3.1 Exploratory Data Analysis

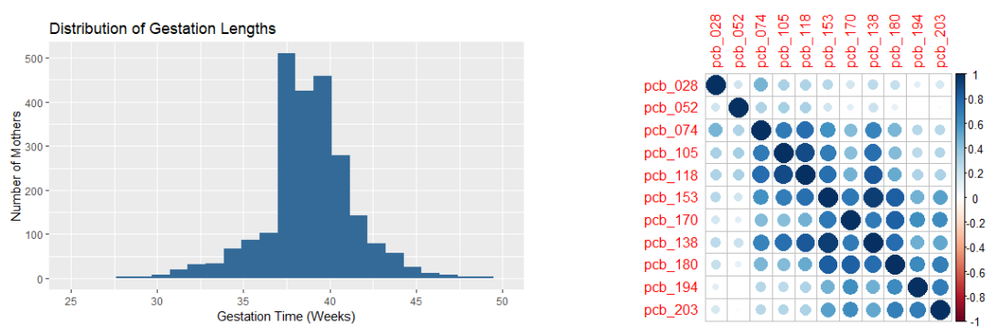

Figure 1: Structure of response variable (gestation length) and some of the predictors (PCB's)

The response variable of interest, gestation length, is displayed above (excluding extreme far-right values, which are presumed to be missing or miscoded). The provided data included a relatively symmetric distribution of gestation lengths centered roughly around 39-40 weeks, which is considered standard term. However, longer gestations than 40 weeks have few if any negative effects, while gestations shorter than 37 weeks [1] can be injurious to neonatal development. It is this need to prioritize prediction of short pregnancies that suggests an alternative to a traditional linear model is needed, along with concerns about having a normal error distribution (see Appendix). These alternatives will be discussed in following sections.

As would be expected in an observational study with lots of variables serving as proxies, there was substantial correlation found amongst the demographic variables in the dataset. Of particular note therein was the PCB collection of variables. The

[1] https://www.nichd.nih.gov/health/topics/pregnancy/conditioninfo

various PCB measures included were all stable breakdown products of the singular pesticide PCB, hence it was unsurprising to find a strong positive correlation structure between the breakdown products. While the boxes shown here may have suggested several principal components were needed to fully capture the PCB data, multiple PCB variables was deemed unnecessary for purposes of model interpretation and due to their lack of notable variance explanation. Two major ways this was dealt with were a sum of all the PCB breakdown products and the most important principal component of the breakdown products. The sum maintained units, and seemed reasonable given the similarity of the PCB distributions to each other. The principal component was used to check that important features of the PCB data were not being excluded.

3.2 Quantile Regression

Since the assumptions for simple linear regression are violated, it is necessary to appeal to alternate methods for examining the relationship between DDE, PCB and gestational age. It is also important to remember that the scope of our question is limited to the effect of these pesticides on the prevalence of premature births. Therefore, it would be ideal if we could focus our analysis on the lower side of the gestational age distribution (in order to pinpoint the effect of DDE and PCB on those most at risk of being premature). Quantile regression accomplishes both of these goals. Not only does this method lack any assumptions on the error distribution, it also allows us to easily examine the effect of the covariates on different gestational age quantiles. The interpretation of quantile regression results is fairly intuitive as well: While normal linear regression estimates the effect of covariates on the conditional mean of the response, quantile regression estimates the effect on the conditional quantile of the response distribution. The algorithm accomplishes this goal by minimizing equation 1.

$$\tau \sum_{y_i > \hat{\beta_\tau^T} X_i} |y_i - \hat{\beta_\tau^T} X_i| + (1-\tau) \sum_{y_i < \hat{\beta_\tau^T} X_i} |y_i - \hat{\beta_\tau^T} X_i| \tag{1}$$

where τ is the quantile being estimated. This equation makes some intuitive sense if we substitute $\tau = 0.5$. In this case, minimizing equation 1 is equivalent to minimizing absolute error. We know that the estimator that minimizes absolute error is the median, which is the 50% quantile, and everything lines up!

3.3 Ordinal Logistic Regression

Another method to model the nonlinear relationship between DDE, PCB and gestational age is to classify the gestational ages. Although gestational age is naturally ordered, the distance between different classes of gestational age, e.g, between preterm and full term, is unknown. Hence it is reasonable to treat the gestational age as ordinal instead of continuous and analyze them with ordinal logistic regression. This method also allows our analysis to focus on premature deliveries by categorizing the responses as "extremely preterm" ($\leq$ 28 weeks), "very preterm" (29-32 weeks), "moderate/late preterm" (33-37 weeks) and "not preterm" ($>$ 37 weeks).

Ordinal logistic regression relies on four basic assumptions: (1) The dependent variable is ordinal, (2) One or more of the independent variables are either continuous, categorical or ordinal, (3) There is no collinearity of independent variables, and (4) Each independent variable has an identical effect at the cumulative split of the ordinal dependent variable. The dataset of interest satisfies all of the assumptions. The non-collinearity assumption of our method is gauranteed by the reasonably small sample correlation, and the proportioal odds assumption is verified by the Brant test.

The basic model of this method is

$$logit P(Y \leq m_j) = \beta_{j0} - \eta_1 x_1 - \cdots - \eta_p x_p, \tag{2}$$

where Y is the gestational age, $x_1, \cdots, x_p$ are the covariates, and m_j is the threshold between the jth and $(j+1)$th category, so the m_j's are 28,32 and 37.

With the estimates $\hat{\beta_{j0}}$'s and $\hat{\eta_k}$'s, the association between gestational age and covariate x_k is quantified by $\exp(-\hat{\eta_k})$, i.e., we expect $100(\exp(-\hat{\eta_k})-1)\%$ increase in the odds of earlier delivery with one unit increase of x_k, holding constant all other variables.

4 Results

4.1 Quantile Regression

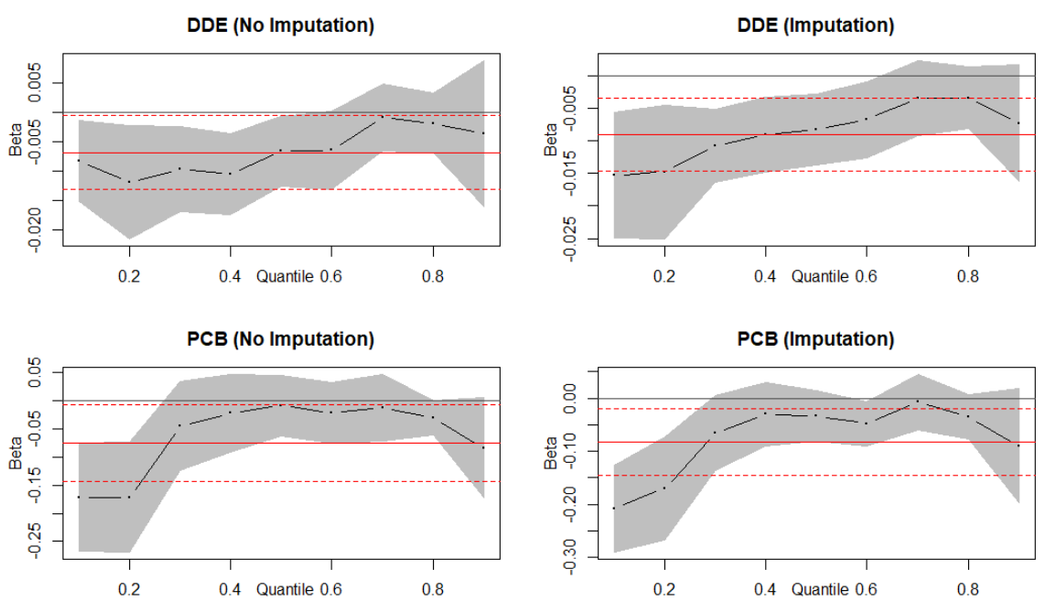

Figure 2: Quantile Coefficient Estimates

Figure 2 contains a lot of interesting information. First it's important to notice the increasing trend in the coefficients associated with PCB and DDE as the quantile increases. This not only demonstrates the larger effect of DDE and PCB on the lower quantiles of gestational age, but also further illustrates the problems inherent to modeling this relationship with a linear regression. For reference, the solid red line and the dashed red lines mark the linear regression estimate and the 95% confidence interval associated with that estimate respectively. If the data was suited to a linear regression model, we would expect the coefficient estimates to stay well within the bounds of the standard error bars and randomly oscillate around the global estimate. Therefore, the clear upwards trend is further evidence that a linear regression would have been inadequate when modeling the effect of DDE and PCB on gestational age. We can also clearly see from the non-imputed data that the pesticides have a much more negative effect on gestational ages in the lower quantiles. Since the 10% and 20% quantiles include premature gestational ages and those on the healthy/unhealthy threshold, we can tentatively conclude that the pesticides significantly increase the prevalence of adverse birth outcomes by both increasing the likelihood of premature births and making even earlier (and therefore riskier) births more likely for those already in the premature quantile. Imputing the socioeconomic scores produces very similar results, but smooths out the increasing trend of the coefficients in an encouraging way (the distribution of the imputed data lines up very well with the distribution of the actual data, seen in figure 3, so we wouldn't expect much analysis to change) .

Finally, let's interpret the quantile regression coefficients in the context of our original question. As we can see in the table 2 and 3 in the appendix, the results suggest that for every 1 unit increase in the sum of the pcb concentration indicators (measured in ng/dL), there is a $0.17 - 0.18$ week decrease in the 10^{th} percentile of gestational age controlling for a host of covariates. For the 20^{th} percentile the effect size does not change much. Similarly, for every 1 unit increase in dde (measured in ug/dL), there is a smaller, but still significant decrease of about 0.01 weeks (also in table 2 and 3 in the appendix) in the 10^{th} and 20^{th} percentiles of gestational age. It's also important to notice that the units imply the effect of the summed PCB concentration indicators are even larger relative to DDE since there are 1000 nanograms to every 1 microgram. Interestingly, after the 10^{th} and 20^{th} percentiles the PCB coefficients becomes generally insignificant while the small DDE coefficient stays significant until the very top of the gestational age distribution. In conclusion, it is clear that DDE and PCB significantly increase the likelihood of premature birth with a noticeable effect size (especially in the case of PCB).

4.2 Ordinal Logistic Regression

In addition to the ordinal logistic regression model with sum of PCB's, we also fitted one with the first principal component of PCB's as a sensitivity analysis. The two models with different covariates of PCB's ended up providing the same estimates of the coefficient of DDE and consistent results, which justified using sum of PCB's in our analysis. More notes on the sum of PCB's are shown in the discussion section.

Holding constant all other variables, for a one unit increase in the predictor, the amount that the odds of earlier delivery (e.g., extremely preterm vs very preterm) is multiplied by are shown in Table 1 and Table 2, where coefficients are estimated without and with imputation respectively. In conclusion, PCB and DDE are associated with premature delivery: with one unit increase in DDE (measured in ug/dL), sum of PCB (measured in ng/dL) or first principal component of PCB, the odds of earlier delivery is expected to increase approximately $0.4\%, 8\%$ and 6% respectively, if the other variables are unchanged. Note that the units of DDE and PCB are different, so the effect of PCB is much stronger than DDE. The sample standard deviation of DDE, sum of PCB and first principal component of PCB are 19.89, 1.87 and 2.44, we conclude that with a one standard deviation increase of DDE, the odds of earlier delivery is expected to increase approximately 9%; a one standard deviation increase in either the sum of PCB or first principal component of PCB is associated with a 16% increase in the odds of earlier delivery.

Table 1: Ordinal Logistic Regression Results (Left: without imputation; Right: with imputation)

predictor	estimate	25%	97.5%
DDE	1.004	0.998	1.010
sum of PCB	1.084	1.009	1.161
PC1 of PCB	1.064	1.007	1.122

predictor	estimate	25%	97.5%
DDE	1.005	1.000	1.010
sum of PCB	1.088	1.022	1.156
PC1 of PCB	1.068	1.017	1.119

5 Discussion

When first analyzing the data, we applied both PCA and imputation to augment the data in a productive way. PCA captures the low-dimensional structure of PCBs, but makes the interpretation of estimated coefficients more difficult. The sum of PCB enjoys better interpretability since it preserves the unit of PCB, but it may fail to capture the variability of PCBs especially when some of them are negatively correlated. Other possible choices to solve the problem of collinearity include Bayesian ordinal regression with shrinkage priors on the coefficients. It also might be interesting to try imputing the data with a Bayesian method that incorporates our uncertainty about the imputed values.

Our inability to appropriately model the data and answer the question of interest with linear regression forced us to explore different methods. While ordinal regression allowed us to pinpoint different preterm birth risk categories and model the effect of DDE and PCB on the likelihood of these outcomes without forcing a linear relationship, it forced us to pick exact cutoffs that seemed arbitrary even though we tried to inform our choices with NIH data. On the other hand, Quantile regression seems to naturally establish groupings and allow us to examine the pesticides effect on the population we are interested in, however it still forces a linear relationship. Both methods have their advantages and disadvantages, however the fact that they both point to the effect of DDE and PCB on preterm deliveries is very encouraging. For future directions, we could try a more flexible model: possibly a non-parametric quantile regression that captures more nuances of the dose-response curve. Additionally, it may be interesting to apply a hierarchical Bayesian alternative to this data in order to capture more of the centers structure in our analysis.

6 Appendix

Imputation Results:

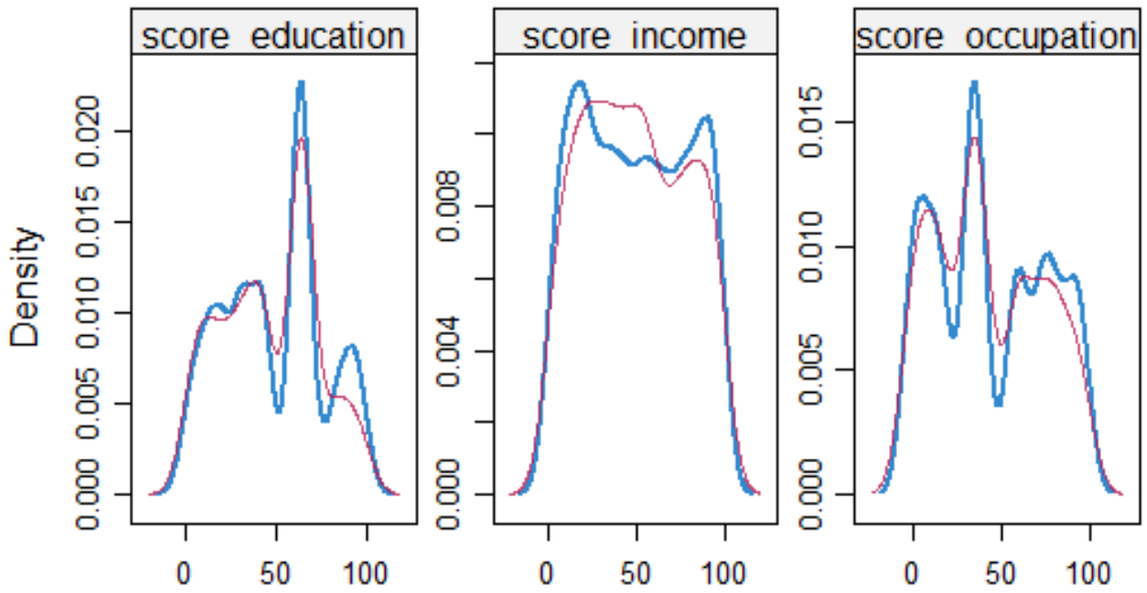

Figure 3: Imputed (Red) vs. Actual (Blue) Densities for the Imputed Score Variables

Histograms of Pesticide Marker Variables

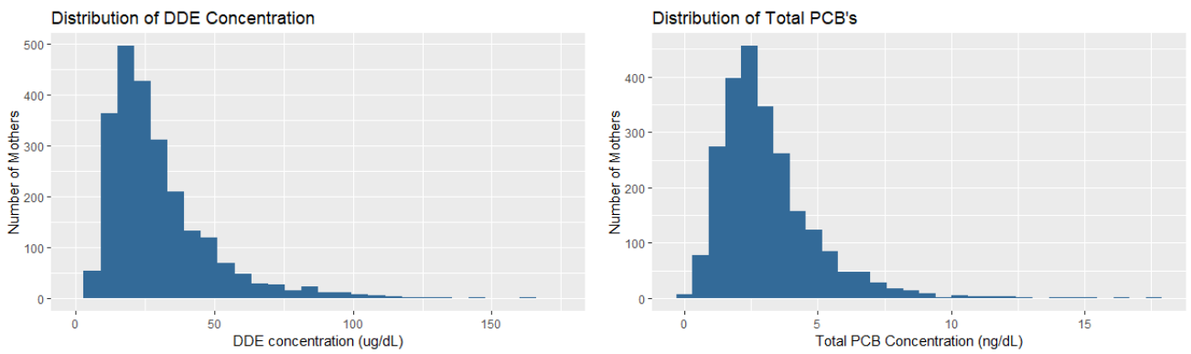

Figure 4: Basic histograms of the pesticide markers. Note the difference in units.

Linear Model Diagnostics

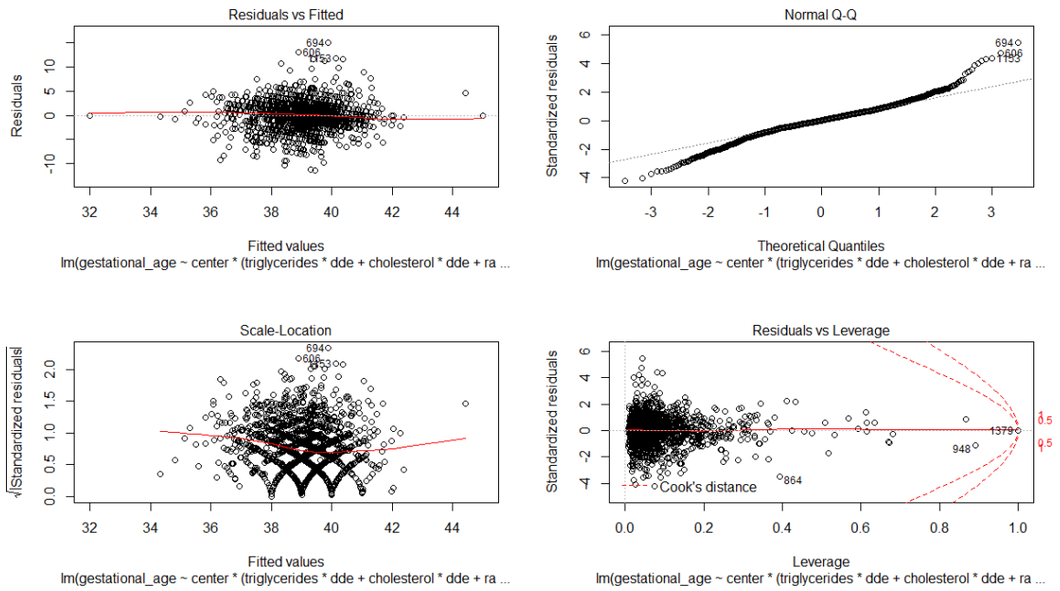

Figure 5: Seen here are the diagnostic plots for a traditional linear regression using the available data. The QQ plot demonstrates that the errors may not be normal, and we can also see some underlying structure to the errors in the scale-location graph.

Table 2: Quantile Regression Results (No Imputation)

Covariates	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(Intercept)	36.504	38.068	39.226	39.579	40.581	41.374	41.704	42.418	43.559
	(0.643)	(0.642)	(0.457)	(0.439)	(0.383)	(0.363)	(0.431)	(0.378)	(0.399)
pcb	− 0.171	− 0.172	− 0.045	− 0.022	− 0.009	− 0.022	− 0.013	− 0.031	− 0.084
	(0.058)	(0.059)	(0.048)	(0.042)	(0.033)	(0.032)	(0.036)	(0.019)	(0.054)
dde	− 0.008	− 0.012	− 0.010	− 0.011	− 0.007	− 0.007	− 0.001	− 0.002	− 0.004
	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.008)
cholesterol	0.003	0.003	0.001	0.001	0.000	− 0.001	− 0.001	− 0.002	0.000
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
triglycerides	− 0.003	− 0.004	− 0.003	− 0.003	− 0.004	− 0.004	− 0.003	− 0.003	− 0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
raceblack	− 0.898	− 0.997	− 0.925	− 0.859	− 0.863	− 0.612	− 0.565	− 0.671	− 0.789
	(0.338)	(0.280)	(0.192)	(0.214)	(0.175)	(0.189)	(0.211)	(0.185)	(0.365)
raceother	− 3.252	− 0.900	− 0.507	− 0.651	− 0.372	− 0.072	0.033	− 0.232	− 0.808
	(0.884)	(0.829)	(0.466)	(0.325)	(0.428)	(0.369)	(0.382)	(0.214)	(0.596)
smoking$_s$*tatus*	− 0.177	− 0.089	− 0.044	− 0.038	− 0.102	− 0.149	− 0.140	− 0.103	− 0.090
	(0.207)	(0.175)	(0.121)	(0.125)	(0.111)	(0.106)	(0.124)	(0.112)	(0.143)
maternal$_a$*ge*	0.000	0.003	− 0.001	− 0.004	− 0.008	− 0.012	− 0.010	− 0.005	− 0.005
	(0.015)	(0.015)	(0.012)	(0.010)	(0.009)	(0.008)	(0.010)	(0.009)	(0.008)
score$_o$*ccupation*	0.006	0.004	0.001	0.000	0.000	− 0.001	0.000	− 0.003	− 0.001
	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
score$_i$*ncome*	0.006	0.003	0.001	0.002	0.001	0.000	0.000	0.001	0.005
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
score$_e$*ducation*	0.005	0.005	0.003	0.003	0.002	0.003	0.000	0.003	− 0.005
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
as.factor(center)10	0.542	0.318	0.161	0.192	0.230	0.271	0.423	0.565	0.768
	(0.323)	(0.309)	(0.190)	(0.230)	(0.207)	(0.260)	(0.291)	(0.276)	(0.135)
as.factor(center)15	− 1.381	− 0.512	− 0.588	0.118	0.102	− 0.107	0.014	− 0.171	0.658
	(1.032)	(0.655)	(0.528)	(0.373)	(0.329)	(0.343)	(0.318)	(0.378)	(0.700)
as.factor(center)31	1.159	0.859	0.243	0.827	0.440	0.296	0.158	0.386	0.798
	(0.991)	(0.301)	(0.392)	(0.236)	(0.259)	(0.268)	(0.240)	(0.437)	(0.461)
as.factor(center)37	− 1.374	− 1.136	− 0.791	− 0.173	− 0.021	0.092	− 0.026	0.079	0.689
	(0.390)	(0.605)	(0.395)	(0.367)	(0.300)	(0.245)	(0.269)	(0.239)	(0.564)
as.factor(center)45	− 0.318	0.557	0.535	0.887	0.453	0.314	0.251	0.938	2.049
	(0.593)	(0.613)	(0.431)	(0.278)	(0.184)	(0.287)	(0.401)	(0.487)	(0.708)
as.factor(center)50	− 0.013	− 0.235	− 0.387	0.209	0.064	0.216	0.484	0.568	0.783
	(0.228)	(0.318)	(0.240)	(0.307)	(0.217)	(0.303)	(0.274)	(0.255)	(0.733)
as.factor(center)55	2.122	0.966	0.682	1.184	0.762	0.700	0.942	0.987	0.778
	(0.391)	(0.388)	(0.729)	(0.582)	(0.369)	(0.418)	(0.612)	(0.247)	(1.650)
as.factor(center)60	− 0.806	− 0.338	0.038	0.069	0.372	0.128	0.295	0.502	1.550
	(1.051)	(0.583)	(0.274)	(0.344)	(0.325)	(0.225)	(0.390)	(0.421)	(0.961)
as.factor(center)66	0.075	0.776	0.196	0.293	0.117	0.113	0.072	0.366	1.440
	(0.484)	(0.376)	(0.221)	(0.251)	(0.234)	(0.232)	(0.283)	(0.214)	(0.411)
as.factor(center)71	0.818	0.383	0.057	0.235	0.172	− 0.036	− 0.296	− 0.031	0.254
	(1.086)	(0.343)	(0.274)	(0.299)	(0.205)	(0.137)	(0.264)	(0.276)	(0.436)
as.factor(center)82	− 0.629	− 0.275	− 0.433	− 0.041	− 0.012	− 0.320	− 0.140	− 0.095	0.272
	(0.706)	(0.448)	(0.424)	(0.395)	(0.314)	(0.324)	(0.289)	(0.350)	(0.524)

Table 3: Quantile Regression Results (Imputation)

Covariates	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(Intercept)	35.805	37.878	39.008	39.506	40.350	41.109	41.736	42.378	44.315
	(0.697)	(0.549)	(0.436)	(0.397)	(0.371)	(0.288)	(0.344)	(0.412)	(0.599)
pcb	− 0.180	− 0.166	− 0.058	− 0.035	− 0.029	− 0.041	− 0.026	− 0.030	− 0.052
	(0.047)	(0.061)	(0.044)	(0.039)	(0.031)	(0.027)	(0.034)	(0.024)	(0.069)
dde	− 0.013	− 0.014	− 0.011	− 0.009	− 0.008	− 0.007	− 0.002	− 0.004	− 0.011
	(0.005)	(0.006)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.006)
cholesterol	0.004	0.003	0.003	0.001	0.001	0.000	− 0.001	− 0.002	− 0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
triglycerides	− 0.004	− 0.003	− 0.003	− 0.003	− 0.004	− 0.003	− 0.003	− 0.003	− 0.005
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
raceblack	− 0.447	− 0.345	− 0.710	− 0.475	− 0.679	− 0.453	− 0.537	− 0.599	− 0.629
	(0.330)	(0.290)	(0.162)	(0.185)	(0.166)	(0.150)	(0.156)	(0.169)	(0.356)
raceother	− 0.522	− 0.224	− 0.510	− 0.322	− 0.236	− 0.146	0.320	0.428	0.356
	(1.096)	(0.608)	(0.356)	(0.268)	(0.347)	(0.357)	(0.499)	(0.377)	(0.591)
smoking$_s$*tatus*	− 0.004	− 0.128	− 0.022	− 0.072	− 0.067	− 0.141	− 0.115	− 0.033	− 0.098
	(0.194)	(0.162)	(0.108)	(0.114)	(0.104)	(0.083)	(0.102)	(0.119)	(0.176)
maternal$_a$*ge*	0.036	0.009	0.004	0.003	0.001	− 0.009	− 0.009	− 0.006	− 0.012
	(0.018)	(0.013)	(0.011)	(0.009)	(0.008)	(0.007)	(0.008)	(0.009)	(0.015)
score$_o$*ccupation*	0.002	0.004	0.002	0.002	0.001	0.001	0.001	− 0.001	0.002
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
score$_i$*ncome*	0.006	0.001	0.000	0.000	0.000	− 0.001	− 0.001	0.000	0.001
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
score$_e$*ducation*	0.005	0.005	0.002	0.001	0.001	0.002	0.001	0.002	− 0.008
	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
as.factor(center)10	0.489	0.243	0.148	0.028	0.175	0.239	0.304	0.498	0.686
	(0.239)	(0.273)	(0.190)	(0.266)	(0.233)	(0.267)	(0.227)	(0.268)	(0.229)
as.factor(center)15	− 2.083	− 1.366	− 0.960	− 0.366	− 0.083	− 0.345	− 0.055	0.052	0.908
	(0.969)	(0.591)	(0.499)	(0.374)	(0.287)	(0.301)	(0.215)	(0.405)	(0.639)
as.factor(center)31	0.263	0.205	0.019	0.346	0.213	0.158	− 0.163	− 0.230	0.500
	(0.440)	(0.314)	(0.479)	(0.238)	(0.263)	(0.105)	(0.176)	(0.316)	(0.539)
as.factor(center)37	− 1.421	− 1.446	− 0.947	− 0.577	− 0.173	− 0.005	− 0.093	0.054	0.192
	(0.486)	(0.472)	(0.336)	(0.312)	(0.318)	(0.200)	(0.207)	(0.206)	(0.567)
as.factor(center)45	− 0.137	− 0.135	0.241	0.348	0.264	0.202	0.213	0.930	1.642
	(0.387)	(0.496)	(0.399)	(0.251)	(0.222)	(0.232)	(0.375)	(0.415)	(0.555)
as.factor(center)50	− 0.330	− 0.285	− 0.341	− 0.274	− 0.111	− 0.099	0.130	0.297	0.464
	(0.274)	(0.279)	(0.259)	(0.260)	(0.202)	(0.175)	(0.287)	(0.279)	(0.705)
as.factor(center)55	− 1.141	− 0.936	− 0.169	0.134	0.363	0.314	0.313	0.614	1.115
	(1.027)	(0.830)	(0.383)	(0.325)	(0.369)	(0.259)	(0.428)	(0.295)	(0.698)
as.factor(center)60	− 1.256	− 0.268	− 0.159	− 0.159	0.120	0.018	0.395	0.798	1.470
	(0.638)	(0.527)	(0.290)	(0.291)	(0.261)	(0.236)	(0.348)	(0.432)	(0.582)
as.factor(center)66	− 0.665	− 0.081	− 0.055	− 0.222	− 0.171	− 0.083	− 0.117	0.260	0.805
	(0.399)	(0.346)	(0.195)	(0.218)	(0.208)	(0.202)	(0.200)	(0.222)	(0.419)
as.factor(center)71	− 0.470	0.036	− 0.094	0.123	0.052	− 0.084	− 0.292	− 0.157	− 0.098
	(1.070)	(0.387)	(0.194)	(0.283)	(0.195)	(0.122)	(0.210)	(0.175)	(0.396)
as.factor(center)82	− 1.049	− 0.932	− 0.702	− 0.698	− 0.237	− 0.529	− 0.294	− 0.257	− 0.019
	(0.691)	(0.426)	(0.348)	(0.358)	(0.338)	(0.296)	(0.327)	(0.288)	(0.513)

Table 4: Brant Test Results (PCA)

Test for	X2	df	probability
Omnibus	24.27	44.00	0.99
dde	1.85	2.00	0.40
triglycerides	0.60	2.00	0.74
raceblack	1.87	2.00	0.39
raceother	8.16	2.00	0.02
score_education	0.26	2.00	0.88
score_income	0.94	2.00	0.62
score_occupation	1.24	2.00	0.54
maternal_age	0.60	2.00	0.74
smoking_status1	3.16	2.00	0.21
cholesterol	1.07	2.00	0.59
center10	0.00	2.00	1.00
center15	0.43	2.00	0.81
center31	0.00	2.00	1.00
center37	1.31	2.00	0.52
center45	1.54	2.00	0.46
center50	0.91	2.00	0.63
center55	1.52	2.00	0.47
center60	0.13	2.00	0.94
center66	0.66	2.00	0.72
center71	0.30	2.00	0.86
center82	1.11	2.00	0.57
pc1	4.06	2.00	0.13

Table 5: Brant Test Results (sum of PCBs)

Test for	X2	df	probability
Omnibus	24.60	44.00	0.99
dde	2.04	2.00	0.36
triglycerides	0.55	2.00	0.76
raceblack	1.87	2.00	0.39
raceother	8.21	2.00	0.02
score_education	0.28	2.00	0.87
score_income	0.88	2.00	0.64
score_occupation	1.40	2.00	0.50
maternal_age	0.50	2.00	0.78
smoking_status1	3.11	2.00	0.21
cholesterol	1.15	2.00	0.56
center10	0.00	2.00	1.00
center15	0.35	2.00	0.84
center31	0.00	2.00	1.00
center37	1.22	2.00	0.54
center45	1.41	2.00	0.49
center50	0.85	2.00	0.65
center55	1.45	2.00	0.48
center60	0.20	2.00	0.91
center66	0.60	2.00	0.74
center71	0.26	2.00	0.88
center82	1.03	2.00	0.60
pcbsum	4.74	2.00	0.09