Assessing Effects of Exposures to DDE and PCBs on Premature Delivery via Ordinal Logistic Regression

Raphael Morsomme Rihui Ou Alessandro Zito

Case Study 1 - Stat 723

January 20, 2020

Overview

Introduction

- **Framework**:
  Dichlorodiphenyldichloroethylene (DDE) and *Polychlorinated Biphenyls* (PCBs) are chemicals that persist in the envirnoment and get stored in fatty depositis in the human tissues.
  $\Longrightarrow$ Potential adverse effect on health
- **Question**:
  Is exposure to DDE and PBCs associated with a higher chance of premature delivery in pregnant women?

Pregnancy timeline

- **Dangerous preterm**: delivery at 34 weeks or before (when main organs are underdeveloped)
- **Preterm**: delivery beween 35 and 37 week
- **At term**: delivery after 37 weeks

Data

Data collected by 12 centers contained gestational age (in weeks) of the mother, the DDE and PCBs concentration, socio-economic info and scores (race, occupation, education, income), amount of triglycerides and cholesterol in blood and smoking status.

Preprocessing:

- Drop obs. with gestational age > 45 (the world record)
- Standardize and average levels of PCBs[1]

$$PCB_i = \frac{1}{11}\sum_{j=1}^{11}\frac{PCB_{ij} - mean_i(PCB_{ij})}{sd_i(PCB_{ij})}$$

- Mean impute of occupation, education and income scores
- Aggregate race into $race = 1$ if white and $race = 0$ if non-white

$\Longrightarrow$ Total obs. $=$ **2336**

[1] This avoids the correlation between the PCBs. See the appendix.

Data

- **Our dependent varible is**:

$$gestgroup_i = \begin{cases} 0 & if \text{ Dangerous preterm} \\ 1 & if \text{ Preterm} \\ 2 & if \text{ At term} \end{cases}$$

- To account for triglyceredes and cholesterol, we introduce an **adjusted measure for** *PCB* **and** *DDE* by:
  1. Computing total lipids using Phillips et al.(1989) and Bernert et al.(2007) forumula

$$lipid_i = 2.27 * cholesterol_i + triglycerides_i + 0.623$$

  2. Setting[2]

$$adjDDE_i = \frac{DDE_i}{log(lipid_i)} \qquad adjPCB_i = \frac{PCB_i}{log(lipid_i)}$$

[2]The choice of the log comes from a Box-Cox analysis of the log-likelihood, as in Li, Longnecker and Dunson (2013)

EDA (I) - Exposures and gestational groups by race

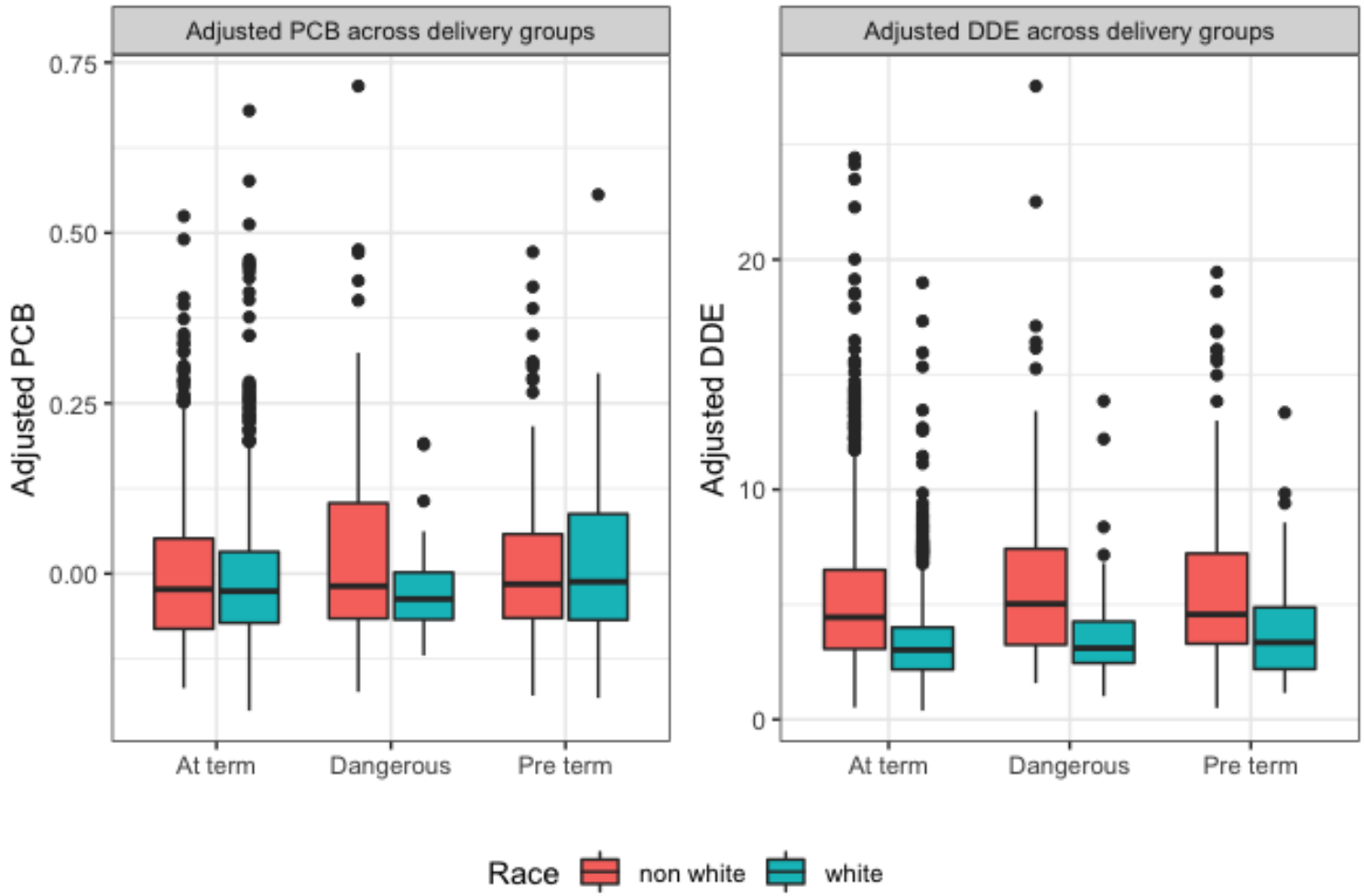

Figure: Relationship between delivery group and adjusted exposures, by race

EDA (II) - Exposure across centers

Model (I) - Ordinal Logistic Regression

We run the following ordinal logistic regression model:

$$\text{logit}(P(\textit{gestgroup} \leq j)) = \beta_{0j} - \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $j = 0, 1, 2$ corresponds to the outcome level, and $\mathbf{X}$ contains:

- *adjDDE*, *adjPCB*, *race*, *center*, *smoke*, the 3 scores, and mother age [main effects]
- (*adjDDE* + *adjPCB*) * (*race* + *center*) [interactions].

AIC-based backward variable selection:

- Maintain *DDE*, *PCB*, *smoke*, *center*, *race*, (*PCB* + *DDE*) * *race*
- Drop (*DDE* + *PCB*) * *center* , mother age,

Model assumptions are checked in the appendix.

Model (II) - Bayesian Ordinal Logistic Regression

Numerical Results

	5%	95%
dde_env_bc	-0.0513582	0.0097459
pcb_env_bc	-2.7673825	-0.7043851
race_aggwhite	0.0854250	0.8561411
center10	-0.1122673	0.7602643
center15	-1.0956887	-0.2863208
center31	-0.2567563	0.8024497
center37	-1.0273528	-0.3027346
center45	-0.4292124	0.3496075
center50	-0.5401077	0.2402548
center55	-0.7698277	0.0759588
center60	-0.6561040	0.1663142
center66	-0.4817386	0.1849701
center71	-0.4542238	0.2900110
center82	-1.0930024	-0.2869819
smoking_status1	-0.3170003	-0.0014069
dde_env_bc:race_aggwhite	-0.1174506	0.0160157

Graphical Results

Conclusions

Appendix

More EDA

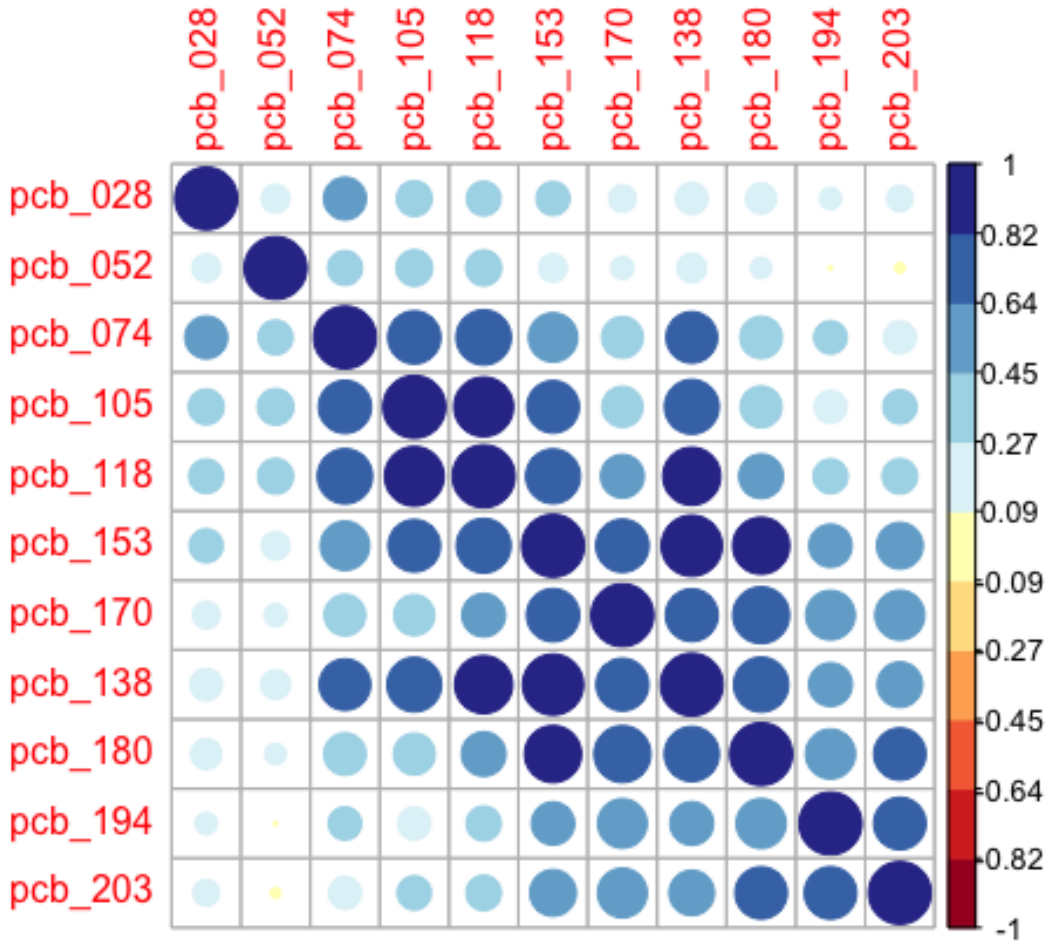

Figure: Correlation plot across PCBs

Appendix

Frequentist Model Checking

We can check the assumption of the (frequentist) ordinal logistic model by looking at the Surrogate residuals. ADD CITATION HERE

If the model assumptions are correct, then the surrogate residuals R_S will have three properties:

1. $E(R_S|X) = 0$
2. $Var(R_S|X) = c$, the conditional variance of R_S is constant
3. The emiprical distribution of R_S resembles an explicit distribution that is related to the link function $G^{-1}(\cdot)$. Specifically, $R_S \sim G(c + \int u dG(u))$.

Appendix

Frequentist Model Checking

Assumptions (i) and (ii) are checked with the Surrogate residuals plot. Both are satisfied in this case.

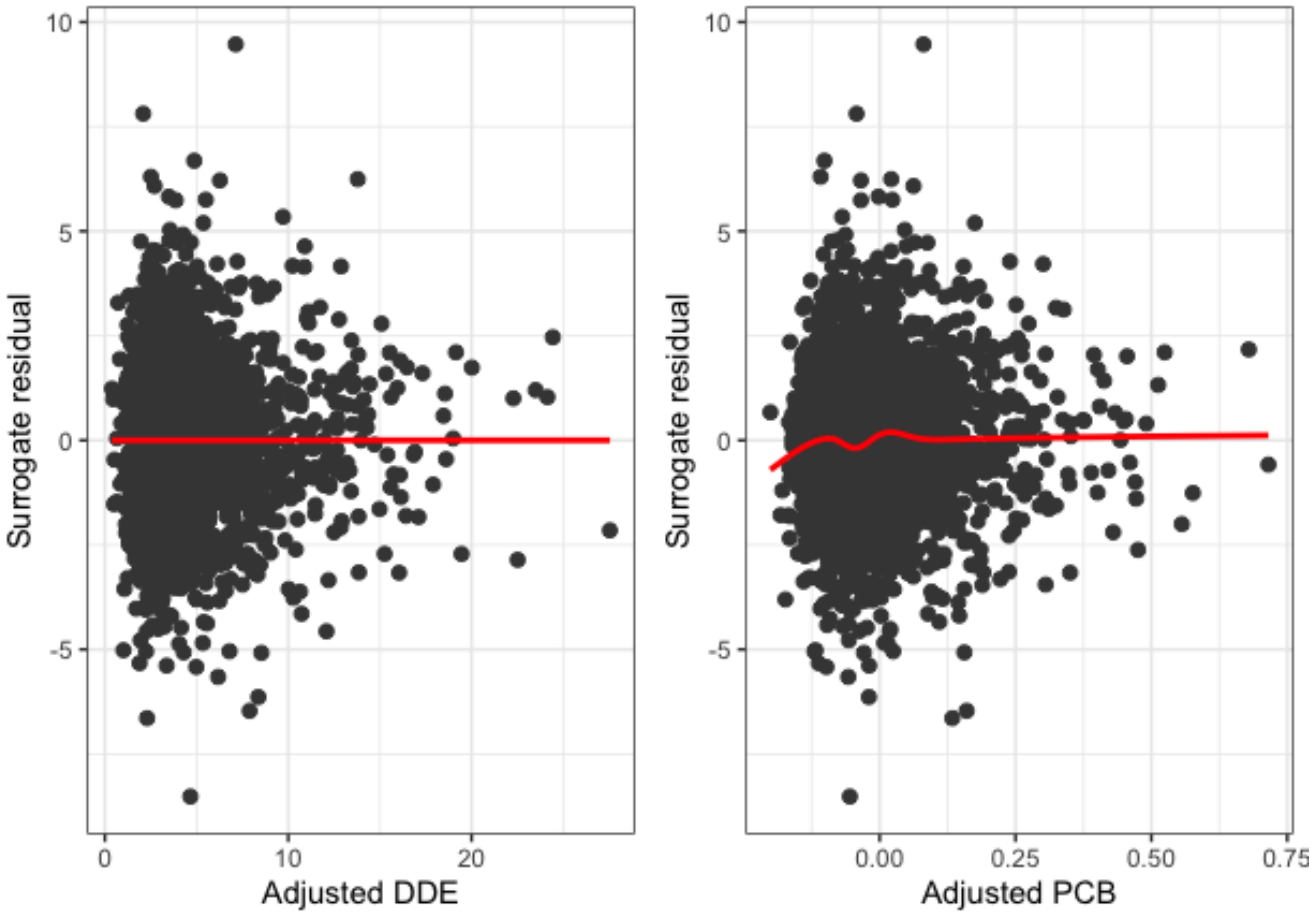

Figure: Surrogate residuals of DDE and PCB

Appendix

Frequentist Model Checking

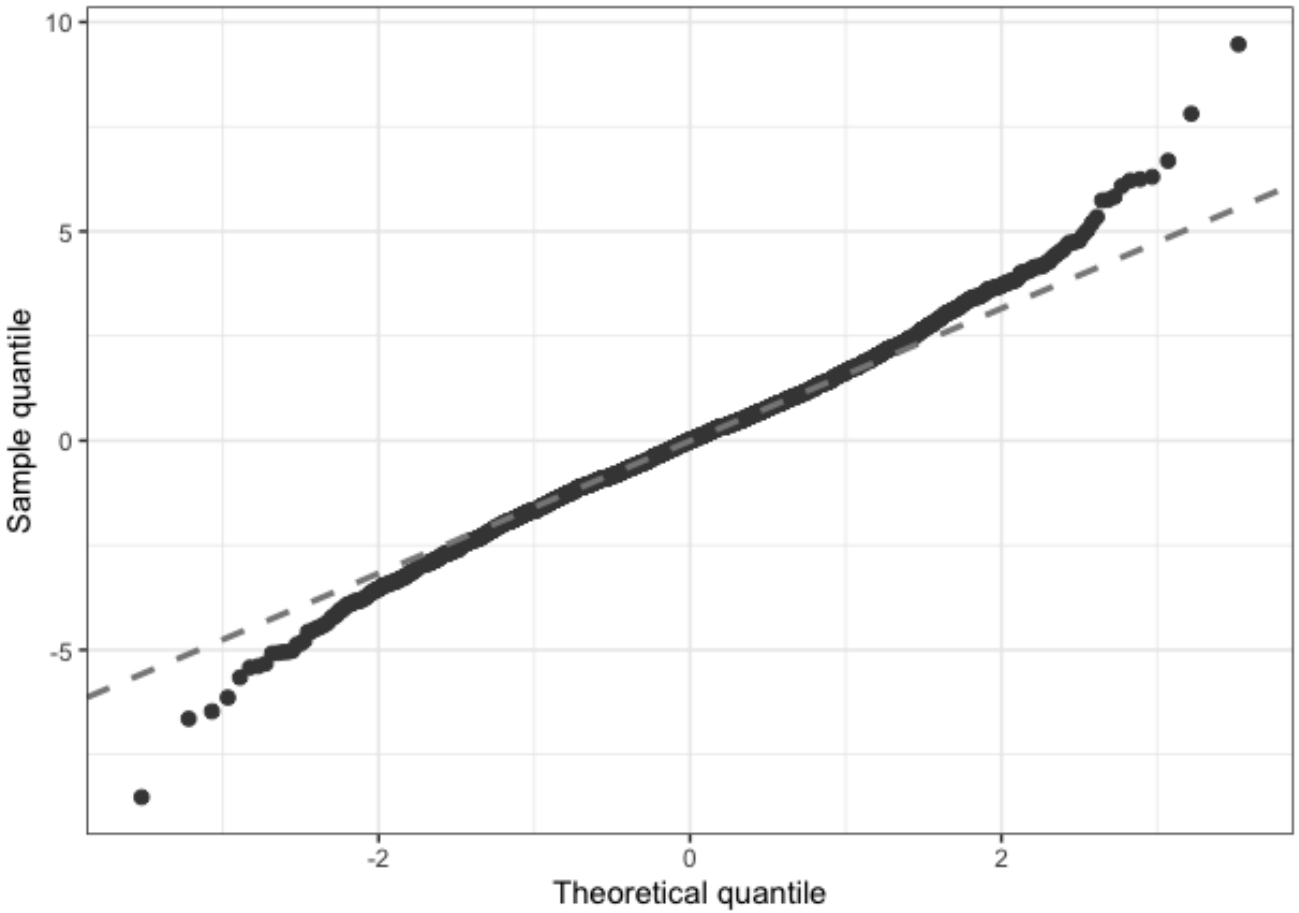

Figure: QQ plot of the Surrogate residuals

Appendix

Box-Cox transformation of lipid

Following Li, Longnecker and Dunson (2013), we adjust the exposures by dividing for a Box-Cox tranformation of the level of lipids. We see that the op-