Supplemental Information for

Increasing Sample Size in Prospective Birth Cohorts: Back-Extrapolating Prenatal Levels of Persistent Organic Pollutants in Newly Enrolled Children

Marc-André Verner1,2\*, Fraser W. Gaspar3, Jonathan Chevrier3,4, Robert B. Gunier3, Andreas Sjödin5, Asa Bradman3, Brenda Eskenazi3

1Department of Occupational and Environmental Health, School of Public Health, Université de Montréal, Montreal, Canada

2Université de Montréal Public Health Research Institute (IRSPUM), Université de Montréal, Montreal, Canada

3Center for Environmental Research and Children’s Health (CERCH), School of Public Health, University of California, Berkeley, Berkeley, CA, USA.

4Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montreal, QC, Canada.

5Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, USA.

**TABLE OF CONTENTS**

[Table S1. Cross-validated algorithm MSEs (weights) in the Super Learner by subset S4](#_Toc408235692)

[Table S2. DSA prediction equations for p,p’-DDT/E and PBDEs using child 9-year serum levels (n=161) S8](#_Toc408235693)

[Table S3. DSA prediction equations for *p,p’*-DDT/E and PBDEs using both maternal and child 9-year serum levels (n=89) S11](#_Toc408235694)

[Figure S1. Scatterplots of back-extrapolated versus measured prenatal (~26 weeks gestation) serum concentrations of *p,p’*-DDT/E and PBDEs using 9-year child serum levels (n=161). Root mean squared errors (RMSEs) and correlation coefficients (R2s) were calculated comparing the linear fit of back-extrapolated and measured log10 serum levels (red line). Black line represents perfect back-extrapolation. SL = Super Learner, DSA = Deletion/ Substitution/ Addition, and PK = pharmacokinetic model. S13](#_Toc408235695)

[Figure S2. Scatterplots of back-extrapolated versus measured prenatal (~26 weeks gestation) serum concentrations of *p,p’*-DDT/E and PBDEs using both 9-year maternal and child serum levels (n=89). Root mean squared errors (RMSEs) and correlation coefficients (R2s) were calculated comparing the linear fit of back-extrapolated and measured log10 serum levels (red line). Black line represents perfect back-extrapolation. SL = Super Learner and DSA = Deletion/ Substitution/ Addition. S14](#_Toc408235696)

Table S1. Cross-validated algorithm MSEs (weights) in the Super Learner by subset

|  |
| --- |
| **Maternal 9-year subset (n=94)** |
| **Method** | ***p,p’*-DDT** | ***p,p’*-DDE** | **BDE-47** | **BDE-99** | **BDE-100** | **BDE-153** |
| GLM | 0.090 (0.000) | 0.046 (0.000) | 0.123 (0.054) | 0.220 (0.000) | 0.111 (0.210) | 0.067 (0.275) |
| sGLM | 0.083 (0.000) | 0.050 (0.000) | 0.094 (0.000) | 0.151 (0.000) | 0.092 (0.137) | 0.069 (0.000) |
| GAM | 0.099 (0.198) | 0.048 (0.000) | 0.134 (0.000) | 0.258 (0.000) | 0.118 (0.000) | 0.076 (0.000) |
| sGAM | 0.076 (0.319) | 0.052 (0.000) | 0.094 (0.406) | 0.178 (0.000) | 0.090 (0.273) | 0.070 (0.106) |
| BayesGLM | 0.083 (0.000) | 0.046 (0.079) | 0.122 (0.000) | 0.214 (0.000) | 0.107 (0.000) | 0.066 (0.000) |
| sBayesGLM | 0.082 (0.000) | 0.050 (0.000) | 0.094 (0.000) | 0.149 (0.428) | 0.092 (0.000) | 0.069 (0.030) |
| GLMNET | 0.074 (0.000) | 0.037 (0.261) | 0.100 (0.000) | 0.156 (0.000) | 0.109 (0.000) | 0.058 (0.259) |
| RandomForest | 0.197 (0.160) | 0.086 (0.000) | 0.118 (0.000) | 0.171 (0.000) | 0.119 (0.021) | 0.100 (0.000) |
| LOESS | 0.128 (0.000) | 0.052 (0.006) | 0.165 (0.039) | 0.218 (0.000) | 0.178 (0.057) | 0.168 (0.000) |
| RPART | 0.166 (0.073) | 0.071 (0.277) | 0.160 (0.000) | 0.218 (0.000) | 0.148 (0.000) | 0.099 (0.008) |
| NNET | 0.567 (0.000) | 0.288 (0.000) | 0.195 (0.052) | 0.242 (0.000) | 0.194 (0.000) | 0.157 (0.005) |
| Polymars | 0.120 (0.048) | 0.050 (0.198) | 0.114 (0.163) | 0.295 (0.000) | 0.120 (0.162) | 0.087 (0.120) |
| SVM | 0.294 (0.000) | 0.103 (0.000) | 0.143 (0.000) | 0.185 (0.028) | 0.130 (0.059) | 0.113 (0.000) |
| DSA | 0.107 (0.201) | 0.046 (0.179) | 0.098 (0.285) | 0.142 (0.544) | 0.109 (0.080) | 0.073 (0.198) |

**Table S1 Continued.** Cross-validated algorithm MSEs (weights) in the Super Learner by subset

|  |
| --- |
| **Child 9-year subset (n=161)** |
| **Method** | ***p,p’*-DDT** | ***p,p’*-DDE** | **BDE-47** | **BDE-99** | **BDE-100** | **BDE-153** |
| GLM | 0.171 (0.181) | 0.102 (0.000) | 0.250 (0.000) | 0.280 (0.000) | 0.254 (0.000) | 0.146 (0.307) |
| sGLM | 0.173 (0.000) | 0.120 (0.000) | 0.213 (0.000) | 0.237 (0.000) | 0.207 (0.000) | 0.149 (0.000) |
| GAM | 0.178 (0.031) | 0.098 (0.543) | 0.240 (0.000) | 0.274 (0.000) | 0.242 (0.000) | 0.152 (0.000) |
| sGAM | 0.158 (0.314) | 0.112 (0.051) | 0.216 (0.000) | 0.241 (0.000) | 0.202 (0.000) | 0.154 (0.000) |
| BayesGLM | 0.171 (0.000) | 0.102 (0.000) | 0.248 (0.000) | 0.278 (0.000) | 0.251 (0.000) | 0.146 (0.000) |
| sBayesGLM | 0.173 (0.000) | 0.120 (0.000) | 0.213 (0.000) | 0.236 (0.000) | 0.207 (0.000) | 0.148 (0.000) |
| GLMNET | 0.160 (0.000) | 0.100 (0.000) | 0.213 (0.000) | 0.214 (0.000) | 0.181 (0.000) | 0.136 (0.000) |
| RandomForest | 0.190 (0.216) | 0.132 (0.000) | 0.170 (0.912) | 0.205 (0.715) | 0.164 (0.622) | 0.131 (0.549) |
| LOESS | 0.215 (0.000) | 0.132 (0.158) | 0.204 (0.000) | 0.235 (0.000) | 0.197 (0.000) | 0.167 (0.000) |
| RPART | 0.239 (0.134) | 0.147 (0.030) | 0.250 (0.000) | 0.283 (0.000) | 0.225 (0.086) | 0.197 (0.000) |
| NNET | 0.667 (0.000) | 0.359 (0.000) | 0.205 (0.088) | 0.212 (0.285) | 0.206 (0.000) | 0.176 (0.000) |
| Polymars | 0.236 (0.115) | 0.137 (0.117) | 0.439 (0.000) | 1.107 (0.000) | 0.387 (0.000) | 0.185 (0.119) |
| SVM | 0.232 (0.010) | 0.132 (0.101) | 0.181 (0.000) | 0.220 (0.000) | 0.183 (0.000) | 0.135 (0.000) |
| DSA | 0.185 (0.000) | 0.128 (0.000) | 0.195 (0.000) | 0.225 (0.000) | 0.171 (0.292) | 0.146 (0.026) |

**Table S1 Continued.** Cross-validated algorithm MSEs (weights) in the Super Learner by subset

|  |
| --- |
| **Maternal and Child 9-year subset (n=89)** |
| **Method** | ***p,p’*-DDT** | ***p,p’*-DDE** | **BDE-47** | **BDE-99** | **BDE-100** | **BDE-153** |
| GLM | 0.116 (0.000) | 0.047 (0.000) | 0.184 (0.000) | 0.282 (0.000) | 0.162 (0.000) | 0.069 (0.000) |
| sGLM | 0.079 (0.000) | 0.043 (0.000) | 0.097 (0.000) | 0.139 (0.425) | 0.087 (0.281) | 0.074 (0.000) |
| GAM | 0.102 (0.020) | 0.040 (0.000) | 0.203 (0.000) | 0.323 (0.000) | 0.146 (0.076) | 0.090 (0.000) |
| sGAM | 0.066 (0.636) | 0.037 (0.014) | 0.096 (0.447) | 0.155 (0.000) | 0.086 (0.000) | 0.071 (0.114) |
| BayesGLM | 0.106 (0.000) | 0.045 (0.000) | 0.170 (0.000) | 0.227 (0.000) | 0.151 (0.000) | 0.065 (0.318) |
| sBayesGLM | 0.078 (0.000) | 0.043 (0.000) | 0.097 (0.108) | 0.138 (0.263) | 0.087 (0.000) | 0.073 (0.000) |
| GLMNET | 0.083 (0.000) | 0.031 (0.000) | 0.104 (0.000) | 0.167 (0.000) | 0.094 (0.000) | 0.058 (0.000) |
| RandomForest | 0.126 (0.038) | 0.066 (0.000) | 0.125 (0.000) | 0.175 (0.067) | 0.128 (0.000) | 0.107 (0.000) |
| LOESS | 0.191 (0.000) | 0.102 (0.020) | 0.234 (0.000) | 0.283 (0.000) | 0.238 (0.000) | 0.205 (0.000) |
| RPART | 0.153 (0.152) | 0.075 (0.138) | 0.186 (0.000) | 0.212 (0.000) | 0.148 (0.073) | 0.098 (0.044) |
| NNET | 0.608 (0.000) | 0.290 (0.000) | 0.208 (0.042) | 0.247 (0.000) | 0.202 (0.040) | 0.164 (0.033) |
| Polymars | 0.108 (0.154) | 0.031 (0.269) | 0.187 (0.000) | 0.300 (0.000) | 0.086 (0.355) | 0.059 (0.285) |
| SVM | 0.295 (0.000) | 0.100 (0.070) | 0.138 (0.000) | 0.190 (0.057) | 0.144 (0.000) | 0.108 (0.000) |
| DSA | 0.098 (0.000) | 0.028 (0.489) | 0.101 (0.403) | 0.159 (0.188) | 0.093 (0.175) | 0.063 (0.207) |
| GLM= generalized linear modelsGLM= “screened” GLM. Variables only selected if significantly correlated with prenatal serum levels in bivariate analysis (p≤0.1)GAM= generalized additive modelsGAM= “screened” GAM. Variables only selected if significantly correlated with prenatal serum levels in bivariate analysis (p≤0.1)BayesGLM= Bayesian linear modelsBayesGLM= “screened” BayesGLM. Variables only selected if significantly correlated with prenatal serum levels in bivariate analysis (p≤0.1)GLMNET= elastic netLOESS= local polynomial regressionRPART= recursive partitioning and regression treesNNET= neural networkPolyMars= polynomial spline regressionSVM= support vector machineDSA= deletion/substitution/addition |

Table S2. DSA prediction equations for p,p’-DDT/E and PBDEs using child 9-year serum levels (n=161)

|  |  |
| --- | --- |
| **Equation** | **Description** |
|  |  = maternal *p,p’*-DDT serum concentration at 26 weeks gestation (ng/g-lipid) = index child *p,p’*-DDT serum concentration at 9-year visit (ng/g-lipid) = index child *p,p’*-DDE serum concentration at 9-year visit (ng/g-lipid) = index child BMI z-score at 9-year visit (continuous) = years in the United States prior to index child (years) |
|  |  = maternal *p,p’*-DDE serum concentration at 26 weeks gestation (ng/g-lipid) = index child *p,p’*-DDE serum concentration at 9-year visit (ng/g-lipid) = index child BMI z-score at 9-year visit (continuous) = maternal parity before index child (#)= breastfeeding duration of index child (months) = maternal age at delivery (years) = maternal pregnancy weight gain (kg) |
|  |  = maternal BDE-47 serum concentration at 26 weeks gestation (ng/g-lipid) |
|  |  = maternal BDE-99 serum concentration at 26 weeks gestation (ng/g-lipid) |
|  |  = maternal BDE-100 serum concentration at 26 weeks gestation (ng/g-lipid) = years in the United States prior to index child (years) |
|  |  = maternal BDE-153 serum concentration at 26 weeks gestation (ng/g-lipid)= index child BDE-153 serum concentration at 9-year visit (ng/g-lipid) = years in the United States prior to index child (years) |

Table S3. DSA prediction equations for *p,p’*-DDT/E and PBDEs using both maternal and child 9-year serum levels (n=89)

|  |  |
| --- | --- |
| **Equation** | **Description** |
|  |  = maternal *p,p’*-DDT serum concentration at 26 weeks gestation (ng/g-lipid) = maternal *p,p’*-DDT serum concentration at 9-year visit (ng/g-lipid) = index child *p,p’*-DDE serum concentration at 9-year visit (ng/g-lipid) |
|  |  = maternal *p,p’*-DDE serum concentration at 26 weeks gestation (ng/g-lipid) = maternal *p,p’*-DDE serum concentration at 9-year visit (ng/g-lipid)= index child *p,p’*-DDE serum concentration at 9-year visit (ng/g-lipid)= maternal pre-pregnancy weight (kg)= cumulative breastfeeding duration post-index child (months) = child fat percentage at 9-year visit (%)= maternal weight at 9-year visit (kg) = maternal pregnancy weight gain (kg) |
|  |  = maternal BDE-47 serum concentration at 26 weeks gestation (ng/g-lipid) = maternal BDE-153 serum concentration at 9-year visit (ng/g-lipid) = maternal BDE-100 serum concentration at 9-year visit (ng/g-lipid) = poverty at 9-year visit (categorical) = years in the United States prior to index child (years) |
|  |  = maternal BDE-99 serum concentration at 26 weeks gestation (ng/g-lipid) = maternal BDE-100 serum concentration at 9-year visit (ng/g-lipid) = maternal BDE-153 serum concentration at 9-year visit (ng/g-lipid) |
|  |  = maternal BDE-100 serum concentration at 26 weeks gestation (ng/g-lipid) = maternal BDE-100 serum concentration at 9-year visit (ng/g-lipid) = maternal BDE-153 serum concentration at 9-year visit (ng/g-lipid) = maternal BDE-99 serum concentration at 9-year visit (ng/g-lipid) = years in the United States prior to index child (years) = Upholstery furniture present in home at 9-year visit (yes/no) |
|  |  = maternal BDE-153 serum concentration at 26 weeks gestation (ng/g-lipid) = maternal BDE-153 serum concentration at 9-year visit (ng/g-lipid) = years in the United States prior to index child (years) |



Figure S1. Scatterplots of back-extrapolated versus measured prenatal (~26 weeks gestation) serum concentrations of *p,p’*-DDT/E and PBDEs using 9-year child serum levels (n=161). Root mean squared errors (RMSEs) and correlation coefficients (R2s) were calculated comparing the linear fit of back-extrapolated and measured log10 serum levels (red line). Black line represents perfect back-extrapolation. SL = Super Learner, DSA = Deletion/ Substitution/ Addition, and PK = pharmacokinetic model.



Figure S2. Scatterplots of back-extrapolated versus measured prenatal (~26 weeks gestation) serum concentrations of *p,p’*-DDT/E and PBDEs using both 9-year maternal and child serum levels (n=89). Root mean squared errors (RMSEs) and correlation coefficients (R2s) were calculated comparing the linear fit of back-extrapolated and measured log10 serum levels (red line). Black line represents perfect back-extrapolation. SL = Super Learner and DSA = Deletion/ Substitution/ Addition.