SUPPLEMENTARY MATERIAL

Prenatal exposure to insecticides and weight trajectories among South African children in the VHEMBE birth cohort

Joanne Kim PhD[1], Seungmi Yang PhD[1], Erica E. M. Moodie PhD[1], Muvhulawa Obida[2], Riana Bornman MBChB MD DSc[2], Brenda Eskenazi PhD[3], Jonathan Chevrier PhD[1]

[1]Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, McGill University, Montreal, Quebec, Canada
[2]University of Pretoria Institute for Sustainable Malaria Control, School of Health Systems and Public Health, University of Pretoria, Pretoria, Gauteng, South Africa
[3]Center for Environmental Research and Children's Health, School of Public Health, University of California, Berkeley, California, USA

Contents

eAppendix 1. Selecting the best-fitting SITAR model

Identifying outliers: Prior to fitting the SITAR models, we used the conditional weights approach to identify implausible weight measurements[1] and removed 22 outlier measurements greater than three standard deviations from expected values conditional on preceding measurements; this is an improvement over traditional methods that identify outliers based on cross-sectional population distributions.

As stated in Section 2.5.1, the weight trajectories of each child were modelled using SuperImposition, Translation And Rotation (SITAR)[2]. This model fits a natural cubic spline to the average population growth curve (in this case, weight in kilograms versus age in months). Individual deviations from this mean curve are captured by three random-effects parameters, where:

$$y_{it} = \alpha_i + h\left(\frac{t - \beta_i}{exp(-c\gamma_i)}\right)$$

a) Size (α) indicates the child's mean weight compared to the average (in kilograms), representing vertical translation of the weight curve;
b) Tempo (β) indicates the child's age at peak weight velocity compared to the average (in months), representing horizontal translation of the weight curve; and,
c) Intensity (γ) indicates the child's growth rate compared to the average (expressed as a fraction).

To identify the best-fitting SITAR model, we ran candidate SITAR models with all possible combinations of the following specifications, in the overall cohort and by sex:

i) all three random effects, size and tempo only, or size and intensity only [3 options]
ii) untransformed or log-transformed child weight [2 options]
iii) three to five degrees of freedom for the population average spline [3 options].

We evaluated all candidate SITAR models based on the following criteria: low Akaike information criterion and Bayesian information criterion adjusted for log-transformation (aAIC, aBIC), low correlation between random effects, and high variance explained measured by R^2 generalized to mixed effects models[3]. The diagnostics for all converged models are shown in eTables 1.1 (overall), 1.2 (girls), and 1.3 (boys).

Results: Models with all three random effects did not converge. Among models with two random effects, those with log-transformed weight performed better compared to models based on untransformed weight (eTables 1.1-1.3). The model based on log-transformed weight, random effects for α (size) and β (tempo) random effects, and 3 degrees of freedom for the population average spline explained the highest proportion of variance in the entire cohort as well as in girls only and boys only, therefore these were considered the best-fitting models (eTables 1.1-1.3).

The best-fitting overall and sex-specific models had similar variance explained (75% overall, 76% girls only, 74% boys only; see eTables 1.1-3), similar correlations between random effects (0.44, 0.44, and 0.47, respectively; see eTables 1.1-3), and similar fitted weight velocity curves (see eFigure 1.1); therefore, the more parsimonious overall model was selected as the final model.

eTable 1.1. Diagnostics for converged SITAR models fit on the entire cohort. The colour scale highlights better metrics in green (lower AIC, lower BIC, higher variance explained, lower correlation between random effects), and worse metrics in red (higher AIC, higher BIC, lower variance explained, higher correlation between random effects).

Random effects	Weight variable	df	Adjusted AIC	Adjusted BIC	Variance explained	Correlation between random effects
Size (α), tempo (β)	Weight, kg	4	32,651	32,725	66	0.95
Size (α), tempo (β)	Log(weight)	3	30,788	30,855	75	0.44
Size (α), tempo (β)	Log(weight)	4	30,650	30,724	69	0.44
Size (α), tempo (β)	Log(weight)	5	30,645	30,727	67	0.44
Size (α), intensity (γ)	Weight, kg	4	32,178	32,252	68	0.91
Size (α), intensity (γ)	Weight, kg	5	32,027	32,109	68	0.91
Size (α), intensity (γ)	Log(weight)	4	30,775	30,849	69	0.36
Size (α), intensity (γ)	Log(weight)	5	30,748	30,830	66	0.37

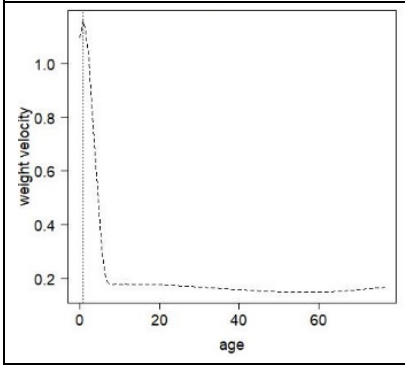
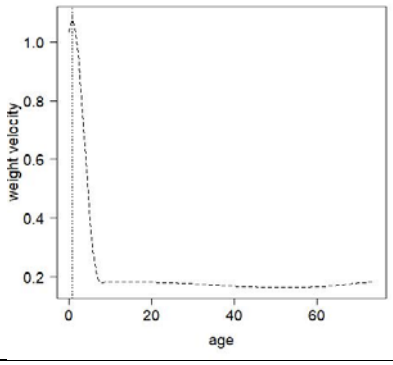
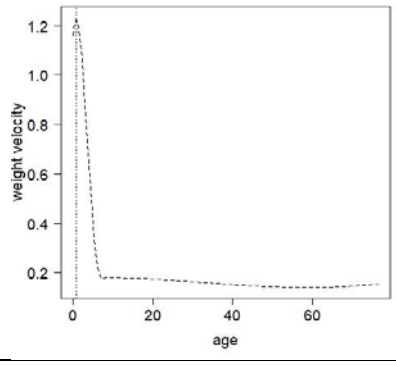
eTable 1.2. Diagnostics for converged SITAR models fit on girls only. The colour scale highlights better metrics in green (lower AIC, lower BIC, higher variance explained, lower correlation between random effects), and worse metrics in red (higher AIC, higher BIC, lower variance explained, higher correlation between random effects).

Random effects	Weight variable	df	Adjusted AIC	Adjusted BIC	Variance explained	Correlation between random effects
Size (α), tempo (β)	Weight, kg	3	16092	16153	67	0.95
Size (α), tempo (β)	Weight, kg	4	16048	16115	66	0.96
Size (α), tempo (β)	Weight, kg	5	16012	16086	66	0.96
Size (α), tempo (β)	Log(weight)	3	14523	14583	76	0.44
Size (α), tempo (β)	Log(weight)	4	14461	14528	71	0.43
Size (α), tempo (β)	Log(weight)	5	14455	14529	69	0.43
Size (α), intensity (γ)	Weight, kg	3	15891	15952	69	0.88
Size (α), intensity (γ)	Weight, kg	4	15538	15605	70	0.9
Size (α), intensity (γ)	Weight, kg	5	15451	15525	70	0.91
Size (α), intensity (γ)	Log(weight)	4	14508	14575	70	0.35
Size (α), intensity (γ)	Log(weight)	5	14502	14576	69	0.35

eTable 1.3. Diagnostics for converged SITAR models fit on boys only. The colour scale highlights better metrics in green (lower AIC, lower BIC, higher variance explained, lower correlation between random effects), and worse metrics in red (higher AIC, higher BIC, lower variance explained, higher correlation between random effects).

Random effects	Weight variable	df	Adjusted AIC	Adjusted BIC	Variance explained	Correlation between random effects
Size (α), tempo (β)	Weight, kg	3	16396	16457	67	0.96
Size (α), tempo (β)	Weight, kg	4	16369	16437	65	0.96
Size (α), tempo (β)	Weight, kg	5	16346	16421	65	0.95
Size (α), tempo (β)	Log(weight)	3	16053	16114	74	0.47
Size (α), tempo (β)	Log(weight)	5	15979	16054	64	0.46
Size (α), intensity (γ)	Weight, kg	3	17158	17219	63	0.89
Size (α), intensity (γ)	Weight, kg	4	16304	16372	66	0.93
Size (α), intensity (γ)	Weight, kg	5	16255	16330	66	0.92
Size (α), intensity (γ)	Log(weight)	4	16055	16123	66	0.40
Size (α), intensity (γ)	Log(weight)	5	16033	16108	63	0.40

eFigure 1.1. Plots of weight velocity (kg/month) vs. age (months) using the best-fitting SITAR models for the entire cohort, girls only, and boys only. All models were based on log-transformed weight, random effects for α (size) and β (tempo) random effects, and 3 degrees of freedom for the population average spline. The dashed vertical line indicates the age at peak weight velocity.

Entire cohort	Girls	Boys
		

eAppendix 2. Multiple imputation by chained equations

We conducted multiple imputation by chained equations using the `mi` suite of commands in Stata version 14.2 (StataCorp, College Station, TX). Continuous variables were imputed using predictive mean matching and binary variables were imputed using logistic regression. We used a burn-in period of 10 iterations and generated 10 imputed datasets[4].

We included in the imputation models all outcomes, exposures and covariates identified in Section 2.5 of the paper, with the exception of two derived variables (food poverty and insufficient maternal energy intake) and included the component variables in the imputation models instead. Specifically, we derived missing values of food poverty from imputed total household income, and derived missing values of insufficient maternal energy intake from imputed maternal age, height, post-delivery weight, and energy intake during pregnancy. We also included auxiliary variables in the imputation models to improve prediction of total household income (auxiliary variable: food poverty at the 1-year study visit) and exclusive breastfeeding (auxiliary variable: total breastfeeding duration)[4].

Finally, we also included variables representing the interaction between the exposures and effect modifiers (sex, food poverty, food insecurity, maternal energy intake sufficiency)[5]. However, to address issues of collinearity and to reduce the number of terms added to the imputation models, we only included interaction terms for three of the seven exposures (*p,p'*-DDT, *cis*-DBCA and *cis*-DCCA) and three of the four effect modifiers (sex, food poverty, and maternal energy intake sufficiency). As stated in our Results section, correlations were high between congeners of DDT/E (Pearson's r= 0.69 to 0.85) and between the pyrethroid metabolites *cis*-DCCA, *trans*-DCCA, and 3-PBA (r= 0.83 to 0.88), therefore the analyte that was most correlated with the other two analytes within the group was selected (*p,p'*-DDT and *cis*-DCCA, respectively); *cis*-DBCA was not correlated with the other pyrethroid metabolites. Among the effect modifiers, food poverty and food insecurity were highly associated with each other ($p<0.001$), therefore only interaction terms with food poverty were created.

eAppendix 3. Inverse-probability of treatment weights (IPTW)

Methods

First, we constructed generalized propensity scores (GPS) for each continuous exposure A, by estimating the conditional density function $f[A|L]$, where L is the vector of covariates. To do so we used multivariable linear regression, with L including the following potential confounders and predictors of the outcomes identified based on a directed acyclic graph (eFigure 3.1): child sex (boy/girl); household food poverty (yes/no), food insecurity (yes/no), and wealth index (continuous); maternal age (years, continuous), height (metres, continuous), post-delivery weight (kg, continuous), education (high school vs. no high school), marital status (married vs. not married), energy intake during pregnancy (insufficient/sufficient), alcohol use during pregnancy (yes/no), HIV status (positive/negative), duration of exclusive breastfeeding (months, continuous), and parity (continuous).

We then generated stabilized inverse-probability of treatment weights (IPTW) with the GPS as the denominator and the marginal exposure density $f[A]$ as the numerator[6]. When the GPS is correctly specified, marginal structural models using stabilized IPTW weights $\frac{f[A]}{f[A|L]}$ result in asymptotically unbiased estimators for the true causal effect[7,8]. To investigate effect measure modification by child sex, food poverty, and energy intake during pregnancy, we fit marginal structural models using IPTW that were stabilized by using the conditional exposure density $f[A|M]$ as the numerator, where M is the effect modifier. Because the GPS already included all effect modifiers M (i.e. M was a subset of L), the GPS was used as the denominator.

Balance diagnostics

We assessed covariate balance before and after weighting by the IPTW for each exposure using three metrics:

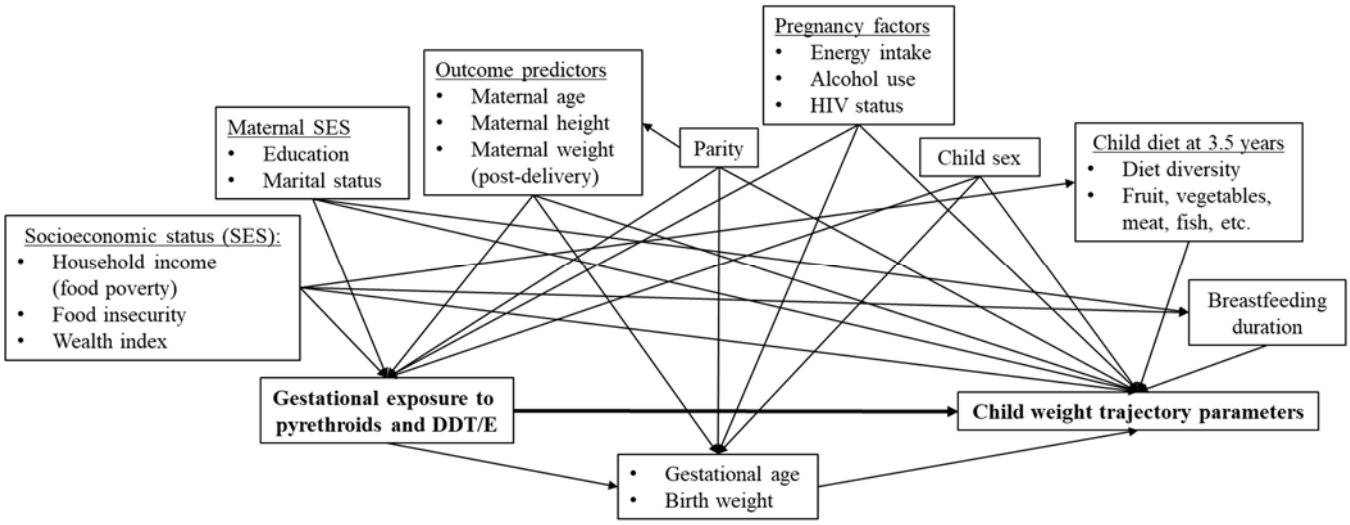
i) Pearson correlation coefficients between the exposure and each continuous covariate, with those below 0.1 indicating balance[9].

ii) Standardized differences comparing all covariates across each quartile of exposure versus all other quartiles, which were then averaged across the four comparisons. Quartiles of exposure were used to avoid small cell sizes and finite sample bias when assessing standardized differences. A threshold of 0.20 for standardized differences has been suggested to indicate balance when conducting balance diagnostics for the GPS[9].

iii) Variance ratios comparing the variance of all covariates across each quartile of exposure versus all other quartiles, which were then averaged across the four comparisons. By definition, a variance ratio of 1.0 describes a covariate which has equal variance across exposure categories, and a threshold of <2.0 has been suggested to indicate balance[10].

Balance was improved through an iterative process. For instance, balance improved for all continuous covariates after $\log_2$-transformation, except for the wealth index, which is a normally-distributed variable and thus was kept un-transformed. In an attempt to further improve balance, we used machine learning algorithms to define our GPS among the candidate confounders and outcome predictors that we identified *a priori*. Using SuperLearner, which creates a weighted average of models fit using different machine learning algorithms and evaluated using cross-validation[11], resulted in overfitting and positivity violations indicated by stabilized weights with a mean not equal to 1 (data not shown) and furthermore did not improve balance diagnostics[12]. Therefore, we proceeded with our more parsimonious original model.

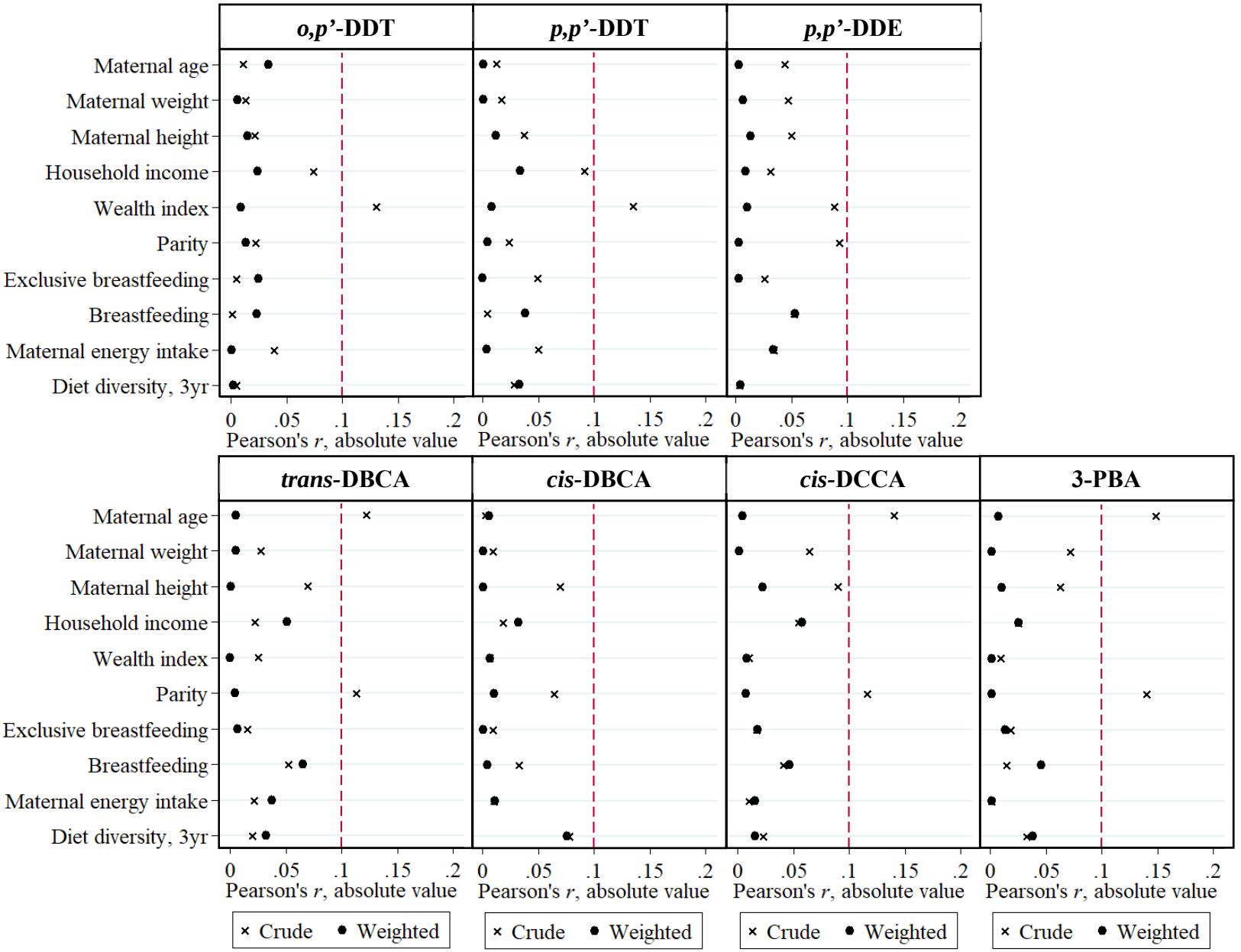
eFigure 3.1. Directed acyclic graph of the relationship between gestational exposure to insecticides and child weight trajectory



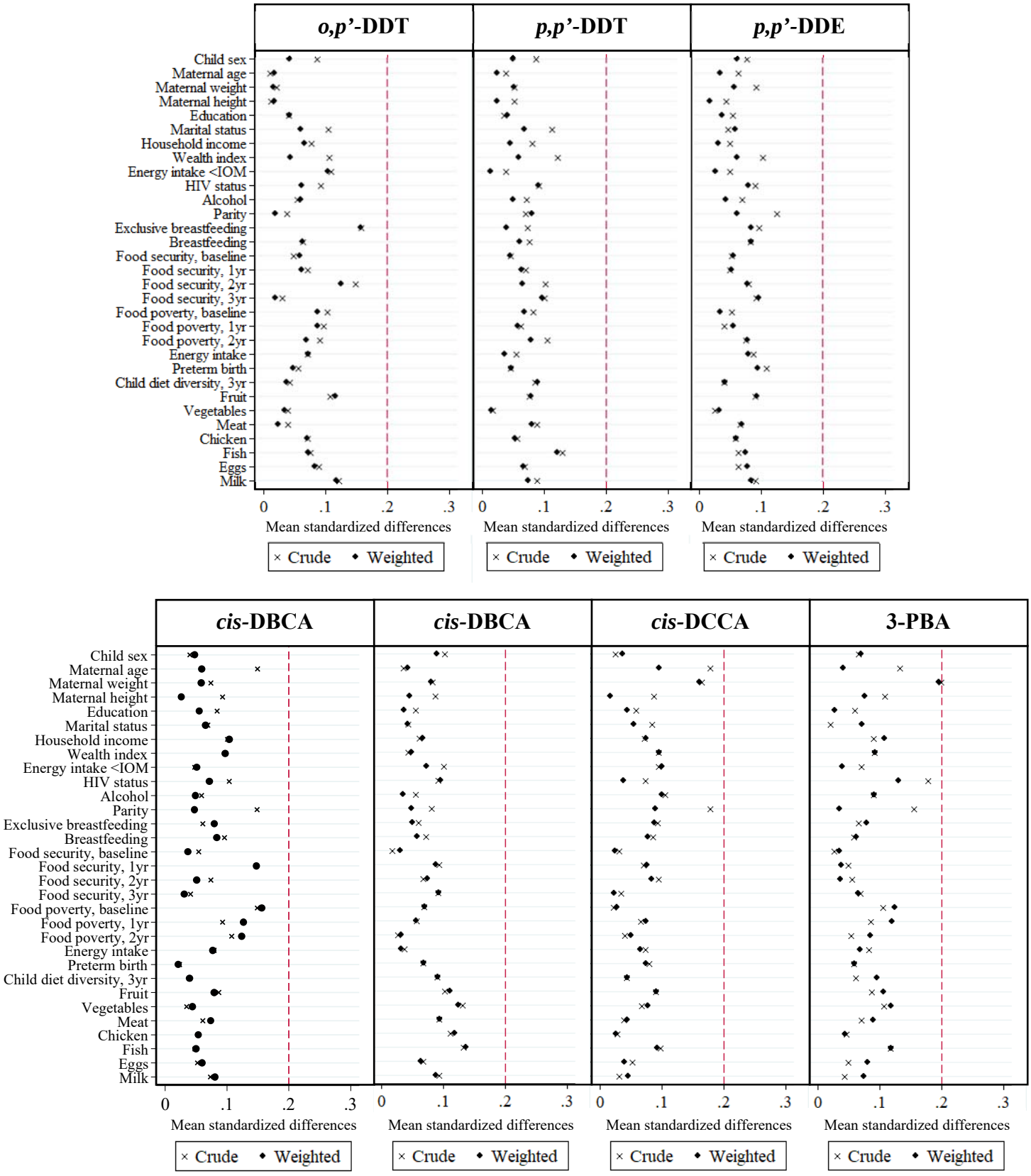
eTable 3.1. Distribution of stabilized inverse probability of treatment weights (IPTW) for each exposure, in overall and effect modification models

Exposure	Model	n	Mean	SD	Min	Max
o,p'-DDT	Overall	751	1.00	0.18	0.27	2.68
	Child sex	751	1.00	0.16	0.22	2.26
	Food poverty	751	1.00	0.17	0.30	2.41
	Food insecurity	751	1.00	0.18	0.29	2.54
	Energy intake	751	1.00	0.18	0.27	2.67
p,p'-DDT	Overall	751	1.00	0.20	0.36	2.29
	Child sex	751	1.00	0.18	0.31	1.97
	Food poverty	751	1.00	0.19	0.40	2.32
	Food insecurity	751	1.00	0.20	0.39	2.45
	Energy intake	751	1.00	0.20	0.38	2.42
p,p'-DDE	Overall	751	1.00	0.20	0.39	2.00
	Child sex	751	1.00	0.20	0.36	2.10
	Food poverty	751	1.00	0.20	0.41	1.92
	Food insecurity	751	1.00	0.20	0.38	2.06
	Energy intake	751	1.00	0.19	0.43	1.90
cis-DBCA	Overall	738	1.00	0.14	0.43	1.76
	Child sex	738	1.00	0.13	0.40	1.85
	Food poverty	738	1.00	0.13	0.46	1.83
	Food insecurity	738	1.00	0.14	0.43	1.78
	Energy intake	738	1.00	0.13	0.41	2.12
cis-DCCA	Overall	738	1.00	0.18	0.29	2.57
	Child sex	738	1.00	0.19	0.30	2.59
	Food poverty	738	1.00	0.19	0.29	2.61
	Food insecurity	738	1.00	0.19	0.27	2.73
	Energy intake	738	1.00	0.19	0.29	2.55
trans-DCCA	Overall	738	1.00	0.17	0.38	2.17
	Child sex	738	1.00	0.17	0.39	2.15
	Food poverty	738	1.00	0.17	0.35	2.11
	Food insecurity	738	1.00	0.16	0.29	1.99
	Energy intake	738	1.00	0.17	0.41	2.30
3-PBA	Overall	737	1.00	0.25	0.18	3.83
	Child sex	737	1.00	0.24	0.22	3.62
	Food poverty	737	1.00	0.25	0.17	3.76
	Food insecurity	737	1.00	0.24	0.15	3.52
	Energy intake	737	1.00	0.24	0.16	3.47

eFigure 3.2. Correlations between each exposure and continuous potential confounders, before (×) and after (●) IPTW-weighting



eFigure 3.3. Standardized differences of all potential confounders, averaged across exposure quartiles, before (×) and after (●) IPTW-weighting



eAppendix 4. Outcome regressions: effect measure modification by food poverty, food insecurity and maternal energy intake during pregnancy

eTable 4.1. Effects of a 10-fold increase in maternal peripartum DDT/E (ng/g lipid) or pyrethroid metabolite (μg/L) concentrations on birth to 5-year weight trajectory parameters, by food poverty status among children participating in the VHEMBE study, Limpopo, South Africa.

	n	No food poverty β (95% CI)	Food poverty β (95% CI)	p-value, interaction
Size (grams)				
o,p'-DDT	751	1.9 (-18.8, 22.6)	1.8 (-13.2, 16.8)	0.99
p,p'-DDT	751	7.0 (-9.2, 23.2)	1.9 (-10.3, 14.1)	0.63
p,p'-DDE	751	6.9 (-13.9, 27.8)	6.7 (-8.7, 22.0)	0.99
cis-DBCA	738	-23.8 (-52.6, 5.0)	-7.6 (-27.6, 12.4)	0.38
cis-DCCA	738	-12.6 (-47.1, 22.0)	-4.0 (-27.8, 19.8)	0.70
trans-DCCA	738	-11.4 (-37.6, 14.9)	-4.8 (-23.3, 13.6)	0.70
3-PBA	737	1.6 (-35.2, 38.4)	0.4 (-23.1, 24.0)	0.96
Tempo (days)				
o,p'-DDT	751	-0.3 (-2.7, 2.1)	-0.3 (-1.9, 1.3)	1.00
p,p'-DDT	751	-0.4 (-2.1, 1.3)	-0.3 (-1.6, 1.0)	0.93
p,p'-DDE	751	-0.1 (-2.4, 2.2)	-0.3 (-1.9, 1.3)	0.92
cis-DBCA	738	-1.0 (-4.6, 2.6)	-2.2 (-4.5, 0.1)	0.59
cis-DCCA	738	-2.5 (-5.2, 0.2)	-1.4 (-4.5, 1.6)	0.62
trans-DCCA	738	-1.4 (-4.0, 1.2)	-1.4 (-3.9, 1.1)	0.99
3-PBA	737	-2.6 (-6.3, 1.1)	-1.8 (-5.3, 1.7)	0.76

Abbreviations: CI, confidence interval; DDE, dichlorodiphenyldichloroethylene; DDT, dichlorodiphenyltrichloroethane; *cis*-DBCA, *cis*-3-(2,2-dibromovinyl)-2,2-dimethyl-cyclopropane carboxylic acid; *cis*-DCCA, *cis*-3-(2,2,-dicholorvinyl)-2,2-dimethyl-cyclopropane carboxylic acid; *trans*-DCCA, *trans*-3-(2,2,-dicholorvinyl)-2,2-dimethyl-cyclopropane carboxylic acid; 3-PBA, 3-phenoxybenzoic acid.

eTable 4.2. Effects of a 10-fold increase in maternal peripartum DDT/E (ng/g lipid) or pyrethroid metabolite (μg/L) concentrations on birth to 5-year weight trajectory parameters, by food insecurity status among children participating in the VHEMBE study, Limpopo, South Africa.

	n	Food secure β (95% CI)	Food insecure β (95% CI)	p-value, interaction
Size (grams)				
o,p'-DDT	751	2.8 (-14.4, 20.0)	0.4 (-17.6, 18.4)	0.86
p,p'-DDT	751	-1.2 (-14.5, 12.1)	8.2 (-6.9, 23.2)	0.38
p,p'-DDE	751	0.0 (-15.8, 15.9)	14.7 (-5.5, 34.9)	0.29
cis-DBCA	738	-16.1 (-38.2, 5.9)	-9.4 (-34.0, 15.2)	0.70
cis-DCCA	738	-6.6 (-33.4, 20.2)	-7.3 (-34.1, 19.4)	0.97
trans-DCCA	738	-4.5 (-24.8, 15.9)	-9.5 (-30.7, 11.8)	0.75
3-PBA	737	2.9 (-21.7, 27.5)	-3.1 (-31.7, 25.6)	0.76
Tempo (days)				
o,p'-DDT	751	-0.2 (-2.1, 1.7)	-0.5 (-2.4, 1.4)	0.84
p,p'-DDT	751	-0.6 (-2.0, 0.7)	-0.1 (-1.5, 1.4)	0.56
p,p'-DDE	751	-0.5 (-2.3, 1.3)	0.1 (-1.8, 1.9)	0.64
cis-DBCA	738	-1.1 (-3.9, 1.6)	-2.6 (-5.3, 0.2)	0.48
cis-DCCA	738	-3.2 (-5.9,-0.5)	-0.2 (-3.6, 3.2)	0.18
trans-DCCA	738	-2.0 (-4.3, 0.3)	-0.7 (-3.5, 2.2)	0.47
3-PBA	737	-3.2 (-6.5, 0.1)	-1.1 (-5.0, 2.9)	0.41

Abbreviations: CI, confidence interval; DDE, dichlorodiphenyldichloroethylene; DDT, dichlorodiphenyltrichloroethane; *cis*-DBCA, *cis*-3-(2,2-dibromovinyl)-2,2-dimethyl-cyclopropane carboxylic acid; *cis*-DCCA, *cis*-3-(2,2,-dicholorvinyl)-2,2-dimethyl-cyclopropane carboxylic acid; *trans*-DCCA, *trans*-3-(2,2,-dicholorvinyl)-2,2-dimethyl-cyclopropane carboxylic acid; 3-PBA, 3-phenoxybenzoic acid.

eTable 4.3. Effects of a 10-fold increase in maternal peripartum DDT/E (ng/g lipid) or pyrethroid metabolite (μg/L) concentrations on birth to 5-year weight trajectory parameters, by maternal energy intake sufficiency during pregnancy among children participating in the VHEMBE study, Limpopo, South Africa.

	n	Sufficient β (95% CI)	Insufficient β (95% CI)	p-value, interaction
Size (grams)				
o,p'-DDT	751	14.2 (-11.0, 39.5)	-3.2 (-17.2, 10.9)	0.25
p,p'-DDT	751	14.3 (-7.3, 35.9)	-2.3 (-13.0, 8.4)	0.19
p,p'-DDE	751	16.2 (-10.1, 42.6)	1.9 (-11.0, 14.8)	0.35
cis-DBCA	738	-12.7 (-41.5, 16.0)	-13.9 (-34.8, 7.1)	0.95
cis-DCCA	738	-22.3 (-53.5, 9.0)	0.8 (-24.5, 26.1)	0.28
trans-DCCA	738	-13.5 (-37.9, 11.0)	-3.8 (-22.6, 15.0)	0.56
3-PBA	737	-17.0 (-48.0, 14.0)	7.0 (-17.0, 31.1)	0.24
Tempo (days)				
o,p'-DDT	751	-0.4 (-2.9, 2.1)	-0.3 (-2.0, 1.3)	0.98
p,p'-DDT	751	0.0 (-1.8, 1.8)	-0.6 (-1.9, 0.7)	0.59
p,p'-DDE	751	0.5 (-1.5, 2.5)	-0.6 (-2.2, 1.1)	0.41
cis-DBCA	738	-3.1 (-6.3, 0.1)	-1.1 (-3.5, 1.4)	0.34
cis-DCCA	738	-2.3 (-5.3, 0.8)	-1.6 (-4.4, 1.2)	0.76
trans-DCCA	738	-2.1 (-5.0, 0.8)	-1.0 (-3.3, 1.2)	0.55
3-PBA	737	-2.0 (-6.1, 2.1)	-2.3 (-5.5, 0.9)	0.90

Abbreviations: CI, confidence interval; DDE, dichlorodiphenyldichloroethylene; DDT, dichlorodiphenyltrichloroethane; *cis*-DBCA, *cis*-3-(2,2-dibromovinyl)-2,2-dimethyl-cyclopropane carboxylic acid; *cis*-DCCA, *cis*-3-(2,2,-dicholorvinyl)-2,2-dimethyl-cyclopropane carboxylic acid; *trans*-DCCA, *trans*-3-(2,2,-dicholorvinyl)-2,2-dimethyl-cyclopropane carboxylic acid; 3-PBA, 3-phenoxybenzoic acid.

References

1. Yang S, Hutcheon JA. Identifying outliers and implausible values in growth trajectory data. *Ann Epidemiol* 2016;**26**(1):77-80 e1-2.

2. Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. *Int J Epidemiol* 2010;**39**(6):1558-66.

3. Cole TJ. Super Imposition by Translation and Rotation growth curve analysis. R package version 1.1.1. ed. http://cran.r-project.org/package=sitar, 2019.

4. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;**30**(4):377-99.

5. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Med Res Methodol* 2012;**12**:46.

6. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;**168**(6):656-64.

7. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006;**60**(7):578-86.

8. Robins JM. Marginal Structural Models versus Structural nested Models as Tools for Causal inference. In: Halloran ME, Berry D, eds. Statistical Models in Epidemiology, the Environment, and Clinical Trials. New York, NY: Springer New York, 2000;95-133.

9. Austin PC. Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat Methods Med Res* 2019;**28**(5):1365-1377.

10. Zhang Z, Kim HJ, Lonjon G, Zhu Y, written on behalf of AMEB-DCTCG. Balance diagnostics after propensity score matching. *Ann Transl Med* 2019;**7**(1):16.

11. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;**6**:Article25.

12. Alam S, Moodie EEM, Stephens DA. Should a propensity score model be super? The utility of ensemble procedures for causal adjustment. *Statistics in Medicine* 2019;**38**(9):1690-1702.