Case Study 1: National Collaborative Perinatal Project

Background

The data are taken from the National Collaborative Perinatal Project (CPP). Women were enrolled during pregnancy through different medical centers and then the kids were followed in order to collect both pregnancy and childhood development outcomes. We consider a subsample of 2380 women and children for this analysis, which was studied by [Longnecker et al., 2001]. A particular focus of the Longnecker et al substudy was in assaying serum samples from the original larger study to obtain information on exposures in order to assess the relationship between these exposures to the women and adverse pregnancy and developmental outcomes in their children. Two exposures of particular interest are Dichlorodiphenyldichloroethylene (DDE) and Polychlorinated Biphenyls (PCBs), which are breakdown products in the body of chemicals that have been historically used to treat crops to protect them from predation. These chemicals persist in the environment and are lipophilic, building up in fatty deposits in human tissues. Hence, each of us carries around our own body burden of these chemicals, potentially impacting our health.

The data

The dataset contains demographic variables, such as race, age, and socio-economic index, along with smoking status and concentration doses for DDE and PCBs. In addition, data are available on levels of cholesterol and triglycerides in serum; these variables are relevant since DDE/PCBs are stored in fat and cholesterol/triglycerides provide measurements of the levels of circulating fats (being somewhat informal) in serum.

Goal

The overarching goal of the analysis is to assess how DDE and PCBs relate to risk of premature delivery. Premature delivery is typically defined as a gestational age at delivery of 37 weeks or less, but it is important to note that deliveries occurring right at the cutoff have similar clinical outcomes to full term deliveries, while deliveries occurring substantially less than 37 weeks (early preterm) are associated with substantial risk of short and long term morbidity and mortality. Ideally we would like to infer a causal effect of these exposures on risk of premature deliveries of different severities, while investigating the dose response relationship. However, these data are not collected in a randomized trial but are the result of an observational epidemiology study. Hence, epidemiologists typically focus on assessing associations, while adjusting for covariates that may confound exposure-outcome relationships. In addressing the above interests, it is important to take into account heterogeneity across study centers.

Variable key

gestational_age = gestational age (in weeks)

dde = concentration of dde (ug/dL)

pcb_* = concentration of pcb_* (ng/dL)

albumin = concentration of albumin (g/dL)

cholesterol = concentration of cholesterol (g/dL)

triglycerides = concentration of triglycerides (g/dL)

race

score_education

score_income

score_occupation

maternal_age = age of mother

smoking_status = mother smoking

center

```
# Load in data & remove data point with missing PCB information
dat <- readRDS("Longnecker.rds")
dat <- dat[-which(is.na(dat$pcb_028)==TRUE),]
dat$race <- as.factor(dat$race)
dat$smoking_status <- as.factor(dat$smoking_status)
dat$center <- as.factor(dat$center)
dat0 <- dat # store original data
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Indicator for preterm
dat <- dat %>%
       mutate(preterm_ind = if_else(gestational_age<33,2,
                                 if_else(gestational_age<37 &
                                             gestational_age>32,1,0)))
dat$preterm_ind <- ordered(dat$preterm_ind)
```

```
# Data Imputation (MICE - Predictive Mean Matching)
library(mice)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'mice'
```

```
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```
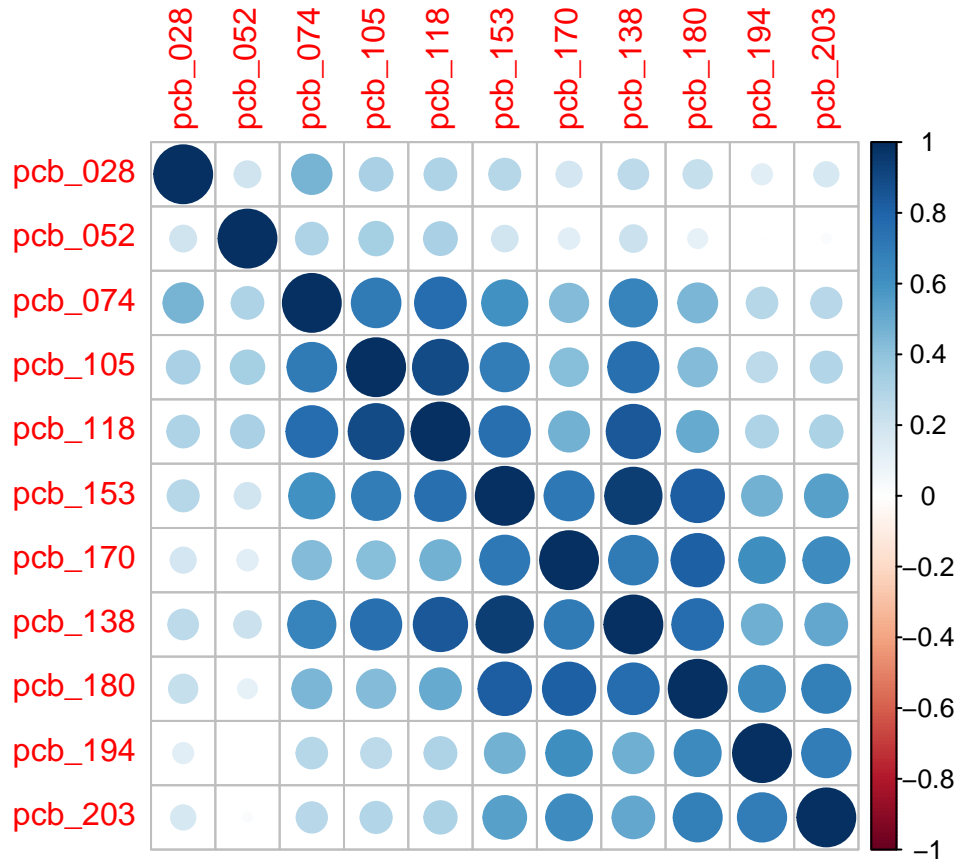
```
# apply(is.na(dat), 2, sum)
# remove albumin; impute score_education, score_income, score_occupation
dat_mice <- mice(dat[,-13], seed = 12345, printFlag = FALSE)
imp_dat <- complete(dat_mice)
```

```r
pcb_col <- grep("pcb", names(imp_dat))
orig_pcb <- imp_dat[,pcb_col]

library(corrplot)
```

```
## corrplot 0.84 loaded
```

```r
corrplot(cor(orig_pcb))
```



```r
# Check for multicolinearity
library(car)
```

```
## Loading required package: carData
```

```
## Registered S3 methods overwritten by 'car':
##   method                            from
##   influence.merMod                  lme4
##   cooks.distance.influence.merMod   lme4
##   dfbeta.influence.merMod           lme4
##   dfbetas.influence.merMod          lme4
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
vif(lm(gestational_age ~ dde + pcb_028 + pcb_052 + pcb_074 + pcb_105 + pcb_118 + pcb_153 + pcb_170 + pcl
```

```
##      dde   pcb_028   pcb_052   pcb_074   pcb_105   pcb_118   pcb_153   pcb_170
##  1.254391  1.345643  1.177619  2.933636  5.592635 11.086467 13.635187  3.391085
##   pcb_138   pcb_180   pcb_194   pcb_203
## 17.715106  5.991696  2.290034  2.468200
```

```r
# Combine pcb_118, pcb_153, pcb_138 into one pcb_sum3
imp_dat <- imp_dat %>%
        mutate(pcb_sum3 = pcb_118 + pcb_153 + pcb_138)

vif(lm(gestational_age ~ dde + pcb_028 + pcb_052 + pcb_074 + pcb_105 + pcb_170 + pcb_180 + pcb_194 + pcl
```

```
##      dde  pcb_028  pcb_052  pcb_074  pcb_105  pcb_170  pcb_180  pcb_194
## 1.170381 1.324664 1.166064 2.732240 4.435526 3.321836 4.920617 2.223868
##  pcb_203 pcb_sum3
## 2.461198 7.703803
```

```r
# all vif's < 10
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
ord_model <- MASS::polr(preterm_ind ~ dde + pcb_028 + pcb_052 + pcb_074 + pcb_105 + pcb_170 + pcb_180 +

summary(ord_model)
```

```
## Call:
## MASS::polr(formula = preterm_ind ~ dde + pcb_028 + pcb_052 +
##     pcb_074 + pcb_105 + pcb_170 + pcb_180 + pcb_194 + pcb_203 +
##     pcb_sum3 + triglycerides + race + score_education + score_income +
##     score_occupation + maternal_age + smoking_status + cholesterol +
##     center, data = imp_dat, Hess = TRUE, method = "logistic")
##
## Coefficients:
##                   Value Std. Error  t value
## dde            0.007952  0.0029821  2.66641
## pcb_028       -0.061710  0.4707846 -0.13108
## pcb_052        0.322277  0.7047924  0.45726
## pcb_074        0.693482  0.4325690  1.60317
```

4

```
## pcb_105          -0.904956  0.8938970 -1.01237
## pcb_170          -1.732233  1.0584996 -1.63650
## pcb_180           0.599487  0.7274387  0.82411
## pcb_194          -0.041209  0.6460882 -0.06378
## pcb_203           0.855138  0.8427519  1.01470
## pcb_sum3          0.167320  0.1328918  1.25907
## triglycerides     0.003014  0.0007512  4.01289
## raceblack         0.244730  0.2132808  1.14745
## raceother         0.453063  0.3549704  1.27634
## score_education  -0.003168  0.0027250 -1.16273
## score_income     -0.001701  0.0024169 -0.70372
## score_occupation -0.003262  0.0024274 -1.34393
## maternal_age     -0.014091  0.0103294 -1.36415
## smoking_status1   0.127490  0.1237035  1.03061
## cholesterol      -0.002451  0.0010008 -2.44915
## center10         -0.993575  0.4955247 -2.00510
## center15          0.813018  0.3256701  2.49645
## center31         -0.588773  0.4879523 -1.20662
## center37          0.776721  0.2766705  2.80739
## center45          0.159297  0.3198840  0.49798
## center50          0.075756  0.3417471  0.22167
## center55          0.351418  0.3562212  0.98652
## center60          0.419574  0.3169149  1.32393
## center66          0.213894  0.2727173  0.78431
## center71         -0.010925  0.3146670 -0.03472
## center82          0.584635  0.3222990  1.81395
##
## Intercepts:
##     Value   Std. Error t value
## 0|1  2.2812  0.4659     4.8966
## 1|2  4.1354  0.4801     8.6141
##
## Residual Deviance: 2245.893
## AIC: 2309.893
```

```
confint(ord_model)
```

```
## Waiting for profiling to be done...
```

```
##                        2.5 %        97.5 %
## dde             0.002007329  0.0137655123
## pcb_028        -1.019695726  0.8333353211
## pcb_052        -1.205562938  1.6305068484
## pcb_074        -0.182384979  1.5287499012
## pcb_105        -2.667921547  0.8380678488
## pcb_170        -3.822559305  0.3805953823
## pcb_180        -0.834437859  2.0302733222
## pcb_194        -2.890251764  2.7561422971
## pcb_203        -1.751858711  3.4391270930
## pcb_sum3       -0.096856301  0.4260980038
## triglycerides   0.001535227  0.0044686776
## raceblack      -0.172286821  0.6639886305
## raceother      -0.254849791  1.1391761711
```

```
## score_education  -0.008513955  0.0021753086
## score_income     -0.006451759  0.0030292512
## score_occupation -0.008043357  0.0014795557
## maternal_age      -0.034628347  0.0060420605
## smoking_status1   -0.115540026  0.3696127496
## cholesterol       -0.004451355 -0.0005012654
## center10          -2.089912150 -0.1108361699
## center15           0.172977998  1.4549758461
## center31          -1.636907586  0.3025196124
## center37           0.230839125  1.3229655931
## center45          -0.476133028  0.7802569608
## center50          -0.632554827  0.7221234709
## center55          -0.356893869  1.0456717659
## center60          -0.222230103  1.0284298259
## center66          -0.318833289  0.7510774611
## center71          -0.652153805  0.5889997110
## center82          -0.050541628  1.2195537111
```

```r
library(car)
Anova(ord_model)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: preterm_ind
##                 LR Chisq Df Pr(>Chisq)
## dde               6.8064  1   0.009083 **
## pcb_028           0.0174  1   0.895046
## pcb_052           0.1985  1   0.655961
## pcb_074           2.4357  1   0.118602
## pcb_105           1.0315  1   0.309802
## pcb_170           2.5884  1   0.107648
## pcb_180           0.6753  1   0.411201
## pcb_194           0.0007  1   0.978897
## pcb_203           0.4152  1   0.519353
## pcb_sum3          1.5532  1   0.212667
## triglycerides    15.3285  1  9.034e-05 ***
## race              2.0762  2   0.354134
## score_education   1.3538  1   0.244617
## score_income      0.4959  1   0.481309
## score_occupation  1.8145  1   0.177969
## maternal_age      1.8741  1   0.171004
## smoking_status    1.0584  1   0.303585
## cholesterol       6.1064  1   0.013470 *
## center           28.5341 11   0.002680 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
poTest(ord_model)
```

```
##
## Tests for Proportional Odds
## MASS::polr(formula = preterm_ind ~ dde + pcb_028 + pcb_052 +
##     pcb_074 + pcb_105 + pcb_170 + pcb_180 + pcb_194 + pcb_203 +
```

```
##       pcb_sum3 + triglycerides + race + score_education + score_income +
##       score_occupation + maternal_age + smoking_status + cholesterol +
##       center, data = imp_dat, Hess = TRUE, method = "logistic")
##
##                       b[polr]      b[>0]      b[>1] Chisquare df Pr(>Chisq)
## Overall                                             16.53 30      0.978
## dde                  7.95e-03   8.36e-03  -4.34e-05   1.78  1      0.182
## pcb_028             -6.17e-02  -8.42e-02   1.60e-01   0.08  1      0.776
## pcb_052              3.22e-01   3.79e-01   3.44e-01   0.00  1      0.980
## pcb_074              6.93e-01   5.89e-01   1.05e+00   0.38  1      0.537
## pcb_105             -9.05e-01  -6.69e-01  -3.16e+00   2.14  1      0.144
## pcb_170             -1.73e+00  -1.74e+00  -2.48e+00   0.13  1      0.722
## pcb_180              5.99e-01   7.25e-01   2.85e-01   0.10  1      0.747
## pcb_194             -4.12e-02  -1.27e-01   1.27e-02   0.00  1      0.959
## pcb_203              8.55e-01   8.56e-01   2.19e+00   0.29  1      0.589
## pcb_sum3             1.67e-01   1.39e-01   4.21e-01   1.40  1      0.237
## triglycerides        3.01e-03   3.17e-03   2.36e-03   0.27  1      0.603
## raceblack            2.45e-01   2.19e-01   1.01e+00   3.36  1      0.067 .
## raceother            4.53e-01   4.22e-01   1.77e+00   4.15  1      0.042 *
## score_education     -3.17e-03  -3.30e-03  -3.15e-03   0.00  1      0.978
## score_income        -1.70e-03  -1.43e-03  -6.38e-03   1.07  1      0.301
## score_occupation    -3.26e-03  -3.11e-03  -1.53e-03   0.11  1      0.736
## maternal_age        -1.41e-02  -1.50e-02  -1.45e-02   0.00  1      0.982
## smoking_status1      1.27e-01   1.35e-01   1.82e-03   0.31  1      0.580
## cholesterol         -2.45e-03  -2.51e-03  -3.44e-03   0.20  1      0.652
## center10            -9.94e-01  -9.79e-01  -1.57e+01   0.00  1      0.992
## center15             8.13e-01   8.17e-01   7.57e-03   1.93  1      0.164
## center31            -5.89e-01  -5.63e-01  -1.67e+01   0.00  1      0.993
## center37             7.77e-01   7.93e-01  -1.14e-01   2.85  1      0.091 .
## center45             1.59e-01   1.76e-01  -1.09e+00   3.11  1      0.078 .
## center50             7.58e-02   6.77e-02  -1.73e-01   0.11  1      0.746
## center55             3.51e-01   3.64e-01  -9.15e-01   3.24  1      0.072 .
## center60             4.20e-01   3.84e-01   4.05e-01   0.00  1      0.970
## center66             2.14e-01   2.35e-01  -7.48e-01   3.84  1      0.050 .
## center71            -1.09e-02  -2.51e-02  -4.50e-01   0.45  1      0.504
## center82             5.85e-01   5.79e-01  -1.37e-01   1.50  1      0.221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
library(effects)
```

```
## Use the command
##     lattice::trellis.par.set(effectsTheme())
##   to customize lattice options for effects plots.
## See ?effectsTheme for details.
```
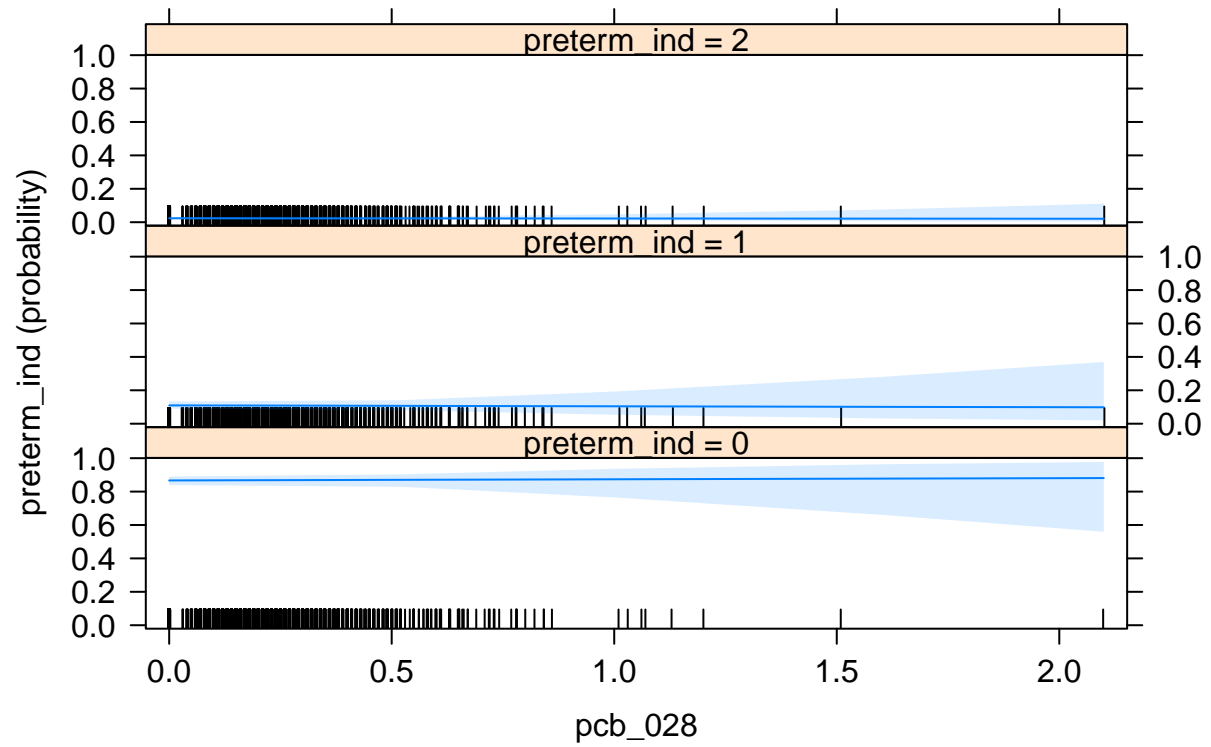
```r
par(mfrow=c(1,2))
plot(Effect("dde", ord_model))
```
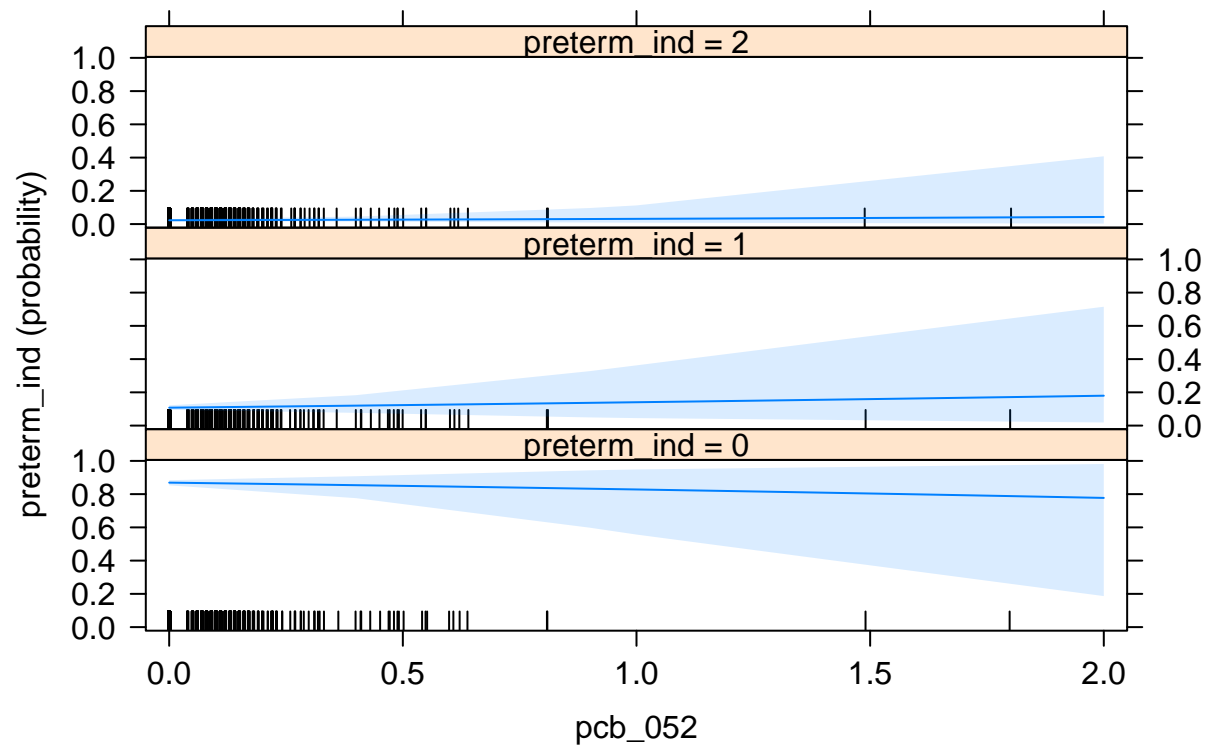
dde effect plot
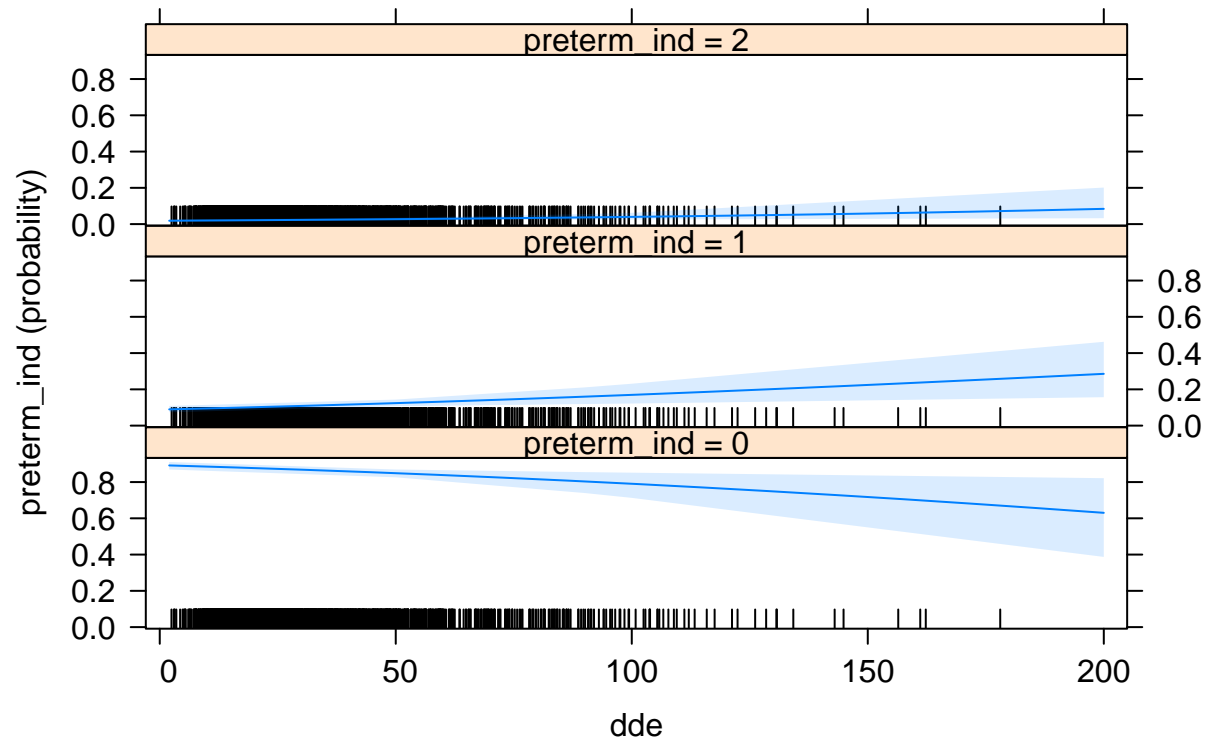

```r
plot(Effect("pcb_028", ord_model))
```
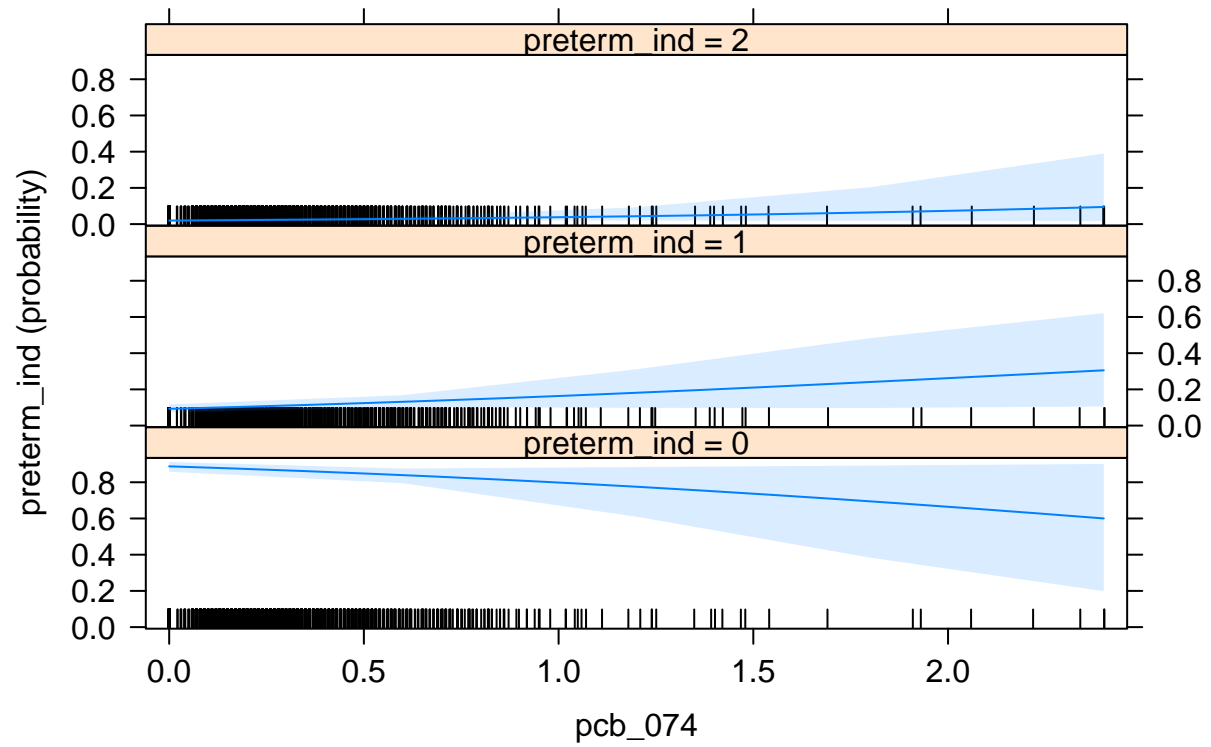
pcb_028 effect plot

```
plot(Effect("pcb_052", ord_model))
```
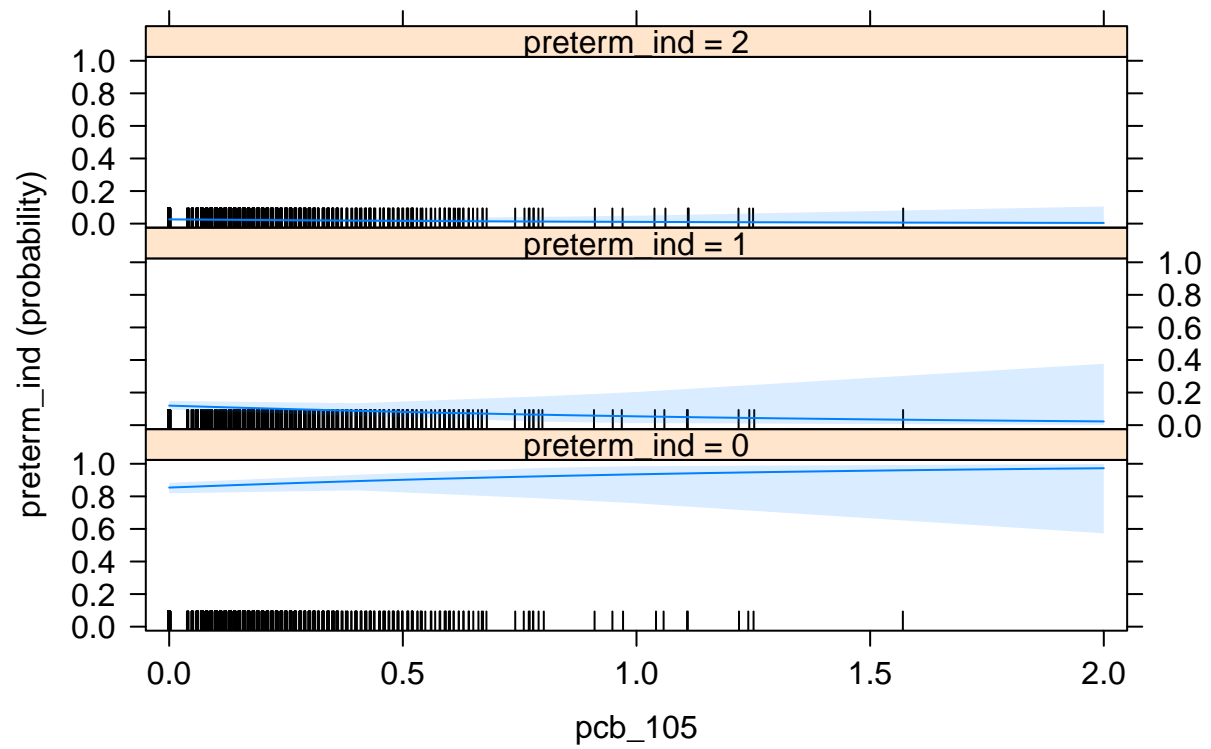
pcb_052 effect plot


```
plot(Effect("pcb_074", ord_model))
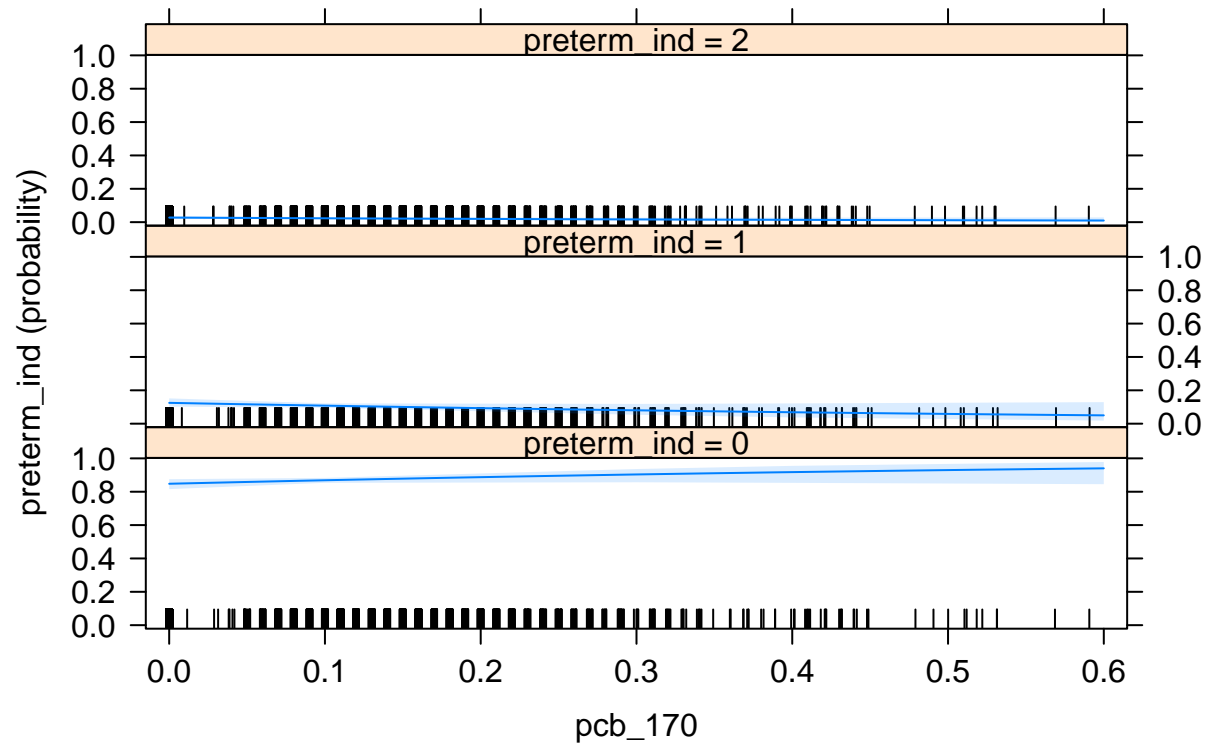```

**pcb_074 effect plot**



```r
plot(Effect("pcb_105", ord_model))
```
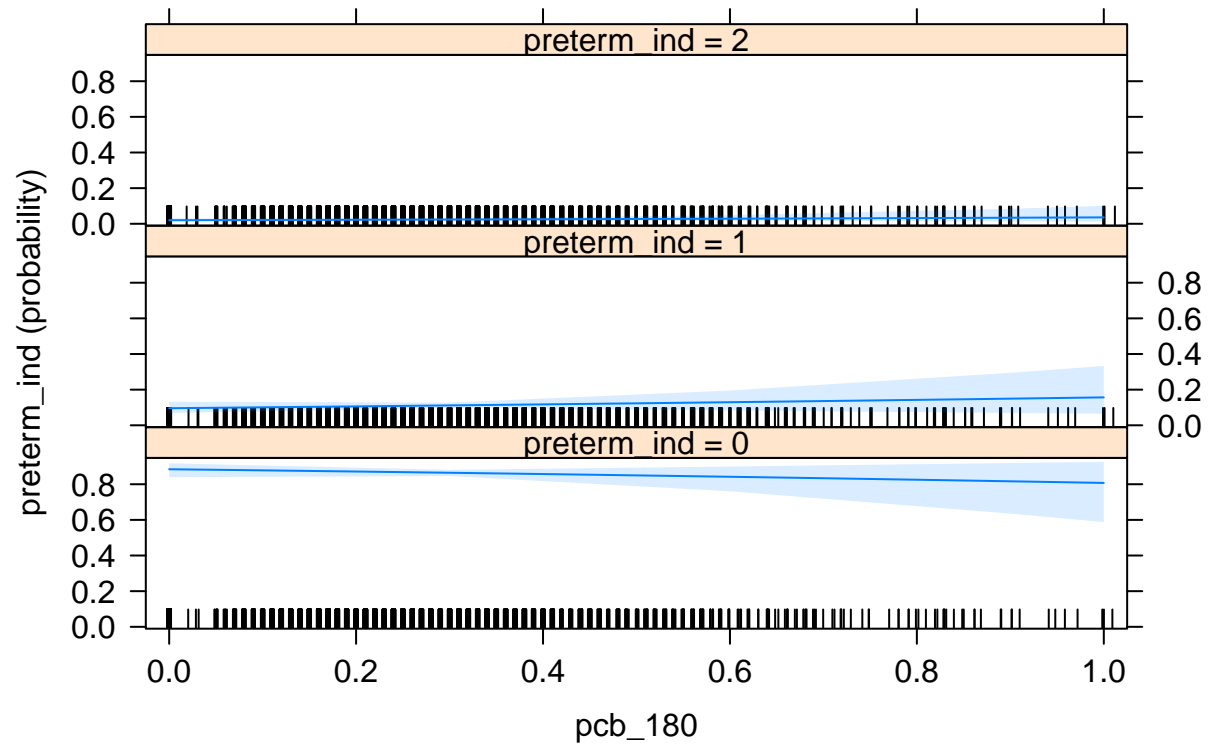
pcb_105 effect plot

```
plot(Effect("pcb_170", ord_model))
```
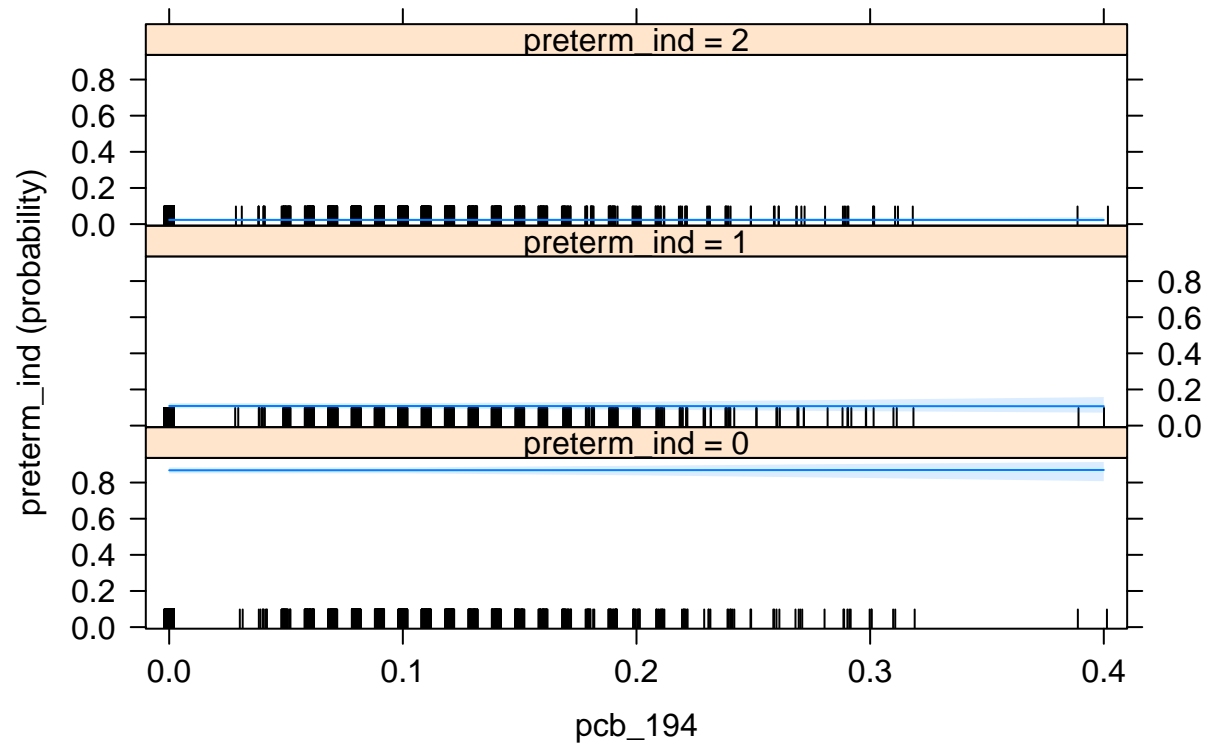
pcb_170 effect plot


```
plot(Effect("pcb_180", ord_model))
```
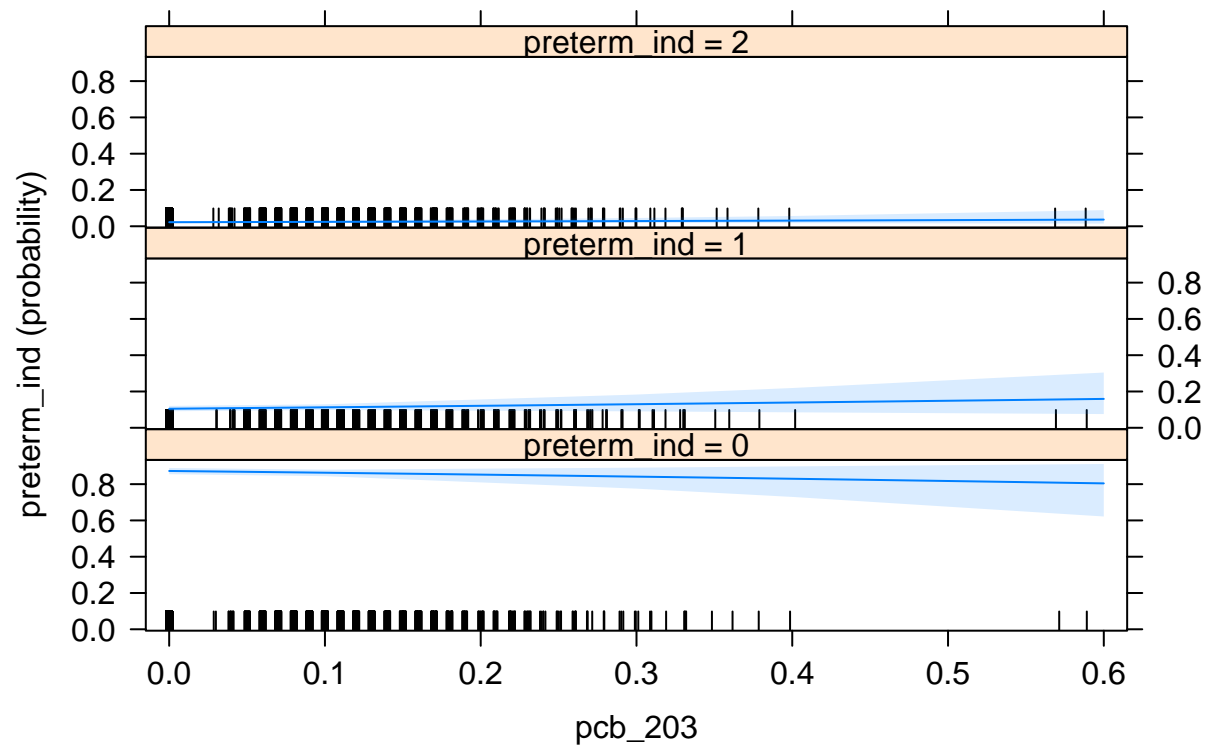
pcb_180 effect plot


```
plot(Effect("pcb_194", ord_model))
```

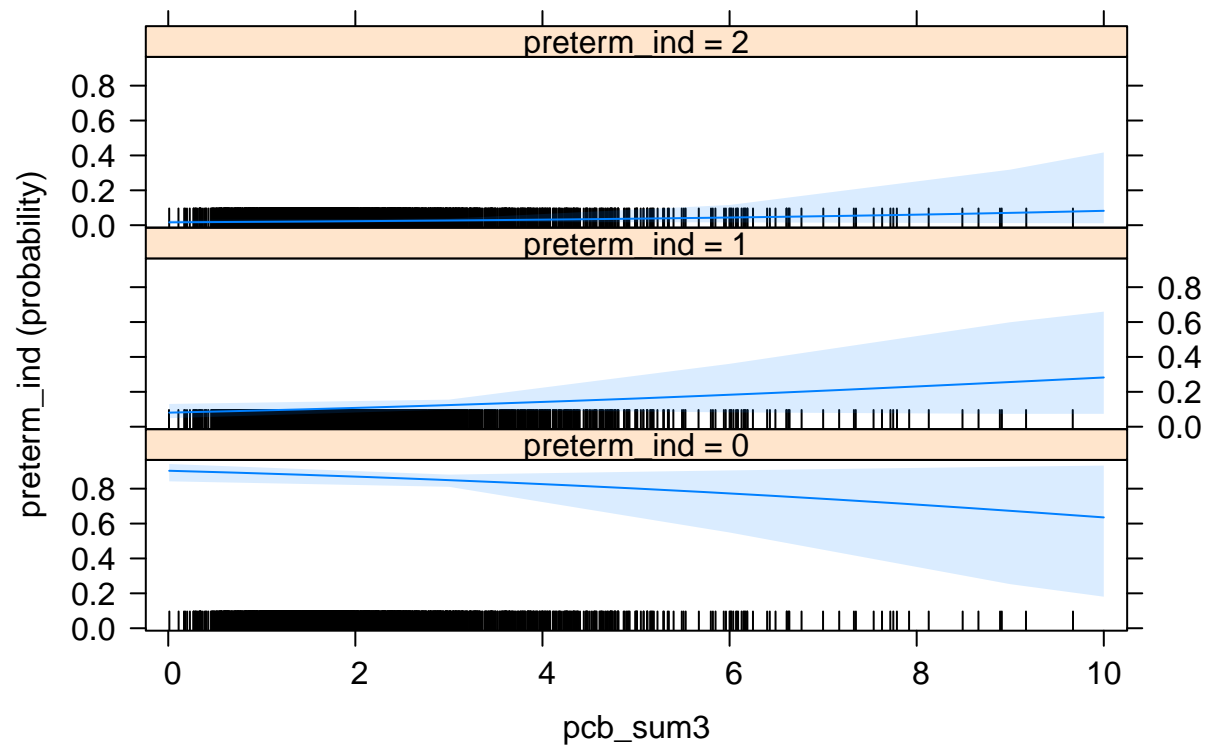**pcb_194 effect plot**

```
plot(Effect("pcb_203", ord_model))
```

pcb_203 effect plot


```r
plot(Effect("pcb_sum3", ord_model))
```

pcb_sum3 effect plot


```
# plot(Effect("pcb_sum3", ord_model), style='stacked', key.args=list(x=.55, y=.9))
```